21

CHARACTERISATION OF THE HUMAN α2(I) PROCOLLAGEN PROMOTER-BINDING PROTEINS

Thesis presented by
MALCOLM ROBERT COLLINS

in fulfillment of the degree of
DOCTOR OF PHILOSOPHY

in
MEDICAL BIOCHEMISTRY

in the
FACULTY OF MEDICINE
UNIVERSITY OF CAPE TOWN

September 1993

CERTIFICATE OF SUPERVISOR

In terms of paragraph eight of the 'General regulations for the degree of Ph.D.', I, as supervisor of the candidate M.R. COLLINS, certify that I approve of the incorporation into this thesis of material that has already been published or submitted for publication.

Prof. M.I. Parker
Department of Medical Biochemistry
D.G. Murray Trust Chair of Cancer Research

ACKNOWLEDGEMENTS

My supervisor, Prof. M.I. Parker, for his guidance, encouragement and patience throughout the duration of this thesis.

Mrs Z. Rossouw for excellent technical assistance in tissue culture.

Drs E. Merrifield and T. Haylett for helpful suggestions on gel filtration and heparin-agarose chromatography.

Dr L. Kellaway for the use of his computer and Laser printer to print the figures and graphs in this thesis.

My other past and present colleagues in the laboratory (A. Smith, Q. Eichbaum, P. Goldberg, J. te Water Naude, V. Leaner, D. Bhoolia, H. Donninger, G. Fenhalls, S. Herman, E. Brice, M. Hall, M. Ziman and M. Madikizela) and the rest of the Department of medical Biochemistry for helpful suggestions and providing a pleasant working environment.

Judy - thank you for your friendship and support over the years and especially the last few months.

Charles Howell, Barry and Maureen van Eyssen for demonstrating true Christian love and support.

My parents for giving me the opportunity to study all these years.

This work was supported by the Medical Research Council, UCT Research Scholarship and Marion Beatrice Waddell and Yeoman Bequest Bursaries.

CONTENTS

Title Page i
Certificate of Supervisor ii
Acknowledgements iii
Contents iv
Abbreviations viii
Dedication x
Abstract xi

CHAPTER 1 INTRODUCTION

1.1 OVERVIEW 1
1.2 THE COLLAGEN GENES 4
1.2.1 Structure of the fibrillar collagen genes 5
1.2.2 Short chain collagen genes 9
1.2.3 Fibril-associated collagens with interrupted triple-helices (FACIT) and FACIT-like genes 10
1.2.4 Extracellular matrix networks 13
1.3 REGULATION OF COLLAGEN GENE EXPRESSION 16
1.3.1 Chromatin conformation 16
1.3.2 Methylation status 17
1.3.3 Post-transcriptional control mechanisms 17
1.3.4 Regulation by exogenous factors 18
1.3.5 Transformed cells 19
1.3.6 Transcription factors 19
1.3.6.1 Type I collagen gene 22
1.3.6.2 Type II collagen gene 29
1.3.6.3 Type III collagen gene 30
1.3.6.4 Type IV collagen gene 31
1.3.6.5 Type V collagen gene 33
1.3.6.6 Type VI collagen gene 34
1.3.6.7 Type IX collagen gene 35
1.4 THE CCAAT BOX BINDING FACTORS 35
1.4.1 CCAAT transcription factor/nuclear factor 1 (CTF/NF1) 36
1.4.2 CCAAT/enhancer binding protein (C/EBP) and related proteins 37
1.4.3 The heteromultimeric CCAAT binding factors 38
1.4.4 CCAAT binding proteins and cell-cycle regulation 42
1.4.5 Other CCAAT binding proteins 44

CHAPTER 2 IDENTIFICATION OF DNA-BINDING PROTEINS WHICH INTERACT WITH THE PROXIMAL α2(I) PROCOLLAGEN GENE PROMOTER

2.1 INTRODUCTION 45

2.2 RESULTS 48

2.2.1 Characterisation of the SVWI-38 α2(I) proximal promoter 48

2.2.1.1 Subcloning 48

2.2.1.2 Sequence analysis 50

2.2.2 Candidate transcription factors which could regulate expression of the α2(I) procollagen gene 50

2.2.3 Subcloning the proximal human α2(I) procollagen promoter 52

2.2.3.1 Plasmid pNJ-59 52

2.2.3.2 Plasmid pNJ-230 55

2.2.3.3 Plasmid pNJ-170 55

2.2.4 Analysis of DNA-protein interaction by electrophoretic mobility shift assays 55

2.3 DISCUSSION 63

CHAPTER 3 THE PARTIAL PURIFICATION OF FIBROBLAST DNA-BINDING PROTEINS

3.1 INTRODUCTION 65

3.2 RESULTS 67

3.2.1 Differential salt extraction of nuclear proteins 67

3.2.2 Heparin-agarose chromatography 67

3.3 DISCUSSION 76

CHAPTER 4 OVERLAPPING TRANSCRIPTION-FACTOR BINDING-SITES WITHIN THE PROXIMAL HUMAN α2(I) PROCOLLAGEN PROMOTER

4.1 INTRODUCTION 77

4.2 RESULTS 79

4.2.1 Analysis of *cis*-elements in the proximal human α2(I) procollagen promoter 79

4.2.2 Identification of DNA-elements involved in DNA-protein complex formation 80

4.2.3 Mutually exclusive DNA-protein complex formation 90

4.2.4 Methylation interference analysis 93

4.2.5 Mutation analysis 102
4.2.6 Delineation of G/CBE and CME 105
4.3 DISCUSSION 113

CHAPTER 5 CHARACTERISATION OF THE PROMOTER BINDING PROTEINS

5.1 INTRODUCTION 117
5.2 RESULTS 119
5.2.1 Native molecular weight determination 119
5.2.1.1 Analytical gel filtration 119
5.2.1.2 Glycerol gradient sedimentation 124
5.2.1.3 Calculation of native molecular weights 131
5.2.2 Polypeptide composition of the complexes 132
5.2.2.1 UV-crosslinking 132
5.2.2.2 Denaturing gel filtration 140
5.2.2.3 Southwestern blotting 141
5.2.3 Protein-protein interactions are not DNA-dependent 147
5.3 DISCUSSION 152

CHAPTER 6 CONCLUSION 156

CHAPTER 7 MATERIALS AND METHODS

7.1 SUBCLONING AND PLASMID DNA ISOLATION 161
7.1.1 Preparation of DNA fragments 161
7.1.1.1 Crush soak method 161
7.1.1.2 Electro-elution 161
7.1.1.3 Electrophoresis into Whatman 3MM paper 162
7.1.1.4 QIAEX extraction protocol 162
7.1.2 Blunt-end ligation 162
7.1.2.1 Filling of recessed 3'-ends 162
7.1.2.2 Blunt ending with T4 DNA polymerase 163
7.1.2.3 Ligation 163
7.1.3 Preparation and transformation of competent bacterial cells 163
7.1.4 Screening of bacterial colonies 164
7.1.5 Large-scale plasmid DNA isolation 165
7.1.5.1 CsCl-ethidium bromide centrifugation 165

7.1.5.2 QIAGEN columns 166
7.1.6 DNA sequencing 166
7.2 CELL CULTURE 166
7.3 PREPARATION AND FRACTIONATION OF CRUDE NUCLEAR EXTRACTS 167
7.3.1 Preparation of Dignam nuclear extracts 167
7.3.2 Small-scale nuclear extract preparation 168
7.3.3 Differential salt extraction 168
7.3.4 Heparin-agarose chromatography 169
7.4 DNA-PROTEIN INTERACTIONS 169
7.4.1 End-labelling of DNA 169
7.4.2 Electrophoretic mobility shift assays 170
7.4.3 Mutually exclusive binding assays 171
7.4.4 Methylation interference assay 171
7.4.5 Oligonucleotides synthesis and annealing 173
7.5 SDS-POLYACRYLAMIDE GEL ELECTROPHORESIS 174
7.6 PHYSICO-CHEMICAL ANALYSIS OF DNA-BINDING PROTEINS 175
7.6.1 Analytical gel filtration chromatography 175
7.6.2 Glycerol gradient sedimentation 177
7.6.3 UV-crosslinking 178
7.6.3.1 Preparation of Br-dU substituted probes 178
7.6.3.2 *In-situ* DNA-protein crosslinking 178
7.6.3.3 DNA-protein crosslinking prior to electrophoresis 179
7.6.4 Southwestern blotting 179
7.7 BUFFERS AND SOLUTIONS 181

REFERENCES 187

APPENDIX A 207

ABBREVIATIONS

ATCC - American Type Culture Collection
ATP - adenosine triphosphate
BIS - N,N'-methylenediacrylamide
BME - eagle's basal medium
Br-dU - bromodeoxyuridine
Br-dUTP - bromodeoxyuridine triphosphate
BSA - bovine serum albumin
bp - base pairs
cDNA - complementary DNA
CIAP - calf intestinal alkalinal phosphatase
CME - collagen modulating element
dATP - deoxyadenosine triphosphate
dC - deoxycytosine
dCTP - deoxycytidine triphosphate
DEAE - diethylaminoethyl
dGTP - deoxyguanosine triphosphate
DMS - dimethyl sulphate
DNA - deoxyribonucleic acid
DNase - deoxyribonuclease
dNTP - deoxynucleoside triphosphate
DTT - dithiothreitol
dTTP - deoxythymidine triphosphate
EDTA - ethylenediaminetetra-acetic acid
EGTA - ethyleneglycol-bis-(ß-aminoethyl ether) N,N'-tetraacetic acid
EMSA - electrophoretic mobility shift assay
G/CBE - GGAGG/CCAAT binding element
G/CBF - GGAGG/CCAAT box factor
HEPES - N-2-hydroxyethyl piperazine-N'-2-ethanesulfonic acid
Kb - kilobases
KDa - kilodaltons
L-Broth - Luria broth
LTR - long terminal repeat
mRNA - messenger ribonucleic acid
MW - molecular weight
NP-40 - nonident P-40
Oligo - oligonucleotide
PAGE - polyacrylamide gel electrophoresis
PBS - phosphate-buffered saline

PCV - packed cell volume
PIPES - Piperazine-N,N'-bis(2-ethanesulphonic acid)
PMSF - phenylmethylsulphonylfluoride
R_s - Stokes radius
$S_{20,W}$ - sedimentation coefficient at 20^{o} C in water
SDS - sodium dodecyl sulphate
SV40 - simian virus 40
TEMED - N,N,N',N'-tetramethylethylenediamine
Tris - tris(hydroxymethyl)aminomethane
tRNA - transfer ribonucleic acid
UTR - untranslated region
V_e - elution volume
V_o - void volume
V_t - total volume

Nucleotide Codes

A - adenine
C - cytosine
G - guanine
T - Thymine

R - A or G (pyrimidine)
Y - T or C (purine)

K - T or G
M - A or G
S - G or C
W - A or T

B - G, T or C
D - G, A or C
H - A, C or T
V - A, G or C

N - A, C, T or G

IN MEMORY OF THOSE WHO DIED AND WERE INJURED IN THE ATTACK ON ST JAMES CHURCH, KENILWORTH ON 25 JULY 1993

The heavens declare the glory of God;
the skies declare the work of his hands.
Day after day they pour forth speech;
night after night they display knowledge.
There is no speech nor language
where their voice is not heard.
Their voice goes out to all the earth,
their words to the ends of the world.

Psalm 19:1-4

ABSTRACT

In an attempt to elucidate the transcriptional mechanisms that regulate the expression of the human α2(I) procollagen gene, *cis*-acting DNA-elements within the proximal promoter were identified and their corresponding *trans*-acting factors characterised. The fibroblast cell lines used in this study had previously been transformed with either simian virus 40 (SVWI-38) or by γ-radiation (CT-1). The SVWI-38 fibroblasts do not produce any α2(I) collagen chains, whereas the CT-1 cell line produces normal type I collagen.

Previous studies suggested that *trans*-acting factor(s) may be responsible for the inactivation of the α2(I) procollagen gene in SVWI-38 fibroblasts (Parker *et. al.* (1989) *J. Biol. Chem* **264,** 7147-7152; Parker *et. al.* (1992) *Nucleic Acids Res.* **20,** 5825-5830). In this study, the SVWI-38 proximal promoter (-350 to +54) was sequenced and shown to be normal, thereby ruling out any possibility that mutations within this region was responsible for inactivation of the gene.

DNA-protein complexes between the proximal promoter fragment and SVWI-38 and CT-1 nuclear extracts were identified using the electrophoretic mobility shift assay. Nuclear extracts from CT-1 fibroblasts or the parental WI-38 cell line formed two DNA-protein complexes (complexes I and III) with the -110 to -46 promoter fragment. SVWI-38 nuclear extracts on the other hand, formed an additional complex, complex II, which was also present in a transformed mouse macrophage cell line, $P388D_1$. SVWI-38 fibroblasts produced much lower levels of complex III compared to CT-1 cells, while $P388D_1$ cells had no detectable levels of this complex. It is therefore tempting to speculate that the relative ratios of complexes II and III may modulate the levels of α2(I) procollagen gene expression.

Competition and methylation interference assays identified two distinct but overlapping DNA-recognition elements within this 60 bp region of the human α2(I) procollagen promoter. Complex I formation involved a 21 bp element, $^{5'}$GCCCTCCCATTGGTGGAGGCC$^{3'}$, extending from nucleotides -92 to -72. This element, the GGAGG/CCAAT box element (G/CBE), contains a downstream GGAGG box, and an upstream inverted

GGAGG box separated by seven nucleotides containing an inverted CCAAT box. The second *cis*-element, $^{5'}$GGAGGCCCTTTT$^{3'}$, extends from nucleotides -78 to -67 and contains the downstream GGAGG box, which also forms part of the G/CBE, and 3'-flanking sequences. CT-1 nuclear extracts formed one complex with this novel element (complex III), whereas SVWI-38 extracts formed two different complexes (II and III). This 12 bp element has been named the collagen modulating element (CME) because complex II comprises a candidate repressor protein(s).

Further characterisation of the G/CBE suggested that the inverted upstream GGAGG and CCAAT boxes were more important motifs for G/CBF (GGAGG/CCAAT binding factor) (complex I proteins) binding. Deletion of a single cytosine which separates these two motifs abolished or drastically reduced the formation of complex I. Deletion and mutagenesis assays also suggested that the downstream GGAGG box is only weakly involved in DNA-protein contacts and loosely forms part of the G/CBE. Physico-chemical analysis of G/CBF suggested that this factor is identical or related to the previously characterised hetero-multimeric CCAAT-binding factors, NF-Y, CBF, α-CP1 and others. Under non-denaturing conditions, G/CBF had a Stokes radius of 5.22 nm, a sedimentation coefficient of 4.9 S, a calculated native molecular weight of 105 KDa and a frictional ratio of 1.66. UV-crosslinking and denaturing gel filtration data suggested that this factor, like the other hetero-multimeric CCAAT-binding factors, is a multicomponent CCAAT box binding protein.

Mutations of the downstream GGAGG box and the cytosine dinucleotides at positions -72 and -73 resulted in a drastic reduction of complex II and III formation. Other DNA-binding data suggested that the upstream GGAGG and CCAAT boxes were involved in weak DNA-protein contacts with complex II and III proteins. The CME binding proteins (complexes II and III) had Stokes radii of 4.12 and 3.15 nm, sedimentation coefficients of 3.9 and 3.2 S, native molecular weights of 66 and 41 KDa and frictional ratios of 1.54 and 1.38 respectively. The Stokes radii of both these factors did not change during fractionation on gel filtration columns in the presence of 0.8 M KCl. The non-denaturing gel filtration,

Southwestern blotting, and UV-crosslinking data suggested that complex II consists of a stable multimeric protein complex. The physico-chemical analysis suggested that complex III, on the other hand, is monomeric.

Competition binding assays demonstrated that complex I, II and III proteins bound to their respective sites in a mutually exclusive manner. These results indicated that G/CBF and complex III proteins probably fine-tune regulate the expression of the human α2(I) procollagen gene in type I collagen producing cells. Complex II proteins, on the other hand, could repress the expression of this gene via competition with complex III proteins for binding to the CME.

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW

The collagens are a large heterogeneous family of structural proteins which form the major constituents of the extracellular matrix. The structure and function of the different collagen types have been extensively reviewed (for example Bornstein and Sage, 1980; Mayne and Burgeson, 1987; Burgeson, 1988; van der Rest and Garrone, 1991). The collagens are important components of tendon, bone and cornea (types I, III and V), hyaline cartilage (types II and XI), basement membrane (type IV) and the anchoring fibril network (type VII). Other collagens, such as types IX, XII and XIV, associate with the surface of collagenous fibrils where they probably function by establishing and stabilising the spatial arrangement of the fibrils.

To date, 13 distinct collagen types, types I to XII and XIV, have been characterised at both the protein and nucleic acid levels. Although these 13 collagen types are structurally and functionally different, they are all trimers consisting of three polypeptide subunits with continuous or interrupted rod-like triple-helical domains. The triple helix consists of repeating Gly-X-Y tripeptides, where X and Y are often proline and hydroxyproline residues respectively. Some of these collagen molecules, such as types II and III, are homotrimers, while the majority are heterotrimers consisting of two (eg type I) or even three (eg type XI) different types of α-chains. A summary of the α-chain composition of the known collagen types is given in table 1.1. All collagens also contain two or more globular or non-collagenous domains ranging in size from a few to several hundred residues. The non-collagenous domains generally accounts for the major part of the protein. There are exceptions, such as the

Table 1.1 Classification, chain composition and tissue distribution of the different collagen types. The name, chromosomal localisation and the size of the 31 distinct collagen genes and the number of exons are tabulated.

TYPE	CONSTITUENT CHAINS	CHAIN COMPOSITION	GENE LOCUS	CHROMOSOMAL LOCALIZATION	GENE SIZE (Kb)	NUMBER EXONS	TISSUE DISTRIBUTION
(A) FIBRILLAR COLLAGENS							
(i) Major fibrillar collagens							
I	α1(I)	$[\alpha1(I)]_2\alpha2(I)$	COL1A1	17q21.3-22	18	52	Ubiquitous, but mainly bone, skin,
	α2(I)	$[\alpha1(I)]_3$	COL1A2	7q21.3-22	38	52	teeth, tendon and ligaments
II	α1(II)	$[\alpha1(II)]_3$	COL2A1	12q14.3	30	54	Hyaline cartilage
III	α1(III)	$[\alpha1(III)]_3$	COL3A1	2q24.3-31	21	52	Blood vessels, lung and skin
(ii) Minor fibrillar collagens							
V	α1(V)	$[\alpha1(V)]_2\alpha2(V)$	COL5A1	9q34.3			Widely distributed in many tissues
	α2(V)	α1(V)α2(V)α3(V)	COL5A2	2q24.3-31			
	α3(V)	and other forms	COL5A3	n.d.			
	α1'(V)[a]		n.d.	n.d.			
XI	α1(XI)	α1(XI)α2(XI)α3(XI)	COL11A1	1p21			Hyaline cartilage
	α2(XI)	and other forms	COL11A2	6p212			
	α3(XI)		COL2A1	12q13-14	30	54	
(B) NON-FIBRILLAR COLLAGENS							
(i) Short chain collagens							
VIII	α1(VIII)	$[\alpha1(VIII)]_2\alpha2(VIII)$	COL8A1	3q12-13.1		2	Vascular, corneal endothellial
	α2(VIII)	and pos homotrimers	COL8A2	1p32.2-34.3			cells and other tissues
X	α1(X)	$[\alpha1(X)]_3$	COL10A1	6q21-22	6.3	2	Hypertrophic hyaline cartilage
(ii) Fibril-associated collagens with interrupted triple-helices (FACIT)							
IX	α1(IX)	α1(IX)α2(IX)α3(IX)	COL9A1	6q12-14	100[b]		Hyaline cartilage and other
	α2(IX)		COL9A2	1	10[b]	32[b]	tissues
	α3(IX)		COL9A3	n.d.			
XII	α1(XII)	$[\alpha1(XII)]_3$	COL12A1	6			Resembles that of type I
XIV	α1(XIV)	$[\alpha1(XIV)]_3$	COL14A1	n.d.			Resembles that of type I
(iii) FACIT-like collagens							
XVI	α1(XVI)	unknown	COL16A1	1p13-34			
Y	α1(Y)	unknown	COLYA1	6q12-14			
(iv) Extracellular matrix networks							
IV	α1(IV)	$[\alpha1(IV)]_2\alpha2(IV)$	COL4A1	13q34	100	52	Basement membranes
	α2(IV)	and other forms	COL4A2	13q34			
	α3(IV)		COL4A3	2q35-37			
	α4(IV)		COL4A4	2q35-37			
	α5(IV)		COL4A5	Xq21			

[a] Some evidence for its existence (Fessler et. al., 1985; Fessler and Fessler, 1987)

[b] For the chicken gene

Table 1.1 continued

TYPE	CONSTITUENT CHAINS	CHAIN COMPOSITION	GENE LOCUS	CHROMOSOMAL LOCALIZATION	GENE SIZE (Kb)	NUMBER EXONS	TISSUE DISTRIBUTION
VI	α1(VI)	α1(VI)α2(VI)α3(VI)	COL6A1	21q22.3			Wide variety of tissues
	α2(VI)		COL6A2	21q22.3	36	30	
	α3(VI)		COL6A3	2q37			
VII	α1(VII)	$[\alpha1(VII)]_3$	COL7A1	3p21.1-21.3	32	118	Basement membrane of stratified squamous epithelia
(C) PREDICTED COLLAGENS							
XIII	α1(XIII)	unknown	COL13A1	10q11-qter			
XV	α1(XV)	unknown	COL15A1	9q34-35			Resembling that of type I
XVII	α1(XVII)	unknown	COL17A1	10q23.4			

non-collagenous domains of type VI collagen which comprises about 80% of the protein (Saitta *et. al.*, 1992). The existence of five additional putative collagen types, types XIII (Pihlajaniemi and Tamminen, 1990), XV (Myers *et. al.*, 1992), XVI (Pan *et. al.*, 1992), Y (Yoshioka *et. al.*, 1992) and XVII (Li *et. al.*, 1993), has been predicted from deduced amino acid sequences of complementary and/or genomic clones.

Each collagen type has a characteristic cell-type or tissue-specific distribution. Type II collagen is the major collagenous protein of hyaline cartilage and is almost entirely limited to this tissue (van der Rest and Garrone, 1991). Type VII collagen also has a limited anatomical distribution (Parente *et. al.*, 1991), whereas, type I collagen is ubiquitous, with high levels in bone, skin, tendons and ligaments (van der Rest and Garrone, 1991). Types XII, XIV and XV collagens also have a tissue distribution pattern which resembles that of type I collagen (Shaw and Olsen, 1991; Myers *et. al.*, 1992). Type III collagen is usually co-expressed with type I but at different ratios in different organs and tissues (Fessler and Fessler, 1978; Kuhn, 1981).

Collagen levels differ during embryogenesis, development and in the adult organism (reviewed by Bornstein and Sage, 1989). During murine embryogenesis, for example, inactivation of the α1(I) procollagen gene is lethal (Schnieke *et. al.*, 1983; Hartung *et. al.*, 1986). Finally, the shift from the synthesis of type I collagen to type II collagen during the differentiation of mesenchymal cells into chondrocytes is an example of altered collagen gene expression during differentiation (reviewed by von der Mark, 1980; Linsenmayer, 1981).

The steady state levels of the different α-chains making up the heterotrimeric collagens are co-ordinately regulated. In the case of type I collagen, the steady state mRNA levels has a stoichometry of 2:1 which is identical to the ratio of the α1(I) and α2(I) chains in the triple helix (Vuust *et. al.*, 1985). Furthermore, collagen levels are altered during pathological processes such as wound healing, inflammatory conditions, arthritis and fibrosis (reviewed by Bornstein and Sage, 1980). The expression of the various procollagen genes, therefore, has to be tightly and co-ordinately regulated in specific cell and tissue types during the normal development of an organism. Numerous studies have been initiated to characterise the *cis*-acting DNA elements and *trans*-acting factors that modulate the spatial and/or temporal patterns of collagen gene expression under normal or pathological conditions (section 1.3).

1.2 THE COLLAGEN GENES

The 18 collagen types are encoded by at least 31 distinct genes, which are dispersed throughout the genome (Table 1.1). To date, the collagen genes have been mapped to 11 different human autosomes. The COL4A5 gene is the only known X-linked collagen gene (Hostikka *et. al.*, 1990). At least eight chromosomes contain two or more distinct collagen genes. Five collagen genes have been mapped to chromosome 6, of which three, the COL9A1, COL12A1 and COLYA1 encode FACIT or FACIT-like collagens and may form a cluster on the long arm of this chromosome (section 1.1.3). In support of this hypothesis, two of these genes, the COL9A1 and COLYA1 genes, have both been mapped to 6q12-14 (Kimura *et. al.*, 1989a;

Yoshioka *et. al.*, 1992). The regional localisation of COL12A1 on chromosome 6 has not been determined (Oh *et. al.*, 1992). Four and five genes have been mapped to human chromosomes 1 and 2 respectively. There are only a few other cases where two genes have been mapped to the same locus, namely the COL3A1 and COL5A2 to 2q24.3-31 (Cutting *et. al.*, 1990), COL4A1 and COL4A2 to 13q34 (Soininen *et. al.*, 1988) and COL6A1 and COL6A2 to 21q22.3 (Francomano *et. al.*, 1991). More recently, COL15A1 has been mapped to the same region on chromosome 9 as COL5A1 (Huebner *et. al.*, 1992). Finally, at least three collagen genes, the COL4A3, the COL4A4 and possibly the COL6A3 genes, could form a second cluster on chromosome 2 (Mariyama *et. al.*, 1992a; for review see Vuorio and de Crombrugghe, 1990).

The genes that encode the heterotrimeric collagen types are often situated on different chromosomes. For example, the COL1A1 and COL1A2 are located on the long arms of chromosomes 17 and 7 respectively (Huerre *et. al.*, 1982), while at least two of the genes for the heterotrimeric type IX collagen have been mapped to two different chromosomes (Kimura *et. al.*, 1989a; Perälä *et. al.*, 1993). In some cases such as collagen types IV (Emanuel *et. al.*, 1986; Griffin *et. al.*, 1987) and VI (Francomano *et. al.*, 1991), however, two co-ordinately expressed genes are located on the same locus.

These phenomena pose interesting questions about the co-ordinated regulation of collagen genes (section 1.3). The isolation of genomic and/or cDNA clones from human and other species, has made it possible to determine the structure of the different collagen genes.

1.2.1 STRUCTURE OF THE FIBRILLAR COLLAGEN GENES

The five structurally related fibrillar collagen types all consist of three domains, a long triple helical domain of about 1000 amino acid residues and two short terminal globular domains. Once the procollagen helix has been assembled and secreted, specific extracellular N- and C-proteases remove the globular propeptides, leaving short telopeptides attached to each end of the triple helix. Based on their abundance in collagenous tissues, the fibrillar

collagens can be divided into the major (types I, II and III) and the minor fibrillar collagens (types V and XI).

The heterotrimeric type I, $[\alpha1(I)]_2\alpha2(I)$, and the homotrimeric type III collagens, $[\alpha1(III)]_3$, are widely distributed in many tissues and are the major collagenous constituents of fibres and fibrils in the extracellular matrix. Type II collagen, a homotrimer consisting of three α1(II) chains, is the major collagenous component of hyaline cartilage (reviewed by van der Rest and Garrone, 1991). Type V collagen is widely distributed in many tissues where it is a minor component of the extracellular matrix (reviewed by Fessler and Fessler, 1987). Type V collagen can either occur as a homotrimer, $[\alpha1(V)]_3$, or as two heterotrimeric species $[\alpha1(V)]_2\alpha2(V)$ and $\alpha1(V)\alpha2(V)\alpha3(V)$ (Woodbury *et. al.*, 1989 and references therein). Although the genetic identity of a fourth type V chain, α1'(V), remains unresolved, there is some evidence for the existence of an α1'(V) chain (Fessler *et. al.*, 1985; Fessler and Fessler, 1987). Type XI collagen is a minor collagenous component of hyaline cartilage and is composed of three distinct α-chains (Morris and Bächinger, 1987). Since types I and V collagens are co-expressed in non-cartilagenous tissues and types II and XI are co-expressed in cartilage, it has been postulated that types V and XI collagens may play a role in regulating the diameter of collagen fibrils (Birk *et. al.*, 1988).

Nine distinct but structurally related genes encode the ten α-chains making up the fibrillar procollagen types. Except for the α3(XI) and the α1(II) chains which are probably both produced from the COL2A1 gene, the remaining seven α-chains are synthesised from separate genes. The four major human fibrillar collagen genes, α1(I), α2(I), α1(II) and α1(III), have been isolated and characterised in detail and shown to be located as single copies on different chromosomes (Table 1.1) (for review see Vuorio and de Crombrugghe, 1990). The 18 Kb human α1(I) procollagen gene was the first human fibrillar collagen gene to be isolated in its entirety (Chu *et. al.*, 1984; Westerhausen *et. al.*, 1991) and is located on chromosome 17q21.3-22 (Huerre *et. al.*, 1982; Retief *et. al.*, 1985). Subsequently, the entire human α2(I) (de Wet *et. al.*, 1987), α1(II) (Sangiorgi *et. al.*, 1985) and α1(III) (Chu

et. al., 1985; Benson-Chanda *et. al.*, 1989) genes have been isolated and characterised. The COL1A2 gene is located on chromosome 7q21.3-22 (Junien *et. al.*, 1982; Retief *et. al.*, 1985), while the COL2A1 and COL3A1 genes have been mapped to 12q14.3 (Law *et. al.*, 1986) and 2q24.3-31 (Cutting *et. al.*, 1990) respectively. The genomic and/or cDNA clones for these genes have also been isolated from a variety of species. Genomic and/or cDNA clones of most of the human minor fibrillar collagen genes have been characterised (Myers *et. al.*, 1985; Weil *et. al.*, 1987; Bernard *et. al.*, 1988; Kimura *et. al.*, 1989b; Woodbury *et. al.*, 1989; Greenspan *et. al.*, 1991a; Takahara *et. al.*, 1991). The chromosomal localisation of these genes have also been mapped (reviewed in Vuorio and de Crombrugghe, 1990; Caridi *et. al.*, 1992).

All the vertebrate fibrillar collagen genes isolated to date contain 51 to 54 short exons with gene sizes ranging from 16 to 44 Kb. Despite some important differences, comparison between the different gene types and between the different species have shown that there is considerable structural conservation. Not only the number, but the size and distribution of the exons are almost identical in the fibrillar collagen genes. The exons are separated by large introns ranging in length from about 100 to 3000 bp. Variations in the sizes of the introns account for the large differences in the sizes of these genes (reviewed by Ramirez *et. al.*, 1990; Vuorio and de Crombrugghe, 1990).

In all these genes, 42 short exons of varying sizes encode the triple helical domain. The sizes of exons 7 to 48 in the fibrillar genes are all multiples of 9 bp where 23 exons are 54 bp, 8 exons are 108 bp, 1 exon is 162 bp and 5 each are 45 and 99 bp long. The 9 bp subunit codes for the basic Gly-X-Y triplet, so that each exon always starts with a complete Gly codon and ends with a codon for an amino acid in the Y position. With the exception of three cases, the pattern of exon sizes of all the fibrillar genes are the same (reviewed by Ramirez *et. al.*, 1990; Vuorio and de Crombrugghe, 1990).

The globular C-propeptide (243 to 247 amino acids), the carboxyl telopeptide (11 to 27 amino acids) and the end of

the triple helical domain are coded for by 4 exons (exons 49 to 52). Exons 51 and 52 of all the fibrillar genes are identical in size, both between the different types and different species. There is also a high degree of sequence identity between exons 51 (Yamada *et. al.*, 1983). Small variations in the sizes of exons 48 and 49 are responsible for the variations in the lengths of the C-propeptide domains. Exon 49, the C-joining exon, encodes for the C-terminal end of the triple helix (ranging from 45 to 63 bp), the C-telopeptide and beginning of the C-propeptide. The joining exon of the α2(XI) procollagen gene contains only 18 bp of triple helical sequences (two Gly-X-Y repeats). The sizes of exons 48 and 49 in the COL11A2 gene are also much smaller than the equivalent exons of the other fibrillar collagen genes (Kimura *et. al.*, 1989b). The slight variations in the lengths of the various triple-helical domains (1014 to 1029 amino acids) is caused by differences in the number of Gly-X-Y repeats in the C- and N-joining exons (reviewed by Ramirez *et. al.*, 1990; Vuorio and de Crombrugghe, 1990).

The size, number and organisation of the various exons encoding the N-propeptide domain of the α-chains differ substantially between the genes and between species. The N-propeptide domains can be subdivided into at least four regions, namely the signal peptide, cys-rich globular region, the short triple-helical region and the N-telopeptide. The signal peptides are coded for by exon 1, while exon 2 in the α1(I), α1(II) and α1(III) procollagen genes code for the 67 to 71 amino acid cys-rich globular subdomains. Since this subdomain is only found in a fraction of the α1(II) procollagen mRNA, it is probably removed by alternative splicing. Two short exons (2 and 3) of 11 and 15 bp, respectively, code for a globular subdomain in the α2(I) chain. The exons that code for the short triple helical subdomain of 39 to 79 amino acids, differ in size and number. This region is coded for by two exons within the α2(I) (exons 4 and 5) and α1(III) (exons 3 and 4/5) gene. Exons (3, 4 and 5) and exons (3, 4A, 4B, 5A and 5B) code for this region in the α1(I) and α1(II) procollagen genes respectively. Exon 6 is the N-joining exon and codes for the N-telopeptide and the

beginning of the triple helix (reviewed by Ramirez *et. al.*, 1990; Vuorio and de Crombrugghe, 1990).

The fibrillar collagen genes also contain several polyadenylation sites which gives rise to different mRNA species (reviewed by Vuorio and de Crombrugghe, 1990).

1.2.2 SHORT CHAIN COLLAGEN GENES

The two short chain collagens, types VIII and X, are encoded by three genetically distinct genes, namely, the COL8A1, COL8A2 and COL10A1. This subclass has been named short chain collagens due to the relatively small size of the triple helical domain which is about half as long as in fibrillar collagens (Yamaguchi *et. al.*, 1991). The short chain collagens contain a short triple helical domain of about 450 to 460 amino acids flanked by N- and C-globular domains consisting of 40 to 120 and about 170 residues respectively. Although it is conceivable that each of the type VIII collagen chains may form homotrimers, it is probably a heterotrimer with a chain composition $[\alpha1(VIII)]_2\alpha2(VIII)$ (Mann *et. al.*, 1990). This collagen is produced by vascular and corneal endothelial cells where it is a component of a specialised basement membrane known as the Descemet's membrane. There is also a much wider tissue distribution of type VIII collagen (Yamaguchi *et. al.*, 1989, 1991 and references therein). The transiently and developmentally regulated type X collagen, on the other hand, is produced by chondrocytes in hypertrophic hyaline cartilage and is a homotrimer made up three $\alpha1(X)$ chains (Schmid and Linsenmayer, 1983).

The entire human COL10A1 gene has been cloned and shown to contain only two exons of 169 bp and approximately 2940 bp, separated by an intron of about 3200 bp. Exon 1 encodes the 5'-untranslated region (15 bp), the signal peptide (54 bp) and 100 bp of the N-terminal globular domain. The second exon encodes for the rest of the N-globular domain (14 bp), the entire triple helical domain of 463 amino acids, the 161 amino acid long C-terminal globular domain and the 3'-untranslated region (1054 bp) (Thomas *et. al.*, 1991; Reichenberger *et. al.*, 1992). The cloning and

characterisation of the entire chicken (Lu valle *et. al.*, 1988) and mouse (Apte *et. al.*, 1992; Apte and Olsen, 1993) α1(X) collagen genes have shown that they, on the other hand, contain three exons. Exon 3 is about 2100 bp and is the largest exon in both of these genes. Like the human α1(X) gene, the human α1(VIII) gene contains two exons. Exon 1 is 331 bp long and the larger exon 2 is about 1907 bp in size (Muragaki *et. al.*, 1991b). Interestingly, the 5.3 Kb rabbit gene consists of four exons of 69, 120, 331 and 2278 bp (Yamagachi *et. al.*, 1991). Although the number and sizes of the exons in the human α2(VIII) gene is not known, the gene has been cloned and shown to be structurally related to the other short chain collagen genes (Muragaki *et. al.*, 1991a).

The three short chain genes have a number of unique characteristics (Muragaki *et. al.*, 1991a, 1991b; Thomas *et. al.*, 1991). The entire triple-helical and C-terminal globular domains are encoded by a single large exon (Ninomiya *et. al.*, 1986; Yamaguchi *et. al.*, 1989, 1991; Thomas *et. al.*, 1991). This is in direct contrast to the fibrillar collagen genes which contain multiple short exons. The rest of the molecule is encoded by one or more short exons. Secondly, the carboxyl three-quarters of the C-terminal globular domains have a very high degree of sequence conservation. Like most of the non-fibrillar collagen genes, the triple helical domain of the short chain collagen genes contain short imperfections in the Gly-X-Y tripeptides. The short chain collagen genes have the same number (usually eight) of Gly-X and/or X-Y imperfections. The relative locations of these imperfections are also constant. The COL8A1, COL8A2 and COL10A1 genes have been mapped to chromosomes 3q12-13.1 (Muragaki *et. al.*, 1991b), 1p32.3-34.3 (Muragaki *et. al.*, 1991a) and 6q21-22 (Apte *et. al.*, 1991; Thomas *et. al.*, 1991) respectively.

1.2.3 FIBRIL-ASSOCIATED COLLAGENS WITH INTERRUPTED TRIPLE-HELICES (FACIT) AND FACIT-LIKE GENES

Type IX collagen, a heterotrimer of genetically distinct α1(IX), α2(IX) and α3(IX) chains, was the first member of the FACIT subgroup to be identified (Noro *et. al.*, 1983). Subsequently, the structurally and functionally related

homotrimeric types XII (Gordon *et. al.*, 1987) and XIV (Dublet and van der Rest, 1991) collagens have also been classified as FACITS. Types XVI (Pan *et. al.*, 1992) and Y (Yoshioka *et. al.*, 1992) collagens, discovered by cDNA cloning and sequencing, are also probably members of the FACIT group . The fibril-associated collagens with interrupted triple-helices have a number of common unique characteristics. Firstly, they do not undergo proteolytic processing like the fibrillar collagens and are therefore not secreted as procollagen molecules. Secondly, they do not form fibrils, instead they are associated with collagenous fibrils or fibres. In hyaline cartilage, for example, type IX collagen associates with and becomes covalently crosslinked to the surface of type-II-containing collagenous fibrils (Eyre *et. al.*, 1987; van der Rest *et. al.*, 1988; Vaughan *et. al.*, 1988).

The FACIT molecules contain more than one triple helical domain separated by short non-triple helical domains. Each of the three type IX collagen chains contains three triple helical domains, designated COL1, containing either 115 (α1 and α2) or 112 (α3) amino acids, COL2 and COL3, consist of 339 and 137 amino acids respectively, and four non-triple helical domains (NC1 to NC4) (Ninomiya *et. al.*, 1990; Har-El *et. al.*, 1992). Type XII collagen, on the other hand, contains only two helical domains of 105 (COL1) and 152 (COL2) amino acids and three non-helical domains (NC1 to NC3) (Gordon *et. al.*, 1989). The structure and function of the FACITS have been extensively reviewed by Shaw and Olsen (1991) and van der Rest and Garrone (1991).

The entire chicken α1(IX) and α2(IX) collagen genes have been isolated and characterised (Lozano *et. al.*, 1985; van der Rest and Mayne, 1987). The α2(IX) gene is 10 Kb long and contains 32 exons, while the α1(IX) gene is much larger, approximately 100 Kb (reviewed by Vuorio and de Crombrugghe, 1990). Since the α2(IX) chain and gene has been the most extensively studied, it serves as a prototype molecule for the FACITS. More recently, the full length cDNA for the chicken α3(IX) chain has been cloned and characterised (Brewton *et. al.*, 1992; Har-El *et. al.* ,1993). The human full length α1(IX) (Kimura *et. al.*, 1989a) and α2(IX) (Perälä

et. al., 1993) cDNA clones have also been isolated and used to map these genes to 6q12-14 (Kimura *et. al.*, 1989a) and chromosome 1 (Perälä *et. al.*, 1993), respectively.

A full length chicken α1(XII) cDNA has been isolated and characterised while only a portion of a genomic clone corresponding to exons 1 to 7, has been sequenced (Gordon *et. al.*, 1989). A portion of the mouse α1(XII) gene has been cloned (Oh *et. al.*, 1992) and used to isolate a human α1(XII)-like gene which contains 3 exons that encode the carboxyl-end of the NC3 domain and the NC3/COL2 junction. The human gene has been mapped to chromosome 6 (Oh *et. al.*, 1992). A partial cDNA bovine type XIV clone has been shown to contain the COL1, NC2 and part of the COL2 domains (Dublet and van der Rest, 1991).

Comparisons between the various FACIT genes have shown that they are very heterogeneous in size and that considerable sequence variations also exist. Never the less the FACIT genes do have a number of common characteristics. The exon/intron organisation of the α1(IX) and α2(IX) genes are similar (Lozano *et. al.*, 1985; van der Rest, 1987). There is also homology in the exon/intron organisation of the partial chicken α1(XII) gene, encoding the NC1, COL1 and NC2 domains, with the corresponding regions of the type IX genes (Gordon *et. al.*, 1987). Most of the exons, encoding the triple helical domains, in these genes adhere to the 9 bp rule. As with the fibrillar collagen genes, a significant portion of these exons are 54 bp long. The α1(IX) and α2(IX) collagen genes contain imperfections in their Gly-X-Y tripeptide repeats, where either a X, Y or Gly residue is deleted (Ninomiya *et. al.*, 1989).

A full length human cDNA clone of the putative α1(XVI) chain, contains 10 collagenous domains alternating with 11 non-collagenous domains, has been isolated and characterised recently (Pan *et. al.*, 1992). This gene is located on chromosome 1p34-35 (Pan *et. al.*, 1992). A homologue of the α1(XII) collagen gene, COLYA1 (D6S228E) has been isolated and mapped to chromosome 6q12-14 (Yoshioka *et. al.*, 1992). In addition to this gene, four collagen genes, COL9A1, COL10A1, COL11A2 and COL12A1, have all been mapped to chromosome 6.

The chicken (Nishimura *et. al.*, 1989) and human (Muragaki *et. al.*, 1990) α1(IX) genes contain two promoters which are preferentially utilised in different tissues to give rise to proteins with different N-terminal (NC4) domains (section 1.3.6.7). The chicken α1(IX) gene also contains a number of polyadenylation sites (Lozano *et. al.*, 1985; Ninomiya *et. al.*, 1989; Nishimura *et. al.*, 1989).

1.2.4 EXTRACELLULAR MATRIX NETWORKS

Although several isoforms of type IV collagen exist in the basement membranes, the most common one has the composition [α1(IV)$_2$α2(IV)]. Both the α1(IV) and α2(IV) chains are present in basement membranes. The minor trimeric isoforms consist of combinations of the α1(IV) chain, the α2(IV) chain and three other chains, designated α3(IV), α4(IV) and α5(IV) chains, but with restricted tissue distribution. The type IV molecule consists of a large C-terminal globular domain (NC1), a long triple helical domain of about 1400 amino acid residues and a short cysteine-rich non-collagenous N-terminal domain (NC2) (reviewed by Schittny and Yurcherico, 1989).

The loci of the human α1(IV) and α2(IV) genes have both been mapped to 13q34 (Emanuel *et. al.*, 1986; Griffin *et. al.*, 1987). In contrast to the other collagen genes and other genes of complex organisms, the human α1(IV) and α2(IV) genes occur in a head-to-head orientation so that they share a common bidirectional promoter (section 1.3.6.4) (Soininen *et. al.*, 1988). The COL4A5 gene, to date the only collagen gene known to be situated on a sex chromosome, has been mapped to Xq21 (Hostikka *et. al.*, 1990). The COL4A3 and COL4A4 genes have both been mapped to chromosome 2q35-37 (Mariyama *et. al.*, 1992a; Turner *et. al.*, 1992).

The complete primary structures of the human α2(IV) (Hostikka and Tryggvason, 1988) and α5(IV) (Pihlajaniemi *et. al.*, 1990; Zhou *et. al.*, 1992) genes have been determined from nucleotide sequence analysis. Only partial cDNA clones of the human α3(IV) (Morrison *et. al.*, 1991; Turner *et. al.*, 1992) and the bovine α4(IV) (Mariyama *et. al.*, 1992b) chains have been isolated and sequenced to date. The entire human

α1(IV) gene has been characterised and shown to contain 52 exons and to be more than 100 kb in length (Soininen *et. al.*, 1989). Based on their structure, the five α-chains can be divided into two classes, namely, the alpha-1-like [α1(IV), α3(IV) and α5(IV)] and alpha-2-like [α2(IV) and α4(IV)] classes (Mariyama *et. al.*, 1992a). The α1(IV) and α2(IV) chains contain a considerable number of interruptions in the triple helical domains. At least half of the interruptions are larger than a deletion or insertion of a single amino acid. Some of the insertions can be up to 24 amino acids long. Most of the exons encoding the triple helical domain contain non-random split codons which always involves the first G of the gly codon. The 9 bp rule is usually adhered to in these exons but the exon sizes are different to those found in the fibrillar collagen genes (Soininen *et. al.*, 1989). There are also extensive differences in the exon organisation of the 3'-end of the α1(IV) and α2(IV) collagen genes (Hostikka and Tryggvason, 1987).

Type VI collagen is a component of the microfibrillar network in the extracellular matrix of a wide variety of tissues. The protein, a heterotrimer consisting of three distinct α-chains [α1(VI), α2(VI) and α3(VI)], forms a dumbell shaped molecule consisting of two large globular domains separated by a short triple helical region. The non-collagenous domains comprise about 80% of the total protein (reviewed by Engel *et. al.*, 1990).

The corresponding cDNAs for all three chains of the human type VI collagen has been cloned and sequenced (Chu *et. al.*, 1988, 1989, 1990). The α1(VI) and α2(VI) chains are similar in structure where the large 230 amino acid N-terminal and the 430 amino acid C-terminal globular domains are separated by a short triple helix of 335 and 336 amino acids respectively. The N- and C-globular domains are divided into subdomains of about 200 amino acid residues. The subdomains are separated from each other or from the collagenous domain by short cysteine-rich connecting segments of 25 to 30 residues long. The entire 36 Kb human α2(VI) collagen gene consists of 30 exons (Saitta *et. al.*, 1990, 1991a, 1992) with exons 1A and 2 separated by the largest intron of 12.5 Kb (Saitta *et. al.*, 1992). Six exons encode the signal peptide

and N-globular domain. Exon 5 is a N-terminal junction exon which encodes for the transition of the non-collagenous to the collagenous domains (Saitta *et. al.*, 1992). The C-terminal globular domain is also encoded by 6 exons and is subject to alternate splicing. In contrast with most of the other collagen genes there is no junction exon in the α1(VI) and α2(VI) genes (Saitta *et. al.*, 1990). All the individual structural domains of the N- and C-terminal globular domains are encoded by separate exons (Saitta *et. al.*, 1990, 1992) The triple helical domain which contains two short interruptions is encoded by the remaining 18 exons (Saitta *et. al.*, 1991). The organisation of the exons that encode the triple helical domains of the human α1(VI) and α2(VI) chains are very similar. With the exception of those encoding regions with imperfections in the triple helix, the exons in both genes adhere to the 9 bp rule and have sizes ranging from 27 to 90 bp. Contrary to the fibrillar collagen genes, the predominant exon size is 63 bp. Contrary to many of the non-fibrillar collagen genes, the introns are positioned between complete codons so that the α1(VI) and α2(VI) genes do not contain split codons in their triple helical domains (Saitta *et. al.*, 1991).

The N-terminal domain of the larger α3(VI) chain contains nine 200 amino acid subdomains while the C-terminal domain contains two 200 amino acid subdomains and three unrelated subdomains (Chu *et. al.*, 1990; Stokes *et. al.*, 1990). A partial 26 Kb genomic clone of the entire N-terminal globular domain of the human α3(VI) chain showed that the N1 subdomain is encoded by two exons of 417 bp and 146 bp, while subdomains N2 to N9 are encoded by separate exons of about 600 bp each (Stokes *et. al.*, 1990). Exons encoding subdomains N7 and N9 are subject to tissue-specific alternate splicing (Stokes *et. al.*, 1990). Single copies of the COL6A1 and COL6A2 genes exist as a cluster on chromosome 21q22.3 (Francomano *et. al.*, 1991), while the COL6A3 gene has been mapped to 2q37 (see Vuorio and de Crombrugghe, 1990).

The homotrimeric type VII collagen, $[\alpha1(VII)]_3$, consists of a central collagenous domain consisting of 52 Gly-X-Y repeats, a C-terminal non-collagenous domain (NC1) and a N-terminal non-collagenous domain (NC2). The NC1 and NC2 domains

consist of 186 and 439 amino acids respectively. This collagen is found in the basement membrane zone beneath the stratified epithelia where it forms part of the anchoring fibrils (Burgeson, 1988; van der Rest and Garrone, 1991). A human cDNA clone corresponding to the α1(VII) chain has been successfully isolated and characterised (Parente *et. al.*, 1991). A 8 kb genomic clone, consisting of 34 exons of the carboxyl end of the molecule, has recently been characterised (Greenspan *et. al.*, 1993). Consistent with other non-fibrillar collagens, the triple helical domain of type VII collagen contains a number of imperfections where 1 to 3 amino acids are inserted or deleted. The COL7A1 gene has been mapped to chromosome 3p21.1-21.3 (Parente *et. al.*, 1991; Greenspan *et. al.*, 1993).9

1.3 REGULATION OF COLLAGEN GENE EXPRESSION

The mechanisms that regulate the production of the various collagen types are complex and diverse. Although transcriptional control mechanisms are probably the most important and predominant, other mechanisms also play significant roles. These other mechanisms, including chromatin structure, DNA methylation status, mRNA processing, mRNA stability, translational control and post-translational processes, will be discussed briefly.

1.3.1 CHROMATIN CONFORMATION

Eukaryotic DNA is packaged into the nucleus in a highly organised manner (reviewed by Fisher, 1989) and unfolding of this structure must proceed transcription. Differences must therefore exist between the structure of transcriptionally active, potentially active and inactive genes. Transcriptionally and potentially active genes have a more open conformation which is easily accessible to the transcription apparatus. Due to its more open conformation, active chromatin is also more sensitive to nucleases such as DNase 1 and S1 nuclease and is a measure of the activation status of genes (reviewed by Weisbrod, 1982; Gross and Garrard, 1987, 1988; Patient and Allan, 1989). DNase 1 hypersensitive sites have been identified in the type I collagen genes and shown to bear a strong correlation with

their expression (reviewed by Bornstein and Sage, 1989; Raghow and Thompson, 1989). This technique has been used to identify discrete regions in the chromatin around the mouse α2(I) procollagen promoter (Liau *et. al.*, 1986). Mc Keon *et. al.* (1984) have found a DNase 1 hypersensitive site in the chicken α2(I) procollagen promoter region in fibroblast but not in brain chromatin, which corresponds to the transcriptional activity of this gene in these tissues.

1.3.2 METHYLATION STATUS

There is an inverse correlation between the methylation status of a gene and its transcriptional activity. Transcriptionally active genes are generally under-methylated at certain CpG dinucleotides within their regulatory elements (reviewed by Doerfler, 1983; Cedar, 1988; Boyes and Bird, 1991). With some exceptions (Mc Keon *et. al.*, 1982), hypermethylation of the type I collagen genes is associated with their inactivation (Parker *et. al.*, 1982, 1986; Smith and Marsilio, 1988; Guenette *et. al.*, 1992). *In vitro* methylation of the human α1(I) procollagen promoter and enhancer also results in transcriptional inactivation of this gene (Thompson *et. al.*, 1991). Although a controversial issue, DNA methylation may be an important mechanism for regulating collagen gene expression.

1.3.3 POST-TRANSCRITIONAL CONTROL MECHANISMS

In comparison with the initiation of transcription (section 1.3.6) and post-transcriptional control mechanisms (reviewed by Prockop *et. al.*, 1979; Bornstein and Sage, 1989), little is known about the regulation of collagen mRNA stability and turnover. These events may include splicing, capping, polyadenylation, sequence editing, nuclear-cytoplasmic transport, stability and turnover (for reviews on mRNA stability and turnover see Cleveland, 1989 and Klausner and Harford, 1989). As illustrated by the following examples, collagen mRNA stability and turnover is regulated at a number of levels. In differentiated chicken chondrocytes, for example, unprocessed α1(I) mRNA is found predominantly in the nucleus and is not transported to the cytoplasm (Saxe *et. al.*, 1985). Dozin *et. al.* (1990) have shown that an

increase in the half-life of α1(II) procollagen mRNA causes accumulation of the message in chicken chondocytes. Since several transcripts, which differ only in their 3'-UTR, are often produced from collagen genes (reviewed by Vuorio and de Crombrugghe, 1990), it is almost certain that the 3'-UTR plays an important role in the stability and turnover of the α1(I) and α2(I) procollagen mRNA. In support of this hypothesis, Maatta *et. al.* (1991) have shown that the 3'-UTR of the COL1A1 gene binds cell-specific nuclear proteins. Finally, Furth *et. al.* (1991) have shown that a post-initiation block prevents the accumulation of type I collagen mRNA in HeLa cells.

1.3.4 REGULATION BY EXOGENOUS FACTORS

Collagen gene expression is also regulated by many exogenous factors such as growth factors, cytokines, hormones and vitamins. These agents can affect collagen mRNA levels by altering either the transcription rates or mRNA stability and turnover. Cytokines such as, interferon γ (IFN-γ), tumour growth factor ß (TGF-ß), tumour necrosis factor α (TNF-α), interleukin 1 and hormones such as glucocorticoids, may modulate collagen mRNA stability and turnover (reviewed by Bornstein and Sage, 1989; Slack *et. al.*, 1993). The effect of exogenous factors on type I collagen gene expression has also been extensively reviewed (Adams, 1989; Bornstein and Sage, 1989; Slack *et. al.*, 1993). Several exogenous factors can regulate collagen synthesis by independent mechanisms and the combined action of these effects play an important role in regulating collagen synthesis during wound healing and tissue repair (Narayanan *et. al.*, 1989). In addition TNF-α and IFN-γ synergistically suppress the stimulatory effect of TGF-ß on type I collagen synthesis by two distinct mechanisms. TNF-α suppresses the activity of the α2(I) procollagen promoter while IFN-γ probably functions at the post-transcriptional level (Kahari *et. al.*, 1990). The synthesis of other collagen types is also affected by exogenous factors. In culture, TGF-ß up-regulates type VII collagen gene expression in normal and transformed epidermal keratinocytes (Ryynanen *et. al.*, 1991) and type VI collagen in human dermal fibroblasts (Heckmann *et. al.*, 1992). Often co-expressed collagen types can be

affected; the cytokine, melanoma growth-stimulatory activity [(MGSA)/GRO], down-regulates the expression of types I and III collagens in a dose-dependent manner in fibroblasts (Unemori *et. al.*, 1993).

1.3.5 TRANSFORMED CELLS

The transformation of fibroblasts by RNA and DNA tumour viruses causes changes in cell morphology, cell growth and synthesis of distinct groups of proteins. The synthesis of type I collagen (reviewed by Bornstein and Sage, 1989) and other collagens, such as type VI (Schreir *et. al.*, 1988), are markedly reduced upon transformation. Secondly, one of the two type I collagen genes is often transcriptionally inhibited when epithelial cells and fibroblasts are chemically transformed (reviewed by Bornstein and Sage, 1989). In addition, the expression of different collagen types, such as types I and III (Liau *et. al.*, 1985), or the expression of the different chain components of a specific collagen type, such as α1(I) and α2(I) chains (Majmudar *et. al.*, 1988), are not always co-ordinately regulated in transformed cells. Both chemically and virally transformed fibroblasts have been used to study the mechanisms that regulate the expression of collagen genes (Parker *et. al.*, 1989; Guenette *et. al.*, 1992).

1.3.6 TRANSCRIPTION FACTORS

Eukaryotic gene transcription is regulated predominantly by the interaction of *cis*-acting DNA elements with sequence-specific DNA-binding proteins. These recognition sequences are located within functionally distinct elements such as promoters, enhancers and silencers (Dynan, 1989). While promoters are situated upstream of the transcription start site, enhancers can function over considerable distances in either orientation to increase promoter activity (reviewed by Khoury and Gruss, 1983; Serfling *et. al.*, 1985; Jeang and Khoury, 1988; Müller *et. al.*, 1989). Silencer elements function in a position and orienatation-independent manner to negatively regulate gene expression. This is achieved by the binding of repressors to their recognition sequences within these elements (Renkawitz, 1990). Silencers were first

identified within a 589 bp region upstream of the porcine PD1 promoter (MHC class I gene) (Ehrlich *et. al.*, 1988). These investigators showed that removal of this PD1 silencer element resulted in increased promoter activity.

One of the most common *cis*-acting DNA elements is the TATA box, which is usually located approximately 30 bp upstream of the transcription start site and is the binding site for RNA polymerase and the rest of the general transcriptional machinery (reviewed by Conaway and Conaway, 1991). Other common recognition elements include the CCAAT and GC boxes while less common *cis*-elements include hormone and growth factor response elements. Numerous lower-abundance, structurally diverse sequence-specific DNA-binding proteins (Sp1, OCT-1, OCT-2, NF-1, Jun, Fos, CBF, etc) bind to their respective recognition elements to either activate or repress gene expression (reviewed by Maniatis *et. al.*, 1987; Mitchell and Tjian, 1989). DNA-binding proteins have a number of functionally distinct domains. The first is a DNA-binding domain, which, in numerous *trans*-acting factors, has been localised to a relatively small region of the protein. Based on their structure, these DNA-binding motifs can be divided into several classes such as helix-turn-helix, zinc-finger, basic region of the leucine-zipper and others (reviewed by Struhl, 1989; Johnson and Mc Knight, 1989; Harrison, 1991). Many *trans*-acting factors bind to recognition elements as multimers and therefore also contain a dimerisation domain such as the leucine zipper (Landschulz, *et. al.*, 1988). All transcription factors also contain a functionally distinct transcriptional activator, or repressor, domain which could be an acidic, a glutamine-rich or a proline-rich domain (Tasset *et. al.*, 1990).

A combination of both positive and negative *trans*-acting factors are frequently responsible and vital for regulating gene expression (Boam *et. al.*, 1990; Chow and Schwartz, 1990). It is generally accepted that transcription activators transmit a positive signal to the transcription initiation complex when they bind to *cis*-elements within promoters and enhancers of a gene. Although, many of the components and mechanisms involved are not known, it is more than likely that intermediate factors, which themselves do

not bind DNA, act as a bridge between the positive factor and the initiation complex. It is also possible that the activator interacts directly with the initiation complex to stimulate transcription (reviewed by Ptashne, 1988; Müller *et. al.*, 1989; Ptashne and Gann, 1990). Transcription repressors on the other hand, can function via one of at least three broad mechanisms which include the inhibition of activator binding, blocking of activation or direct DNA-binding to achieve silencing (reviewed by Levine and Manley, 1989; Renkawitz, 1990). There are many different mechanism whereby repressors prevent the binding of positive factors to their recognition element (reviewed in Renkawitz, 1990). A classical example is the competition of a CCAAT displacement protein (a repressor) with the binding of a CCAAT binding factor (an activator) to overlapping *cis*-elements within the sea-urchin H2B-1 promoter (Barberis *et. al.*, 1987). Interestingly, there are examples where transcriptional repression is caused by competition between two positive factors (reviewed in Renkawitz, 1990). Alternatively, a repressor can prevent an activator from binding to its recognition element by forming stable protein-protein complexes. Inhibitory POU (I-POU), for example, does not bind to DNA, but forms a stable heterodimer with Cf1-a (a POU-domain protein) to inhibit DNA-binding and transcriptional activation ability of Cf1-a (Treacy *et. al.*, 1991, 1992). An inhibitory protein, IkB, prevents the binding of a transcriptional activator, NF-kB, to its *cis*-element via a different mechanism. IkB binds to NF-kB in the cytoplasm and prevents NF-kB from entering the nucleus. After activation, IkB releases NF-kB into the nucleus where it activates several genes (Zabel and Baeuerle, 1990). Other repressors inhibit the transcriptional activation properties of DNA-binding proteins via several mechanisms (reviewed by Renkawitz, 1990). Finally, activators and repressors can be produced from the same gene (reviewed by Foulkes and Sassone-Corsi, 1992) or due to post-translational modification of transcription factors. A twin of I-POU, for example, is produced by alternative splicing and forms a homodimer which activates a distinct group of genes (Treacy *et. al.*, 1992). The regulation of transcription by phosphorylation of transcription factors has been reviewed by Hunter and Karin (1992).

Transient transfection experiments using sequences upstream of the start site of initiation and/or intronic segments fused to reporter genes have been used to define the functional promoter, enhancer and silencer elements of a number of collagen genes. Deletion constructs, together with other DNA-binding techniques, have identified *cis*-acting elements within the promoters, enhancers and silencers of collagen genes. A number of investigators have identified, purified and characterised the *trans*-acting factors which bind to these regulatory elements. The micro-injection of chimeric genes into mouse oocytes to generate transgenic mice has and will continue to play an important role in understanding the tissue-specific, temporal and spatial expression of genes during embryogenesis and in the adult organism. Hopefully, the identification and characterisation of regulatory elements and their DNA-binding proteins, will help to determine the mechanisms governing the co-ordinated expression of the various collagen genes under normal and pathological conditions.

1.3.6.1 TYPE I COLLAGEN GENE

The transcriptional mechanisms which regulate the cell-type and tissue-specific expression of the type I collagen genes have been well studied. The mouse COL1A2 gene is the best studied gene. Transient transfection experiments (Schimdt *et. al.*, 1986) and transgenic mice (Khillan *et. al.*, 1986) were initially used to show that the mouse α2(I) promoter (between -2000 and +54) was sufficient in directing the expression of a reporter gene in a cell- or tissue-specific manner. Although considerably weaker, the expression of the COL1A2-driven reporter gene mimics that of the endogenous gene. It is therefore possible that additional enhancer elements, located either upstream or downstream of this promoter fragment, are required for high-level expression of the mouse α2(I) procollagen gene (Schmidt *et. al.*, 1986; Goldberg, H. *et. al.*, 1992). This 2000 bp promoter fragment, never the less, is sufficient for its tissue-specific expression. The minimal DNA-sequences required for tissue-specific expression of the mouse α2(I) procollagen gene have been shown to be located between -350 and +54 (Niederreither

et. al., 1992). Although the overall activity was less than that of the -2000 bp promoter, the -350 bp proximal promoter was able to mimic the temporal and spatial expression of the endogenous gene during embryogenesis in transgenic mice. Additional upstream sequences seem to be required for the tissue-specific repression of the α2(I) procollagen gene in the brain since expression of this construct was not repressed in this tissue (Goldberg, H. *et. al.*, 1992). Boast *et. al.* (1990) have shown in tissue culture experiments that only the proximal 350 bp of the human α2(I) procollagen promoter is required for the cell-type specific expression of a reporter gene. Interestingly, there is a high degree of sequence identity (86%) between the human and mouse -350 bp proximal promoters (Dickson *et. al.*, 1985). This high sequence identity between these promoters is lost further upstream (M.I. Parker, unpublished data).

Several *cis*-acting DNA-elements such as a TATA box (-30 to -25), an inverted CCAAT box (-84 to -80), a $^{5'}$CAGA$^{3'}$ sequence (-250 to -247) and a NF-1-like site (-315 to -295) have been identified in the mouse proximal α2(I) procollagen promoter (Karsenty *et. al.*, 1988). Interestingly, all these elements are conserved in the human promoter (see chapter 2). DNA-binding factors present in NIH-3T3 nuclear extracts have been shown to bind to these regulatory elements. Although the factor(s) which bind to the $^{5'}$CAGA$^{3'}$ sequence has not been identified, the substitution of this motif with other nucleotides (AAAG or ATAG) almost abolishes COL1A2 promoter-driven expression of a reporter gene and DNA-binding activity (Karsenty *et. al.*, 1988). A heterotrimeric *trans*-acting factor, the activating CCAAT binding factor (CBF), which binds to the inverted CCAAT box, was initially indemnified in NIH-3T3 extracts (Hatamochi *et. al.*, 1988; Maity *et. al.*, 1988, 1992). Subsequently this factor has been purified from rat liver and characterised in some detail (section 1.4.3). A direct correlation exists between the binding of a highly purified preparation of CBF to the CCAAT box and its ability to stimulate transcription in an *in vitro* reconstituted transcription assay (Maity *et. al.*, 1988). CTF/NF-1 or a related protein binds to the NF-1-like site at -300 and has been shown to mediate the stimulatory effect of TGF-ß on type I collagen gene expression (Oikarinen *et. al.*, 1987;

Rossi *et. al.*, 1988). A 3 bp substitution mutation within the NF-1-like binding site also prevents the induction of the mouse α2(I) procollagen promoter by TGF-ß (Rossi *et. al.*, 1988). However, there appears to be no difference in the formation of the of NF-1-like complex between either TGF-ß treated or non-treated cells (de Crombrugghe *et. al.*, 1990). In electrophoretic mobility shift assays, the levels of complexes that binds to the NF-1-like sequence are higher in fibroblast than in myeloma nuclear extracts (Goldberg, H. *et. al.*, 1992). These results suggest that this element probably plays an important role in the cell-specific expression of the mouse α2(I) procollagen gene in fibroblasts. Mutations in both the CCAAT and NF-1-like motifs decreases DNA-binding of the cognate proteins and also decreases promoter activity in transfection assays (Karsenty *et. al.*, 1988). Although NF-1 can bind to sequences that contain a CCAAT motif, CTF/NF-1 and CBF only bind to their respective recognition elements, namely the NF-1-like site or the inverted CCAAT box, within the mouse α2(I) procollagen promoter (Oikarienen *et. al.*, 1987). Ristiniemi and Oikarinen (1989) have purified a factor, which may be histone H1, that binds to the NF-1-like sequence in the mouse α2(I) procollagen promoter.

A segment of the mouse COL1A2 first intron, between +418 and +1524, functions in an orientation independent manner as a transcriptional enhancer in transient transfection experiments in NIH-3T3 cells but not in the S194 lymphoid cell line (Rossi and de Crombrugghe, 1987). Since this intronic enhancer is also active in a number of other cell types, including human cervical carcinoma cells (HeLa) and African Green Monkey cells (CV-1) (Pogulis and Freytag, 1993), it is possible that it may only function in cell types that synthesise type I collagen. Goldberg, H. *et. al.* (1992), on the other hand, have shown that this intronic fragment is not essential for the tissue-specific expression of a COL1A2-driven reporter gene in transgenic mice. Recently, two consensus sequences have been identified within the intronic enhancer (Pogulis and Freytag, 1993). The first element contains a $^{5'}$TGTTTTAA$^{3'}$ motif (initially identified within the c-mos and human papilloma virus enhancers) between nucleotides +992 and +999. A collagen intron-binding factor

I (CIBF-I) which is present in mouse fibroblast, HeLa and CV-1 nuclear extracts, binds to this element. The second *cis*-acting element is a "GT box" (nucleotides +871 to +880) that binds to affinity-purified Sp1 and to a Sp1-like protein present in NIH-3T3 and other nuclear extracts (Pogulis and Freytag, 1993). The binding of both CIBF-I and Sp1 to their respective intronic recognition elements contribute to the stimulation of the mouse α2(I) procollagen gene.

In contrast to the mouse first intron, the human first intron does not contain any enhancer activity: On the contrary, it causes a decrease in promoter activity (Sherwood *et. al.*, 1990). Comparative studies of the human and mouse α2(I) procollagen first introns identified very little sequence homology (Sherwood *et. al.*, 1990). The human first intron contains two potential AP-1 consensus sequences and a long GT stretch, which are both absent in the mouse.

Transient transfection experiments have suggested that the mouse proximal α1(I) procollagen promoter (220 bp) contains sufficient information for the tissue-specific expression of this gene (Rippe *et. al.*, 1989). The human α1(I) proximal promoter (from -300 to +100) is also able to effectively direct transcription of a reporter gene in a tissue-specifc manner (Bornstein *et. al.*, 1987; Rossouw *et. al.*, 1987). The human, chicken and mouse proximal promoters contain an intact TATA box and additional upstream promoter sequences (Finer *et. al.*, 1987; Dickson *et. al.*, 1987; Brenner *et. al.*, 1989). The human minimal promoter comprises the CCAAT box, potential Sp1 binding sites and a pyrimidine-rich element (Rossouw *et. al.*, 1987). The chicken promoter contains CCAAT boxes, but lacks the pyrimidine-rich element (Finer *et. al.*, 1987). Two inverted CCAAT boxes, one between nucleotides -96 and -100 and the other between nucleotides -122 to -126, are also located within the mouse α1(I) proximal promoter (Brenner *et. al.*, 1989; Karsenty and de Crombrugghe, 1990). The transcriptional activator, CBF, of the mouse α2(I) procollagen gene, binds to the downstream, but not the upstream, CCAAT box to activate transcription of the α1(I) gene (Karsenty and de Crombrugghe, 1990). It is possible that CBF plays a vital role in the co-ordinate regulation of the type I collagen genes. The downstream CCAAT box is

flanked by a 12 bp GC-rich direct repeat, $^{5'}$TGGGGGCCGGGC$^{3'}$, extending from nucleotides -124 to -113 and -94 to -83. The 12 bp repeats resemble the binding sites for Sp1, MLTP (a *trans*-activating factor in the adenovirus major late and γ-fibrinogen promoters) or AP-2 (Nehls *et. al.*, 1991). Karsenty and de Crombrugghe (1990), however, have shown that purified Sp1 does not bind to this site or that an AP-2 consensus sequence did not compete with this sequence for DNA-protein complex formation. Inhibitory factor-2 (IF2), a metaloprotein, binds to these 12 bp GC-rich repeats (Karsenty and de Crombrugghe, 1990). Methylation interference assays indicated that the upstream 12 bp repeat is the major region for DNA-protein contact and that the downstream repeat is less important for binding. There is mutually exclusive binding of IF2 and CBF to their respective recognition elements, which could play an important role in the regulation of α1(I) procollagen gene expression. Nehls *et. al.* (1991) have subsequently shown that another factor, NF-1, and not CBF is the major factor that interacts with the mouse α1(I) procollagen promoter. Since there is a striking similarity between the downstream, and to a lesser extent the upstream, CCAAT boxes with the NF-1 consensus sequence, these investigators have classified both these elements as NF-1 sites. Binding and mutation analysis have shown that the downstream NF-1 site is more important in activating the gene (Nehls *et. al.*, 1991). Contrary to previous studies (Karsenty and de Crombrugghe, 1990), Nehls *et. al.* (1991, 1992) have demonstrated that Sp1 actually binds to the 12 bp repeat and that NF-1 and Sp1, like CBF and IF2, bind in a mutually exclusive manner to their respective binding sites. Over-expression of Sp1 in NIH-3T3 fibroblasts reduces basal promoter activity by about 2-fold, whereas over-expression of NF-1 resulted in enhanced transcription. These investigators therefore postulated that the mouse α1(I) procollagen gene is regulated by the relative levels of Sp1 and NF-1.

A second transcriptional repressor, inhibitory factor 1 (IF1), binds to two or more upstream adjacent sites. One of these sites, element A, is situated between nucleotides -190 and -170, while the other, element B, extends from nucleotides -160 to -130. Competition experiments have shown that IF1 binds more readily to element A than to element B

(Karsenty and de Crombrugghe, 1990). Transient transfection experiments where 5'-flanking sequences (up to -2500) of the human (Boast *et. al.*, 1990) and mouse (Rippe *et. al.*, 1989; Karsenty *et. al.*, 1990) α1(I) procollagen promoters have been added to their proximal promoters resulted in decreased expression of the chimeric constructs. Contrary to these results, microinjection of constructs into xenopus oocytes caused an increase in transcription (Rossouw *et. al.*, 1987). These investigators have suggested that both positive and negative elements are present within these 5'-flanking regions. One of these factors is an inhibitory factor present in type I collagen producing cell lines and binds to a recognition element situated between nucleotides -339 to -361 in the mouse promoter. This factor is probably involved, together with cell-specific activators, in the tissue-specific regulation of the mouse α1(I) procollagen expression (Ravazzolo *et. al.*, 1991). An Sp1 consensus sequence (-466 to -459) have been identified within the distal human promoter and binds purified Sp1 (Bornstein *et. al.*, 1987; Liska *et. al.*, 1992). A strong tissue-specific enhancer element has also been identified between -2300 and -444 (Slack *et. al.*, 1991). Ritzenthaler *et. al.* (1991) have identified a TGF-ß activating element (TAE) at -1600 within the rat α1(I) procollagen promoter which resembles NF-1 and AP-2 binding sites. A 82 KDa TAE-binding protein, which is different to either NF-1 or AP-2, has been identified in human fibroblast nuclear extracts (Ritzenthaler *et. al.*, 1993). In an attempt to study the tissue-specific expression of type I collagen in bone, Lichtler *et. al.* (1989) have shown that the rat α1(I) promoter carries the necessary information for the cell-specific regulation by the calcitrophic hormone, 1,25-dihydroxyvitamin D_3, in bone cells but not in fibroblasts. More recently, Pavlin *et. al.* (1992) have identified potential osteoblast and odontoblast specific stimulatory elements located between -3521 and -1672 in the rat promoter.

Sequences within the first introns of the mouse and human α1(I) procollagen genes are also involved in modulating transcription. Unfortunately, controversy exists as to the nature and effect of the intronic sequences on α1(I) procollagen gene expression. Both positive and negative

regulatory elements are located within the first intron of the mouse and human gene (Bornstein *et. al.*, 1987, 1988; Bornstein and Mc Kay, 1988; Rippe *et. al.*, 1989; Liska *et. al.*, 1990; Boast *et. al.*, 1990). A 274 bp fragment (nucleotides +820 to +1093) which inhibits the α1(I) procollagen promoter activity in an orientation-dependent manner, has been identified within the 3'-half of the first intron of the human gene (Bornstein *et. al.*, 1987; Rossouw *et. al.*, 1987). Bornstein *et. al.* (1988) have postulated that this element inhibits gene expression by interacting with promoter sequences through co-operative DNA-protein interactions. Simkevich *et. al.* (1992) have also postulated that orientation-dependent interactions between *cis*-acting elements within the promoter and the first intron determine the tissue-specific expression of the human α1(I) procollagen gene. Several *cis*-elements, including a Sp1 motif (+925 to +934), a viral core enhancer element (+999 to +1007) and a DNase 1 protection footprint extending from nucleotides +951 to +978, have been identified within this negative element (Bornstein *et. al.*, 1987; Rossouw *et. al.*, 1987). Liska *et. al.* (1992) have shown that purified Sp1 binds independently to the Sp1 motif and to a GC-rich region just 5' of the viral core enhancer element. The first intron of the human gene also contains a positive orientation-dependent element extending from +292 to +670. This element contains an AP-1 consensus sequence at approximately +600 (Liska *et. al.*, 1990). Interestingly, the chicken promoter and first intron contains two unique DNA sequences; fifteen tandem repeats of the sequence $^{5'}$GGGGAGA$^{3'}$ have been identified within the first intron and at least 25 copies of a polymorphic, 23 bp tandemly repeated sequence within the distal promoter (Finer *et. al.*, 1987).

Most cells of mesenchymal origin produce type I collagen, while ectodermal, endodermal or hematopoietic lineage cells do not (Adams, 1989). Type I collagen is present in almost all connective tissues although quantitative differences may exist. As an example of the tissue-specific expression of type I collagen, the mechanisms for the inhibition of type I collagen synthesis in hyaline cartilage have been studied. Interestingly, type I collagen mRNA has been identified in cultured chicken chondrocytes and is found exclusively in the

nucleus, apparently unprocessed (Saxe *et. al.*, 1985). In contrast, a smaller α2(I) mRNA product is processed and transported to the cytoplasm where it is found primarily associated with ribosomes (Bennett and Adams, 1987). Chondrocytes synthesise a non-collagenous product from the α2(I) mRNA which is transcribed from an alternate promoter and transcription start site situated in the second intron of the COL1A2 gene. This alternate promoter contains a CCAAT box but lacks a TATA box. The first two exons of the COL1A2 gene are replaced by a novel 96 bp exon so that the new mRNA contains four open reading frames which are all out of frame with the collagen coding sequences (Bennett and Adams, 1990).

1.3.6.2 TYPE II COLLAGEN GENE

During the differentiation of mesenchymal cells into chondrocytes, type II collagen synthesis is initiated and the synthesis of type I collagen ceases (reviewed by von der Mark, 1980; Linsenmayer, 1981). Transient transfection assays, using chondrocytes and other cell types, have been used to identify a weak promoter, a cell-specific enhancer and silencer elements within the type II collagen gene (Horton *et. al.*, 1987; Savanger *et. al.*, 1990). The 550 bp enhancer element, which increases the activity of the rat α1(II) procollagen promoter in chondrocytes, but not in fibroblasts and myoblasts, has been identified within the first intron and shown to be activated during the differentiation of mesenchyme cells to chondrocytes (Horton *et. al.*, 1987). Experiments with transgenic mice have confirmed that the tissue-specific expression of the type II collagen gene is dependent on the enhancer and promoter (reviewed by Yamada *et. al.*, 1990). Furthermore, similar studies have shown that both these elements are sufficient for the regulation of the gene during development (Palmiter *et. al.*, 1987; Breitman *et. al.*, 1987; Bruggeman *et. al.*, 1989). Finally, Savagner *et. al.* (1990) have identified two silencer elements, between -360 and -460, and -620 and -700, within the rat α1(II) procollagen promoter. The tissue- and developmental-specific expression of the type II procollagen gene is therefore controlled by both negative and positive *cis*-elements.

The rat α1(II) procollagen promoter contains a TATA box, several GC-rich sequences and an enhancer core element. The enhancer core element is located around -280, while direct and inverted Sp1 hexanucleotide consensus sequences are situated between -450 and the TATA box (Kohno *et. al.*, 1985). There is considerable homology between the rat and human α1(II) procollagen promoters (Nunez *et. al.*, 1986). The enhancer contains a GC box, a glucocorticoid response element, a NF-1 consensus sequence and many direct and inverted repeats. DNA-binding studies with chondrocyte nuclear extracts suggest that a number of regions within the enhancer interacts with distinct nuclear proteins (Yamada *et. al.*, 1990). Wang *et. al.* (1991) have identified a decamer sequence $^{5'}$CACAATGCAT$^{3'}$ in the middle of the enhancer that binds a chondrocyte-specific protein(s) which is required for enhancer activity. The DNA-binding activity of this protein(s) was induced during differentiation of limb bud mesenchymal cells into chondrocytes. The silencer elements contain consensus sequences that are found in silencers of other genes. DNA-binding studies have shown that *trans*-acting factors in non-chondrocytic nuclear extracts, such as HeLa, but not in chondrocyte nuclear extracts interact with DNA fragments containing the silencer elements (Savagner *et. al.*, 1990).

1.3.6.3 TYPE III COLLAGEN GENE

In contrast to the type I collagen genes, relatively little is known about the mechanisms that regulate the expression of the COL3A1 gene (reviewed by de Crombrugghe *et. al.*, 1990). Although types III and I collagens are co-ordinately expressed in many tissues, the ratios of these two collagen types varies from tissue to tissue and also during development (Fessler and Fessler, 1978; Kuhn, 1981). The levels of both collagen types are similarly affected by TGF-ß (Ignotz and Massague, 1986) and hepatic fibrogenic factor (Choe *et. al.*, 1987). Fibroblasts transformed with Rous sarcoma virus and v-mos oncogenes synthesise less type I and type III collagen mRNA (Howard *et. al.*, 1978; Schmidt *et. al.*, 1985). It can therefore be postulated that the mechanisms that regulate the expression of types III and I

collagen genes may share common regulatory elements, or that they may be controlled by a "master" gene.

A 2.3 Kb mouse α1(III) procollagen promoter fragment fused to a reporter gene was used in transient transfection studies to identify the *cis*-acting elements that regulate the expression of the type III collagen gene (Mudryj and de Crombrugghe, 1988). In contrast to the mouse α2(I) procollagen promoter, deletion analysis of the α1(III) promoter suggested that regulation is predominantly by negative regulatory elements. Although the precise location of these elements is not known, one of them lies within a 50 bp region between -350 and -300 (Mudryj and de Crombrugghe, 1988). Two positive recognition elements, located at -122 to -106 and at -83 to -61, have been identified in the proximal 150 bp of the mouse α1(III) promoter. *Cis*-elements in the mouse α1(I) and α2(I) procollagen promoters, such as the CCAAT box, are not present in the mouse α1(III) promoter and the heterotrimeric CBF does not bind to this 150 bp region of the α1(III) promoter (Hatamochi *et. al.*, 1986; Maity *et. al.*, 1988). Contrary to expectations, these observations suggest that different mechanisms regulate the expression of the type I and type III procollagen genes. AP-1 or a related factor probably binds between -122 and -116 to activate transcription (Ruteshouser and de Crombrugghe, 1989). A monomeric, heat-resistant, 95 KDa factor (the B element-binding factor or BBF) has been purified from HeLa and NIH-3T3 cells and shown to bind between -83 and -61 (Ruteshouser and de Crombrugghe, 1992). Binding of BBF to its recognition element is probably modulated by other factors such as an inhibitor of BBF DNA-binding activity (Ruteshouser and de Crombrugghe, 1992). Additional work needs to be done to fully understand the role of these and other factors in the co-ordinated control of type III and type I collagen gene expression.

1.3.6.4 TYPE IV COLLAGEN GENE

The transcriptional regulation of type IV collagen and other basement membrane protein genes is more extensively studied in F9 teratocarcinoma stem cells which differentiate into parietal endoderm-like cells when treated with retinoic acid and dibutyryl-cAMP (Strickland *et. al.*, 1980). Since

differentiation is associated with a co-ordinated increase in the synthesis of basement membrane proteins, these cells serve as an ideal model system for studying the regulation of type IV collagen genes (Burbelo *et. al.*, 1990 and references therein).

The expression of type IV collagen is tightly regulated, both spatially and temporally, during development when extracellular deposition is important for cell migration and differentiation (Timpl and Dziadek, 1986). The co-ordinate regulation of both the human and mouse α1(IV) and α2(IV) collagen genes have a number of unique characteristics; firstly, both genes are located on the same chromosomal region in a head to head configuration and are separated by approximately 130 bp of DNA (Poschl *et. al.*, 1988; Burbelo *et. al.*, 1988). Comparison of the 130 bp mouse and human region revealed a high level of sequence homology (nearly 90%) (Poschl *et. al.*, 1988; Burbelo *et. al.*, 1988). Since this region is able to direct the expression of a reporter gene in an orientation independent manner, it serves as a common bidirectional promoter for both genes, although with a low transcriptional activity (Poschl *et. al.*, 1988; Burbelo *et. al.*, 1988). Two DNA-binding sites have been identified within the short TATA box-less bidirectional promoter (Bruggeman *et. al.*, 1992). The first is a consensus sequence situated within an interrupted dyad symmetry in the centre of the promoter. The other site is a novel $^{5'}$CCCTCCC$^{3'}$ motif which is present in several other extracellular matrix promoters, including the α2(I) procollagen promoter (chapter 4). Factors from Engelbreth-Holm-Swarm (EHS) mouse sarcoma protein extracts interact with both these recognition elements (Bruggeman *et. al.*, 1992).

A cell-specific 210 bp enhancer, which stimulates the common promoter of both the α1(IV) and α2(IV) collagen genes, has been identified within the first intron of the mouse COL4A1 gene. Two *cis*-acting elements (A and B) have been identified within the enhancer (Burbelo *et. al.*, 1988). Two putative positive acting factors, one of 37 KDa and the other of 94 KDa, which bind to element A ($^{5'}$CCTTATCTCTGATGG$^{3'}$) has been purified and shown to be required for the effective transcription of both genes (Burbelo *et. al.*, 1991). A

putative negative *cis*-element is probably present in the third intron of the human COL4A2 gene (Poschl *et. al.*, 1988). Both positive and negative factors, therefore, probably regulate the co-ordinated expression of the two major type IV collagen genes. Methylation status and/or chromatin structure of the type IV collagen promoter and/or enhancer may also play an important role in silencing the expression of these genes in cells that do not synthesise type IV collagen (Burbelo *et. al.*, 1990, 1991).

1.3.6.5 TYPE V COLLAGEN GENE

As already discussed (section 1.2.1), type V collagen is co-expressed with type I collagen and probably plays a role in regulating the diameter of type I containing collagen fibres. To understand the mechanisms that regulate the expression of type V collagen and to gain insight into its co-ordinated expression with other collagen types, the regulatory elements within the α2(V) procollagen gene have been studied and characterised. Greenspan *et. al.* (1991b) have isolated a 17 kb genomic clone containing the 5'-portion of the human α2(V) procollagen gene and nucleotide sequence analysis revealed remarkable similarity between the promoter, the first exon and the 5'-end of the first intron of the α1(III) gene. The homologous regions include the -120 bp proximal promoters, especially near the TATA box between -37 and -27, the transcription start sites and 5'-untranslated regions of the two genes. This promoter, like the α1(III) promoter, does not contain a CCAAT box (Greenspan *et. al.*, 1991b; Truter *et. al.*, 1992b).

More recently, Truter *et. al.* (1992a) have identified a 52 bp region of the human α2(V) procollagen promoter that is essential for the cell type-specific expression of a reporter gene in transient transfection assays. This region of the human promoter is highly homologous with the equivalent region of the mouse promoter. Two novel nuclear factor binding sites, FP-A (-115 to -99) and FP-B (-149 to -118), have been identified in this short region of the human promoter (Truter *et. al.*, 1992a). Both FP-A and FP-B binding factors are transcriptional activators. Negative *cis*-acting elements have also been identified within the

first intron and 5'-flanking sequences of the human COL5A2 gene (Greenspan *et. al.*, 1991b).

1.3.6.6 TYPE VI COLLAGEN GENE

Type VI collagen gene expression is independently regulated from those of the other major constituents of the extracellular matrix, such as types I and III collagens (Hatamochi *et. al.*, 1988; Oono *et. al.*, 1993). Furthermore, the two type VI collagen genes, α1(VI) and α2(VI), which are located on human chromosome 21 are co-ordinately regulated, whereas the α3(VI) gene on chromosome 2 is regulated independently (Schreir *et. al.*, 1988; Hatamochi *et. al.*, 1989; Oono *et. al.*, 1993).

The human α2(VI) collagen promoter contains a TATA box, located between -40 and -50, flanked by SP1 consensus sequences and two potential CCAAT boxes (Saitta *et. al.*, 1992). Four mRNA species that differ in sequences at their 5'-UTRs are synthesised from the human α2(VI) gene. Transcription of one of the mRNA species starts at exon 1, while three minor species are transcribed from the second exon (exon 1A). It is highly likely that the minor mRNA species are transcribed from an alternative promoter. This hypothetical promoter, which lacks a TATA and CCAAT boxes, contains several SP1 consensus sequences (Saitta *et. al.*, 1992).

Contrary to other collagen genes, the chicken promoter is structurally very different from the human promoter (Koller *et. al.*, 1991; Saitta *et. al.*, 1992). The chicken promoter lacks both the TATA and CCAAT boxes. Instead it contains several Sp1 and ETF (Kageyama *et. al.*, 1989) consensus sequences and, like most TATA box-less genes, has multiple transcription initiation sites distributed over 75 bp of DNA. The CpG dinucleotide occurs with a high frequency within the 5'-flanking region of the gene. A 207 bp region of the promoter containing the entire CpG island and all the transcription initiation sites has strong promoter activity when linked to a reporter gene in transient transfection assays (Koller *et. al.*, 1991). The chicken α2(VI) collagen

upstream promoter also contains a 403 bp purine/pyrimidine motif of unknown function (Koller *et. al.*, 1991).

1.3.6.7 TYPE IX COLLAGEN GENE

Two tissue-specific forms of type IX collagen are produced by the alternative use of two transcription start sites and alternative splicing. The shorter corneal form of the α1(IX) chain lacks the N-terminal globular domain, which is present in the cartilage form. Approximately 20 kb of genomic DNA, containing exons 1 to 6, separate the cartilage and corneal transcription initiation sites in the chicken α1(IX) collagen gene. The corneal-specific promoter is situated near the 3'-end of the sixth intron and contains a TATA box (approximately -13), CCAAT box (-46 to -42) and an AP-1 binding site (-341 to -334). The cartilage-specific promoter is structurally different in that it contains a Sp1 consensus sequence (-93 to -88) and other motifs at -50, -630 and -107 which are similar to sequences within the apoE promoter and the immunoglobulin enhancer (Nishimura *et. al.*, 1989).

1.4 THE CCAAT BOX BINDING FACTORS

Most eukaryotic promoters contain a CCAAT box in either a direct or inverted orientation in which flanking sequences may differ tremendously (Chodosh *et. al.*, 1988a; Dorn *et. al.*, 1988; Santoro *et. al.*, 1988). Some promoters even contain two or more functional CCAAT boxes (Park *et. al.*, 1990; Liu *et. al.*, 1991). A number of structurally distinct families of transcription factors bind to CCAAT boxes to either activate or repress transcription (Dorn *et. al.*, 1988; Raymondjean *et. al.*, 1988). A single cell type can contain several CCAAT binding factors which bind to different CCAAT boxes (Dorn *et. al.*, 1988). This is possible because the flanking sequence of the CCAAT motif probably determine the affinity of a transcription factor for a specific CCAAT box. Based on their structural complexity and function these factors can be divided into several families of CCAAT box binding proteins such as CTF/NF-1, C/EBP, the hetero-multimeric factors and others.

1.4.1 CCAAT TRANSCRIPTION FACTOR/NUCLEAR FACTOR 1 (CTF/NF-1)

Nuclear factor 1 (NF-1), also known as CCAAT transcription factor (CTF) and CTF/NF-1, is a family of sequence-specifc DNA-binding proteins which can either function as positive or negative transcription factors (Jones *et. al.*, 1987; Gil *et. al.*, 1988; Angel *et. al.*, 1988) or as factors in DNA replication (Nagata *et. al.*, 1983). Rosenfeld and Kelly (1988) used DNA-affinity chromatography to purify the factor from HeLa cells to apparent homogeneity, and have shown that NF-1 comprises a population of related polypeptides with apparent molecular weights ranging from 52 to 66 KDa. In humans the individual members of this family of proteins, containing highly conserved N-termini and variable C-termini, may be produced by alternate splicing of a single gene product (Santoro *et. al.*, 1988). Chicken NF-1 proteins, on the other hand, are encoded by at least three distinct genes, designated NF-1-A, NF-1-B and NF-1-C (Rupp *et. al.*, 1990). Both the forms and levels of NF-1 vary in different cell lines (Goyal *et. al.*, 1990).

The 499 amino acid long CTF-1 factor contains a number of functional domains. The proline-rich transcriptional activation domain is situated in the C-terminal region while the DNA-binding, protein dimerisation and DNA-replication domains are all situated in the N-terminal 240 amino acids of the protein (Mermod *et. al.*, 1989). The DNA-binding and dimerisation domains are situated on the N-terminal and C-terminal portions of this 240 amino acid region respectively (Gounari *et. al.*, 1990).

NF-1 binds as a dimer to the recognition sequence $5'TGGN_{6-7}GCCAA3'$ (Jones *et. al.*, 1987; Gounari *et. al.*, 1990). The flanking sequences and the length of the 6 to 7 bp spacer region can greatly modulate binding efficiency (Gronostajski, 1986). NF-1 may also bind to NF-1-like sites, such as the inverted NF-1-like site within the mouse α2(I) procollagen promoter which binds to NF-1 or a NF-1-like protein with lower affinity (Oikarinen *et. al.*, 1987). Zorbas *et. al.* (1992) have shown that NF-1 can bind to the $5'GGGN_6GCCAG3'$ motif in the human α-globin gene promoter. Since the 3'-end of the NF-1 motif resembles a CCAAT box,

NF-1 can bind to some CCAAT boxes and has therefore been classified as a CCAAT box binding factor (Jones *et. al.*, 1987). Often promoters contain both NF-1 and CCAAT motifs which independently bind NF-1 and other CCAAT box binding factors (Oikarinen *et. al.*, 1987; Zorbas *et. al.*, 1992).

1.4.2 CCAAT/ENHANCER BINDING PROTEIN (C/EBP) AND RELATED PROTEINS

C/EBP, which is not related to NF-1, is a sequence-specific DNA-binding protein that was originally shown to bind to the CCAAT box within the herpes virus thymidine kinase promoter and the murine sarcoma virus long terminal repeat and called the CCAAT binding protein (CBP) (Graves *et. al.*, 1986). The same factor, which bound to the enhancer core element ($^{5'}$TGTGGWWWG$^{3'}$) present in many viral enhancers, was purified independently from rat liver nuclei and called the enhancer binding protein (EBP) (Johnson *et. al.*, 1987). This heat-stable 42 KDa factor is now referred to as the CCAAT/enhancer binding protein (C/EBP) and is known to be a family of related factors which include C/EBPα, C/EBPβ, C/EBPδ and C/EBP-γ (Lamb and McKnight, 1991 and references therein). C/EBP can also bind to a third element, which resembles the enhancer core element, the ATF/CRE elements (Bakker and Parker, 1991). The binding sites for C/EBP, therefore, have the consensus sequence $^{5'}$TKNNGVAAK$^{3'}$ (Yuh and Ting, 1991). Several C/EBP-like factors, which have DNA-binding specificities similar to those of C/EBP, have also been identified and are believed to be members of a C/EBP super-family of transcription factors. These factors include AGP/EBP (Chang *et. al.*, 1990), the IL-6DBP family (Poli *et. al.*, 1990), DBP (Muller *et. al.*, 1990), LAP (Descombes *et. al.*, 1990) and the C/ATF family (Vallejo *et. al.*, 1993). The regulation of gene expression during different stages of cell differentiation and growth is probably tightly controlled by the C/EBP and C/EBP-like factors.

The DNA-binding domain of this super-family of factors is bipartite, consisting of a dimerisation interface, the leucine zipper, and a DNA-binding basic region (Landschulz *et. al.*, 1988, 1989). Members of the C/EBP family can interact with each other to form heterodimers (reviewed by

Lamb and Mc Knight, 1991). They can also form heterodimers with C/EBP-like proteins, such as the C/ATF family (Vallejo *et. al.*, 1993). The DNA-binding domain of C/EBP can also be phosphorylated by protein kinase C in a Ca^{2+} and lipid dependent manner (Wegner *et. al.*, 1992; Mahoney *et. al.*, 1992). Post-translational modifications are becoming increasingly important in modulating the DNA-binding ability and the transcription activational properties of C/EBP.

C/EBP and the related proteins are present in various tissues including liver, lung, adipose, gut and placenta (Birkenmeier *et. al.*, 1989). It has been postulated that C/EBP activity is normally associated with only fully differentiated, non-proliferating cells that metabolise lipids and cholesterol at high rates (Birkenmeier *et. al.*, 1989; Christy *et. al.*, 1989, Friedman *et. al.*, 1989). The factor may function exclusively in these cells by activating genes responsible for a specialised phenotype (Friedman *et. al.*, 1989). Although C/EBP activity is present in differentiated cells it does not appear to be expressed early in the differentiation pathway or to regulate the initiation of differentiation, but probably helps to maintain the differentiation status of a cell (Birkenmeier *et. al.*, 1989; Friedman *et. al.*, 1989). A number of tissue-specific genes, like the serum albumin gene (Friedman *et. al.*, 1989), the human clotting factor IX gene (Crossley and Brownlee, 1990) and the phospho-enolpyruvate carboxykinase (PEPCK) gene (Park *et. al.*, 1990;), are activated when C/EBP binds to *cis*-elements within their promoters. In fact C/EBP binds to several sites along the PEPCK promoter (Park *et. al.*, 1990; Liu *et. al.*, 1991). C/EBP may also play a direct role in regulating the genetic programme responsible for the termination of cell growth such as the termination of meiotic growth of adipoblasts (Umek *et. al.*, 1991) and other pluripotent cells (Taylor and Jones, 1979).

1.4.3 THE HETERO-MULTIMERIC CCAAT BINDING FACTORS

A large number of related and in some cases even identical CCAAT binding factors, which are distinct from CTF/NF-1 or C/EBP, have been purified and characterised independently by several of investigators. These factors include NF-Y (Dorn

et. al., 1987), CBF (Hatamochi *et. al.*, 1988), CP1 (Chodosh *et. al.*, 1988a), αCP1 (Kim and Sheffery, 1990), the yeast Hap-2,3,4 factors (Forsbery and Guarente, 1989) and others (table 1.2). Based on their structural complexity, these factors belong to a distinct category of CCAAT binding factors with heterologous subunits. Initially these complexes were identified as heterodimers but recent evidence has shown that an ever increasing number of these factors are actually heterotrimers. The structure and function of a few of these hetero-multimeric factors are explained in detail.

Table 1.2 Subunit composition of the hetero-multimeric factors. Tabulation of the species and tissue or cell types from which these factors have been identified. Subunit C of some of these factors (n.d.) has not been identified or characterised to date.

FACTOR	SPECIES	SOURCE	SUBUNITS		
			B	A	C
α-CP1	Mouse	MEL cells	γ	ß	α
CBF	Rat	liver	CBF-B	CBF-A	CBF-C
CBF	Mouse	NIH-3T3	CBF-B	CBF-A	CBF-C
CP1	Human	HeLa cells	CP1B	CP1A	n.d.
EFI	Chick	Embryos	EFI_B	EFI_A	n.d.
EFI	Rat	liver	EFI_B	EFI_A	n.d.
NF-Y	Mouse	Ubiquitous	NF-YA	NF-YB	n.d.
Hap-2,3,4	Yeast	-	Hap-2	Hap-3	Hap-4

In yeast, a heterotrimeric transcription factor, Hap-2,3,4, binds to a CCAAT box in the upstream activation site, UAS2, of the CYC1 gene to activate transcription (Forsburg and Guarente, 1989 and references therein).

The transcriptional activator, α-CP1, which binds to the α-globin promoter was initially identified and purified from murine erythroleukemia (MEL) cells (Cohen *et. al.*, 1986; Barnhart *et. al.*, 1988). α-CP1 constitutes at least seven polypeptides which have been divided into three distinct classes designated α, ß and γ. The α class consists of a

single 35 KDa polypeptide while two polypeptides (27 KDa and 28 KDa) make up the ß class. The γ class comprises four related polypeptides with apparent molecular masses of 33, 34, 37 and 38 KDa. One polypeptide from each class forms a functional heterotrimeric factor which has a native molecular mass of ±101 KDa. The α and ß polypeptides form an inactive core complex which is stable at moderately high ionic strengths. The γ polypeptide is weakly associated with the stable αß core and required for α-CP1 activity (Kim and Sheffery, 1990).

The transcriptional activator, α-CP1, competes with a stronger monomeric activator, α-CP2, for binding to distinct overlapping recognition sites in a mutually exclusive manner to regulate the α-globin gene (Kim *et. al.*, 1990; Lim *et. al.*, 1992). Interestingly, CCAAT-binding factors can also compete with transcriptional repressors for overlapping binding sites in a mutually exclusive manner to regulate gene expression (Barberis *et. al.*, 1987).

The heterotrimeric CCAAT binding factor, CBF, was initially identified in NIH-3T3 fibroblasts (Hatamochi *et. al.*, 1986; Oikarinen *et. al.*, 1987) and later purified to homogeneity from rat liver nuclear extracts (Hatamochi *et. al.*, 1988). CBF is also a transcription activator which binds to the proximal CCAAT box in the mouse α1(I) and α2(I) procollagen promoters (Maity *et. al.*, 1988). The interaction of three distinct polypeptides (CBF-A, CBF-B and CBF-C) are required for CBF DNA-binding activity. CBF-A forms a stable non-covalent, anionic and heat labile complex with the 40 KDa CBF-C subunit which can only be dissociated by denaturing agents. The cationic and heat stable CBF-B subunit forms an easily dissociable complex with the CBF-A.CBF-C complex (Hatamochi *et. al.*, 1988; Maity *et. al.*, 1992). Full-length CBF-A and CBF-B cDNA clones have been isolated and a protein of 207 amino acid with a calculated molecular weight 25 KDa is encoded by the CBF-A cDNA. There is a high sequence identity between significant portions of CBF-A and the yeast Hap-3 polypeptide (Vuorio *et. al.*, 1990). The CBF-B cDNA clone contains an open reading frame of 1023 bp encoding a protein of 341 amino acids with a calculated molecular mass of 41 KDa. The C-terminal portion of this polypeptide is

highly homologous with a portion of the Hap-2 protein (Maity *et. al.*, 1990).

The transcription factor CP1 has been partially purified from HeLa cell extracts (Chodosh *et. al.*, 1988a). CP1 binds to the consensus sequence, $^{5'}YN_6RRCCAATCANYK^{3'}$, within the human α-globin promoter, the human hsp70 promoter, H-2K promoter, adenovirus MLP and MSV LTR. CP1 consists of at least two subunits, CP1A and CP1B, which are both required for CP1 DNA-binding activity and, like the other hetero-multimeric factors, forms a complex independently of a CP1 binding site. The CP1A and CP1B subunits are functionally similar to the yeast Hap-3 and Hap-2 proteins respectively (Chodosh *et. al.*, 1988b). CP1A and CP1B can also form functionally stable hybrid complexes with their complementary subunits of the yeast CCAAT binding factor Hap-2,3,4.

Enhancer factor I (EFI), a transcriptional activator, which binds to an inverted CCAAT box within the Rous sarcoma virus LTR (Sealy and Chalkley, 1987), is also a hetero-multimeric factor consisting of at least two subunits, EFI_A and EFI_B, which forms a heterodimer that binds with high affinity to its recognition sequences (Faber and Sealy, 1990). A second DNA-binding form of EFI, probably a multimer of EFI_A, is present in avian nuclear extracts under certain conditions (Ozer *et. al.*, 1990). The molecular weight of the avian EFI_A subunit has been reported to be between 43 and 60 KDa (Faber and Sealy, 1990). The rat EFI_A cDNA encodes a 322 amino acid protein (35 KDa) (Faber and Sealy, 1990) and is nearly identical to two previously described human CCAAT box binding proteins, dbpB (Sakura *et. al.*, 1988) and YB-1 (Didier *et. al.*, 1988). These proteins contain a novel structural motif containing alternating regions of positively and negatively charged amino acid residues separated by short glutamine-rich regions (Ozer *et. al.*, 1990).

Nuclear factor Y (NF-Y) is a ubiquitous CCAAT box binding factor which binds to the consensus sequence $^{5'}RRCCAATCAG^{3'}$ either in the direct or inverted orientation in a variety of promoters (Wuarin *et. al.*, 1990; Li *et. al.*, 1992 and references therein). NF-Y was originally identified as a sequence-specific DNA-binding protein that recognises the Y

box within the major histocompatibility complex (MHC) class II gene, E_α (Dorn *et. al.*, 1987). The binding and physical properties and subunit composition (NF-YA and NF-YB) of NF-Y have been reported (Hooft van Huijsduijnen, 1987, 1990). The NF-YA and NF-YB genes contain at least 9 and 5 exons respectively (Li *et. al.*, 1992) and different forms of NF-YA are produced by alternative tissue-specific splicing. Certain forms are more predominant in the brain, liver, lung and in fibroblast and teratocarcinoma cells, while other forms predominate in the thymus and spleen and in B lymphoid cell lines. NF-YA and NF-YB are also homologous to Hap-2 and Hap-3 respectively (Li *et. al.*, 1992).

1.4.4 CCAAT BINDING PROTEINS AND CELL CYCLE REGULATION

A number of cell-cycle regulated genes whose expression increases at the G_1/S phase boundary, contain CCAAT boxes within their proximal promoters. These include, amongst others, the thymidine kinase (TK) (Flemington *et. al.*, 1987), histone H1 (Gallinari *et. al.*, 1989) and hsp70 genes (Wu, B.J. *et. al.*, 1987). The expression of these three genes is coupled with the onset of DNA synthesis (Milarski and Morimoto, 1986). The human proximal TK promoter contains two inverted CCAAT boxes located at -36 and -67, both of which, especially the distal one, are required for TK gene expression (Lipson *et. al.*, 1989; Kim and Lee, 1991). Inverted and/or tandem CCAAT boxes are also found in other cell-cycle regulated genes (Busslinger *et. al.*, 1980; Harvey *et. al.*, 1982; Sierra *et. al.*, 1983) and are recognised by specific CCAAT box binding factors which stimulate their expression (Lum *et. al.*, 1990; Gallinari *et. al.*, 1989).

Several CCAAT binding proteins which activate cell-cycle regulated genes have been purified and/or characterised from a variety of cell lines and tissues and have been shown to be distinct from C/EBP or from the hetero-multimeric factors (table 1.3). In contrast a factor which may be related to ubiquitous NF-Y binds to the CCAAT boxes within the human TK promoter (Arcot *et. al.*, 1989). Investigators have also shown that several distinct or related CCAAT binding proteins can bind *in vitro* to the same promoter element (table 1.3)

(Jones *et. al.*, 1985; Graves *et. al.*, 1986; Morgan *et. al.*, 1987; van Wijnen *et. al.*, 1988; Lum *et. al.*, 1990)

Table 1.3 CCAAT box binding proteins and cell cycle regulation. Several cell-cycle genes are regulated by different CCAAT box binding proteins. The molecular weights (MW) of these factors are given in KDa.

GENE/PROMOTER	FACTOR	MW	SOURCE	SPECIES
Human TK	CBP/tk		Fibroblast	Human
Human TK	-	33	Fibroblast	Mouse
Human TK	"NFY-like"		HeLa	Human
HSV-1 TK	CTF		HeLa	Human
HSV-1 TK	CBP	43	Liver	Rat
Human hsp70	CTF		HeLa	Human
Human hsp70	CP1			
Human hsp70	CBF	114	Fibroblast	Human
Human histone H1	H1TF2	47	HeLa	Human
Human histone H1	HiNF-B			Mouse

The transcriptional activation properties of some or even all of these CCAAT binding factors are cell-cycle regulated and correlates with the activation of their target genes. These factors may therefore stimulate transcription of the cell-cycle regulated genes during the G_1/S phase boundary of the cell cycle. The cell-cycle dependent DNA-binding activity of the CCAAT binding protein for the TK gene (CBP/tk) may activate the TK gene prior to the onset of DNA synthesis (Pang and Chen, 1993). Alternatively, during the cell-cycle when these genes are not expressed, a repressor protein(s) binds directly to the DNA-bound CCAAT-binding protein to inhibit transcription. The repressor protein(s) dissociates from the CCAAT binding factor during the G_1/S phase transition and the cell-cycle genes are activated. At the end of the activation period the repressor protein reassociates with the bound CCAAT binding factor to inactivate the gene. Knight *et. al.* (1987) have shown, for example, that although a murine CCAAT binding factor interacts with the human TK promoter throughout the cell-cycle, the nature of this DNA-protein complex is cell-cycle

dependent. Secondly, the human hsp70 gene is transcriptionally repressed when wild-type p53 interacts with CBF (Agoff *et. al.*, 1993). The adenovirus E1a protein, on the other hand, activates the hsp70 gene via protein-protein interactions with DNA-bound CBF. E1a could nullify the activity of the repressor and allow CBF to activate the gene or, alternatively, it may mimic the function of a coactivator (Lum *et. al.*, 1992).

1.4.5 OTHER CCAAT BINDING PROTEINS

The α subunit CCAAT binding factor (αCBF) activates the α subunit gene of the pituitary and placental glycoprotein hormone. This transcription factor was identified in human choriocarcinoma cells and is also present in other cells such as HeLa and HepG2 cells. Monomeric αCBF has an apparent molecular weight of 53 KDa and is probably not related to CTF/NF-1, C/EBP or one of the hetero-multimeric CCAAT-binding factors (Kennedy *et. al.*, 1990). CCAAT-binding proteins may also regulate the age-dependent decreased expression of some genes. The age-related decrease in TK gene expression, for example, is coupled to a similar decrease in CBP/tk DNA-binding activity (Pang and Chen, 1993).

In summary, the CCAAT box is not recognised by a single protein but by a class of structurally distinct proteins whose inter-relationships are not clear. CCAAT binding proteins play an important role in a variety of physiological processes, which include regulation of the cell-cycle, lipid and cholesterol metabolism, DNA replication, cell growth arrest, etc. Interestingly, related factors appear to regulate a specific subset of these functions and can also compete with other transcription factors (both activators and repressors) for binding to distinct sites in a mutually exclusive manner to regulate gene expression. Activators and repressors can also bind directly to CCAAT-binding factors via protein-protein interactions to regulate gene expression.

CHAPTER 2

IDENTIFICATION OF DNA-BINDING PROTEINS WHICH INTERACT WITH THE PROXIMAL α2(I) PROCOLLAGEN PROMOTER

2.1 INTRODUCTION

The synthesis of type I collagen by a normal diploid human embryonic lung fibroblast cell line (WI-38) (ATCC CCL-75) and two transformed WI-38 cell lines, one transformed with simian virus 40 (SVWI-38) (de Haan *et. al.*, 1986) and the other by γ-radiation (CT-1) (Namba *et. al.*, 1980), has previously been investigated (Parker *et. al.*, 1989). This study showed that the WI-38 and CT-1 fibroblasts produce similar quantities of the heterotrimeric type I collagen, while the SVWI-38 cells produce only 20-25% of the total collagen synthesised by their normal counterparts. Similar observations were reported by other investigators, who have also shown a decrease in collagen synthesis in various transformed cell lines (Hata and Peterkofsky, 1977 and Sandmeyer *et. al.*, 1981).

SDS-polyacrylamide gel electrophoresis and cyanogen bromide peptide analysis showed that the SVWI-38 fibroblasts produced only overmodified α1(I) chains, while no detectable α2(I) collagen chains were produced. DEAE-cellulose ion exchange chromatography identified the product as a homotrimer of type I collagen consisting of three overmodified α1(I) chains (Parker *et. al.*, 1989). Sundarraj and Church (1978) have also demonstrated an increase in the post-translational modification of type I procollagen in another SV40-transformed human fibroblast line (GM637).

Nothern blotting analysis revealed that α1(I) procollagen mRNA was present in SVWI-38 and WI-38 cells, while α2(I) transcripts could only be detected in WI-38 fibroblasts. Although SVWI-38 fibroblasts produce only 20%-25% of the collagen produced by their WI-38 counterparts, the α1(I)

procollagen mRNA levels in these cells were similar to those in WI-38 cells (Parker *et. al.*, 1989). If the translation rate of the α1(I) procollagen mRNA is normal in SVWI-38 fibroblasts then the excess α1(I) chains produced by these cells are probably degraded prior to homotrimer formation. Alternatively, an accumulation of free α1(I) procollagen chains, due to the slower rate of type I homotrimer formation, or an increase in degradation products of the excess α1(I) chains may play a role in feedback inhibition of translation of α1(I) procollagen mRNA. Bennett and Adams (1987) have demonstrated such a control mechanism for α2(I) procollagen mRNA translation in chicken vertebral chondroblasts.

Karyotype analysis revealed that SVWI-38 cells contain only 28 normal chromosomes while gross chromosomal aberrations were noted in the remaining 18 chromosomes. Chromosomes 7 and 17, which contain the α2(I) and α1(I) procollagen genes respectively, appeared to be normal (Junien *et. al.*, 1982; Huerre *et. al.*, 1982; Retief *et. al.*, 1985) . The SV40 genome is covalently integrated within a stretch of 21 Kb of the SVWI-38 fibroblast genome as a series of partial tandem repeats (Goldberg, Y.P. *et. al.*, 1992). Restriction mapping of the entire α2(I) procollagen gene and its promoter did not reveal any gross deletions, insertions, rearrangements or integration of large fragments of SV40 DNA within the body of the gene or its promoter (Parker *et. al.*, 1989). It is possible that the inactivation of this gene is due to the integration of a small fragment or fragments of SV40 DNA, which cannot be easily detected by restriction mapping. However, it is unlikely that both alleles of the same gene would be affected in the same way in the same cell.

Previous studies have shown that hypermethylation of the type I procollagen genes is associated with the inactivation of these genes (Parker *et. al.*, 1982; Smith and Marsilio, 1988; Guenette *et. al.*, 1992). The methylation of the α2(I) procollagen gene in WI-38 and SVWI-38 fibroblasts showed that the gene was methylated to the same extent in both cell types in both the 5'- and 3'-regions. Although Guenette *et. al.* (1992) have demonstrated that DNA hypermethylation is responsible for inactivation of the α2(I) procollagen gene in

a chemically transformed rat liver epithelial-like cell line, W8, which also secretes a type I collagen homotrimer, aberrant methylation is probably not responsible for inactivation of this gene in SVWI-38 fibroblasts.

Transient transfection studies indicated that the α2(I) promoter is inactive in SVWI-38 cells. Since the α2(I) procollagen gene in SVWI-38 fibroblasts is intact, a *trans*-acting factor or factors is most probably responsible for the repression of this gene in this cell line (Parker *et. al.*, 1989 and Smith, 1989). In an attempt to identify regions of the α2(I) procollagen gene to which these negative *trans*-acting factors bind, the activity of 5'-deletions of the wild type promoter was measured in transient transfection assays (Parker *et. al.*, 1992). The transfection data on CT-1 cells were similar to those obtained for normal fibroblasts (Boast *et. al.*, 1990), while the overall activity of all the constructs was much lower in SVWI-38 fibroblasts. Although several areas containing negative *cis*-acting elements were identified in the upstream sequences, the smallest construct containing the proximal promoter (-107 to +58) showed the greatest relative inhibitory effect in SVWI-38 cells.

This overall decrease in promoter activity in SVWI-38 cells could be titrated out when these cells were co-transfected with an excess α2(I) procollagen promoter (nucleotides -3800 to +58). A smaller fragment of the proximal promoter (nucleotides -107 to +58) containing an inverted CCAAT box and flanking sequences was also able to titrate out this inhibitory effect.

Since a negative *trans*-acting factor or factors are responsible for the inactivation of the α2(I) procollagen gene in SVWI-38 fibroblasts, the aim of this study was to identify, characterise and purify the *trans*-acting factors which regulate the expression of this gene in SVWI-38 and CT-1 fibroblasts.

Subcloning and sequencing of the SVWI-38 proximal α2(I) procollagen promoter and identification of protein complexes from a number of cell lines which bind to this region of the promoter is described in this chapter.

2.2 RESULTS

2.2.1 Characterisation of the SVWI-38 α2(I) proximal promoter

Extensive restriction mapping of the entire α2(I) procollagen gene and its promoter did not reveal any gross deletions, insertions, rearrangements or integration of large fragments of the SV40 DNA within this gene or its promoter (Parker *et. al.*, 1989). Secondly, the transfection experiments suggested that a *trans*-acting factor(s) which binds between nucleotides -107 to +58 of the α2(I) procollagen gene plays an important role in the inactivation of this gene in SVWI-38 fibroblasts (Parker *et. al.*, 1992). Although both these sets of experiments strongly suggest that *trans*-acting factors are responsible for the inactivation of this gene in SVWI-38 fibroblasts, the -351 bp proximal promoter of the gene from these cells was subcloned and sequenced to rule out any possibility that point mutations, small deletions or insertions in this functionally important region of the promoter could play a role in the inactivation of this gene.

2.2.1.1 Subcloning

Plasmids pUCT-1 and pUCT-2, containing a 4.2 Kb Eco R1 fragment of the α2(I) procollagen promoter isolated from SVWI-38 fibroblasts (Smith, 1989), were digested with Pst 1 and Eco R1 to release, amongst others, a 651 bp Pst 1 - Eco R1 fragment. The digested plasmids were resolved on a 1% low melting agarose gel and the 651 bp fragment eluted using the QIAEX extraction protocol as described in Materials and Methods (7.1.1.4). The eluted Pst 1 - Eco R1 fragments were cloned into pUC19 to produce plasmids pUCT-1-600 and pUCT-2-600 respectively (Figure 2.1).

These plasmids contain the -351 bp proximal promoter, the first exon and the 5'-portion of the first intron of the SVWI-38 α2(I) procollagen gene (nucleotides -351 to +300).

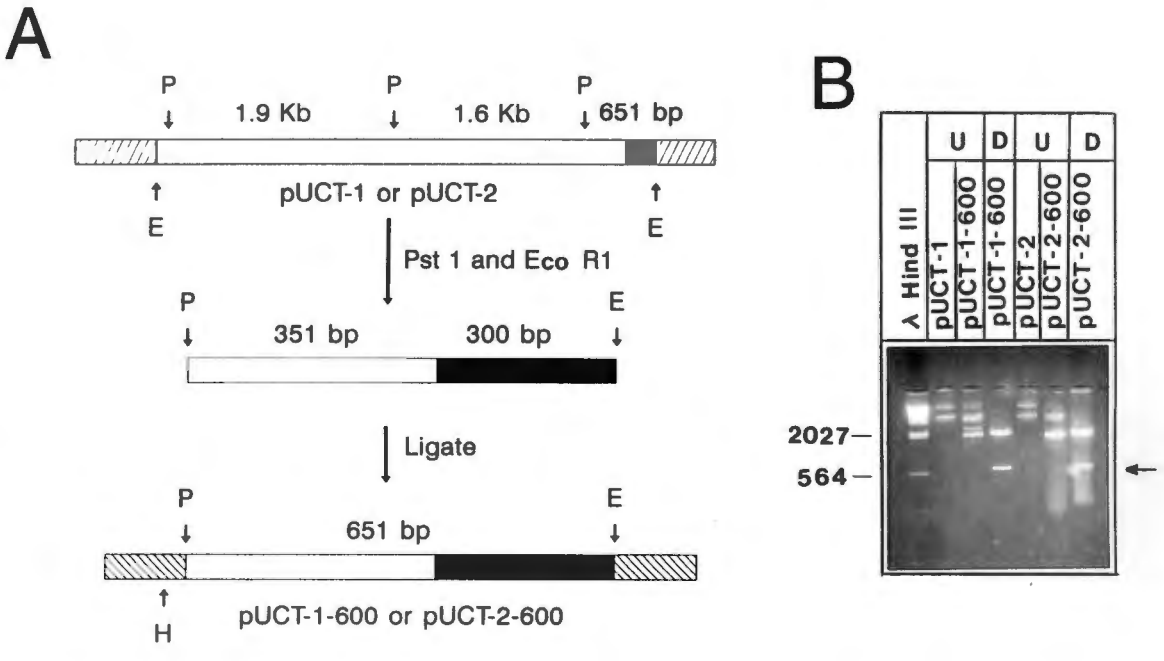


Fig 2.1 Subcloning of the α2(I) proximal promoter. (A) A schematic outline showing the subcloning of the 651 bp Pst 1 (P) - Eco R1 (E) fragment from plasmids pUCT-1 and pUCT-2 into pUC19 (refer to Materials and Methods section 2.2.1.1) to produce plasmids pUCT-1-600 and pUCT-2-600 respectively. The 651 bp fragment contains the proximal promoter up to -351 (clear box), exon 1 and the 5' portion of the first intron (shaded box). The polylinker of pUC19 is shown as a hatched box, while the Hind III site is indicated with a H. **(B)** An ethidium bromide stained 1% agarose gel containing undigested (U) plasmids pUCT-1, pUCT-1-600, pUCT-2 and pUCT-2-600 and Eco R1 and Hind III digested (D) pUCT-1-600 and pUCT-2-600. The 651 bp fragment produced by digestion is indicated by the arrow.

2.2.1.2 Sequence analysis

Plasmids pUCT-1-600 and pUCT-2-600 were sequenced using the Sanger dideoxy method as described in Materials and Methods (7.1.6). The nucleotide sequence of the -351 bp proximal promoter, exon 1 and the 5'-portion of intron 1 of the α2(I) procollagen gene from SVWI-38 fibroblasts is shown in figure 2.2. The sequences of plasmids pUCT-1-600 and pUCT-2-600 were identical to the sequence of the corresponding region of the gene from normal cell lines which produce the α2 chain (Dickson *et. al.*; 1985). This result supports the hypothesis that *trans*-acting factors, and not mutations in this functionally important region of the promoter, are responsible for inactivation of the α2(I) procollagen gene in SVWI-38 fibroblasts.

2.2.2 Candidate transcription factors which could regulate expression of the α2(I) procollagen gene

In order to identify candidate transcription factors which could regulate the expression of the human α2(I) procollagen gene, a database of transcription factors TFD sites (release 4.3) (Ghosh, 1990), was used to analyse the -351 bp proximal promoter of this gene.

A CCAAT box is located between nucleotides -201 and -195, while an inverted CCAAT box is present at the conventional distance between nucleotides -84 and -80. A total of three hexanucleotide Sp1 sequences are present in the -351 bp proximal promoter extending from nucleotides -305 to -298, -295 to -288, and -278 to -271. However, none of these Sp1 sequences are the true decanucleotide consensus sequences required for Sp1 binding (Kadonaga *et. al.*, 1986). A potential inverted AP-1 sequence and an AP-4 consensus sequence were identified between nucleotides -258 and -251 and nucleotides -123 and -114 respectively. Furthermore, four AP-2 sequences are situated between nucleotides -104 and -95, -115 and -108, -136 and -129, and -187 and -180, while two inverted AP-2 sequences occur between nucleotides -282 and -273, and -305 and -296. A TATA box consensus sequence is present between nucleotides -33 and -27.

```
          -350       -340       -330       -320       -310
        C TGCAGAGCAC TCCGACGTGT CCCATAGTGT TTCCAAACTT GGAAAGGGCG

-300       -290       -280       -270       -260       -250
GGGGAGGGCG GGAGGATGCG GAGGGCGGAG GTATGCAGAC AACGAGTCAG AGTTTCCCCT

-240       -230       -220       -210       -200       -190
TGAAAGCCTC AAAAGTGTCC ACGTCCTCAA AAAGAATGGA ACCAATTTAA GAAGCCAGCC

-180       -170       -160       -150       -140       -130
CCGTGGCCAC GTCCCTTCCC CCATTCGCTC CCTCCTCTGC GCCCCCGCAG GCTCCTCCCA

-120       -110       -100       -90        -80        -70
GCTGTGGCTG CCCGGGCCCC CAGCCCCAGC CCTCCCATTG GTGGAGGCCC TTTTGGAGGC

-60        -50        -40        -30        -20        -10
ACCCTAGGGC CAGGGAAACT TTTGCCGTAT AAATAGGGCA GATCCGGGCT TTATTATTTT

+1         +10        +20        +30        +40        +50        +60
AGCACCACGG CAGCAGGAGG TTTCGGCTAA GTTGGAGGTA CTGGCCACGA CTGCATGCCC

           +70        +80        +90        +100       +110       +120
GCGCCCGCCA GGTGATACCT CCGCCGGTGA CCCAGGGGCT CTGCGACACA AGGAGTCTGC

           +130       +140       +150       +160       +170       +180
ATGTCTAAGT GCTAGACATG CTCAGCTTTG TGGATACGCG GACTTTGTTG CTGCTTGCAG

           +190       +200       +210       +220       +230       +240
TAACCTTATG CCTAGCAACA TGCCAATgta agtgccttca gcttgtttgg gggagactgg

           +250       +260       +270       +280       +290       +300
gtagagaggt tagatgggag ggcaccctgc cctgaaaagg aaaacctgta acctgaattc
```

Fig 2.2 Nucleotide sequence of a 651 bp Pst 1 - Eco R1 fragment of plasmids pUCT-1-600 and pUCT-2-600. This fragment contains the -351 bp proximal promoter (uppercase letters), the first exon (bold) and the 5' portion of the first intron (lower case letters) of the α2(I) procollagen gene isolated from SVWI-38 fibroblasts. The 651 bp Pst 1 - Eco R1 fragment was subcloned into pUC19 as described in section 2.2.1.1 and sequenced using the universal M13 forward and reverse primers, oligonucleotide C1 (doubly underlined) and oligonucleotide D2 (underlined) as described in section 7.1.6.

In addition to the inverted CCAAT box, two other elements have been identified in the mouse proximal promoter as being important in the regulation of α2(I) procollagen gene expression (Karsenty *et. al.*, 1988). One of these sequences, a CAGA element situated between nucleotides -253 and -250 is also present in the human promoter. The other sequence, an inverted NF-1-like binding site ($^{5'}$TCGCCCTTGCCAA$^{3'}$), is present in the human promoter between nucleotides -311 and -299, but with a two-base mismatch ($^{5'}$CCGCCCTTTCCAA$^{3'}$). NF-1 binds to this sequence in the mouse promoter to activate transcription (Oikarinen *et. al.*, 1987).

The result of this analysis is tabulated in appendix A and summarised in figure 2.3.

2.2.3 Subcloning the proximal human α2(I) procollagen promoter

In order to identify, characterise and purify the *trans*-acting factors which regulate the expression of the human α2(I) procollagen gene, several DNA fragments of the α2(I) procollagen gene promoter were subcloned and used in DNA-binding assays.

Plasmids pNJ-59, pNJ-230 and pNJ-170 were subcloned from either plasmids pNJs400 19A or pNJs400 19B, containing a 409 bp fragment cloned into pUC19 in both orientations as shown in figure 2.4 (Smith, 1989).

2.2.3.1 Plasmid pNJ-59

Plasmid pNJs400 19B was digested with Sma 1 to release a 191 bp Fragment (nucleotides -107 to +58) which was eluted from a 7% non-denaturing polyacrylamide gel as described in Materials and Methods (7.1.1.1) and digested with Bst N1 to produce a 59 bp Sma 1 - Bst N1 and a 132 bp Bst N1 - Sma 1 fragments. The 59 bp fragment was subcloned into the Sma 1 site of pUC19 after blunt ending with Klenow DNA polymerase (section 7.1.2.1) to produce plasmid pNJ-59 (Figure 2.4). This plasmid contains a 65 bp fragment (-110 to -46) of the human α2(I) procollagen promoter.

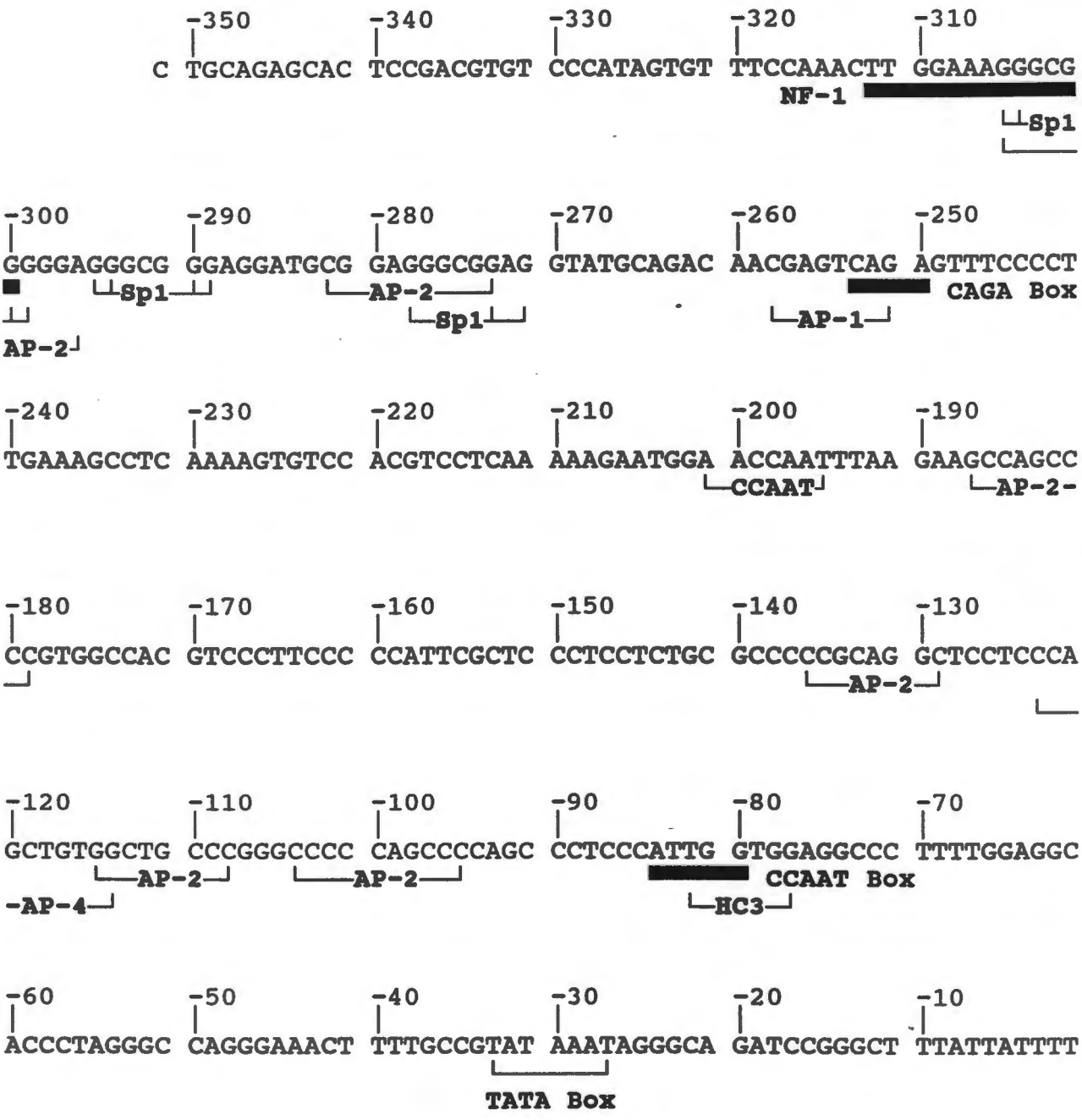
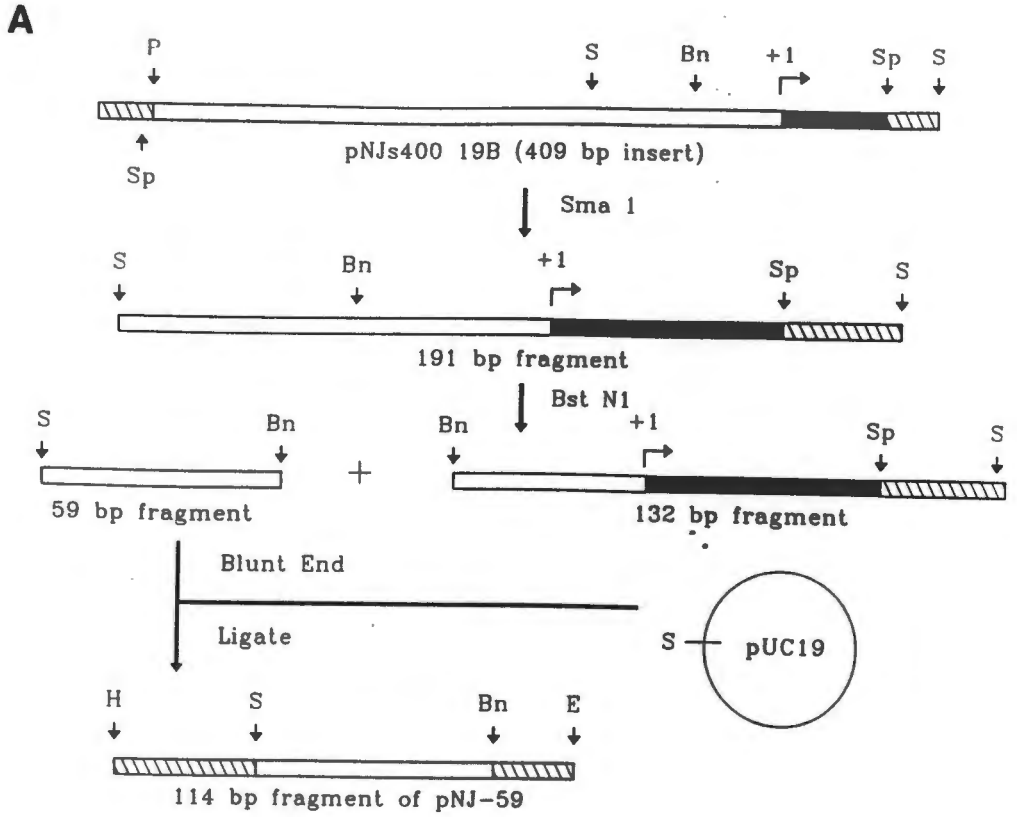


Fig 2.3 A summary of the major transcription factor binding motifs (underlined ┗━━┛), in the -351 bp proximal promoter of the α2(I) procollagen gene. These sequences were identified by searching a database of transcription factors tabulated in appendix A. The DNA sequence elements identified in the mouse gene as being important in regulation of this gene are shown in bold (▬▬).



B

-110 -100

aagcttgcat gcctgcaggt cgactctaga ggatcCCCGG GCCCCCAGCC

Hind III

Sma 1

-90 -80 -70 -60 -50 -46

CCAGCCCTCC CATTGGTGGA GGCCCTTTTG GAGGCACCCT AGGGCCAGGG

Bst N1

taccgagctc gaattc

Eco R1

Plasmid pNJ-59

Figure 2.4 Subcloning plasmid pNJ-59. **(A)** A schematic diagram of the subcloning of the 59 bp Sma 1 (S) - Bst N1 (Bn) fragment from plasmid pNJs400 19B into the Sma 1 site of pUC19 to produce plasmid pNJ-59, as described in section 2.2.3.1. The α2(I) procollagen promoter, the 5' portion of exon 1 and the polylinker of pUC19 are shown as clear boxes, shaded boxes and hatched boxes respectively. The restriction sites Pvu II (P), Sph 1 (Sp), Hind III (H) and Eco R1 (E) are indicated with arrows. **(B)** Nucleotide sequence of the 116 bp Hind III - Eco R1 fragment of plasmid pNJ-59, containing the polylinker of pUC19 (lowercase letters) and a 65 bp fragment of the promoter from nucleotides -110 to -46 (uppercase letters).

2.2.3.2 Plasmid pNJ-230

pNJs400 19A was linearised with Bam H1 and digested with Bst X1 to remove a 193 bp fragment containing 166 bp of the promoter (-351 to -180) (A. Smith, unpublished data). The fragments were separated on a 1% agarose gel, the larger fragment (vector and -179 to +58) was electro-eluted from the gel, blunt ended with Klenow DNA polymerase as described in Materials and Methods (7.1.1.2 and 7.1.2.1) and self ligated to produce plasmid pNJ-230 containing 58 bp of exon 1 and 179 bp of the collagen proximal promoter (nucleotides -179 to +58) (Figure 2.5).

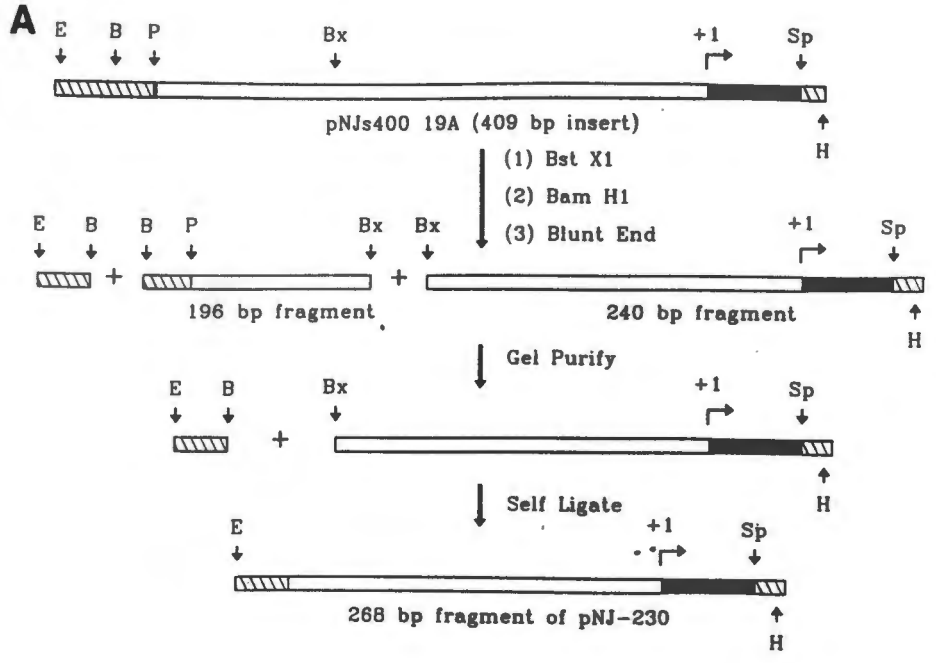
2.2.3.3 Plasmid pNJ-170

pNJs400 19B was digested with Bam H1 and Bst X1. The Bam H1 and Bst X1 sticky ends were blunt ended with T4 DNA polymerase as described (section 7.1.2.2). The two fragments were resolved on a 1% agarose gel and the larger fragment was eluted according to the method of Errington (1990) as described in Materials and Methods (7.1.1.3). This fragment was self ligated to produce plasmid pNJ-170 containing a 169 bp Pst 1 - Bst X1 (nucleotides -351 to -183) fragment of the human α2(I) procollagen promoter (figure 2.6).

2.2.4 Analysis of DNA-Protein interaction by electrophoretic mobility shift assays (EMSA)

EMSA (Fried and Crothes, 1981) was used to identify DNA-protein complex formation from crude nuclear extracts of SVWI-38, CT-1 and WI-38 fibroblasts and of a non-collagen producing mouse tumour macrophage cell line, $p388D_1$ (Koren *et. al.*, 1975). This simple, rapid and sensitive technique proved to be ideal for the initial identification of DNA-protein complexes (section 7.4.2).

The methods of Dignam *et. al.* (1983) or Lee and Green (1990), as described in Materials and Methods (7.3), were used to isolate nuclear proteins from SVWI-38, CT-1, WI-38 and $p388D_1$ cells. The integrity of these proteins were checked on 10%

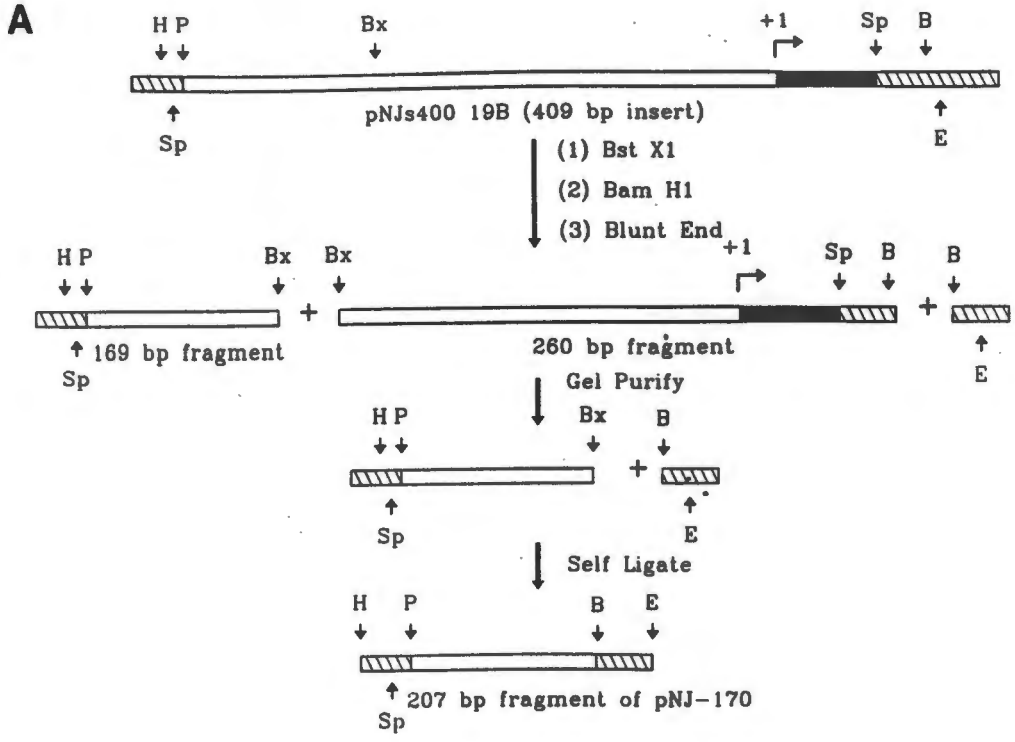


B

```
                                                    -179          -170          -160
gaattcgagc tcggtacccg gggatCGTGG CCACGTCCCT TCCCCCATTC
ECo R1              Sma 1

          -150          -140          -130          -120          -110
GCTCCCTCCT CTGCGCCCCC GCAGGCTCCT CCCAGCTGTG GCTGCCCGGG
                                                        Sma 1

          -100          -90           -80           -70           -60
CCCCCAGCCC CAGCCCTCCC ATTGGTGGAG GCCCTTTTGG AGGCACCCTA

          -50           -40           -30           -20           -10
GGGCCAGGGA AACTTTTGCC GTATAAATAG GCAGATCCG GGCTTTATTA

          +1            +10           +20           +30           +40
TTTTAGCACC ACGGCAGCAG GAGGTTTCGG CTAAGTTGGA GGTACTGGCC

          +50           +58
ACGACTGCAT GCaagctt
           Sph 1 Hind III
```

Plasmid pNJ-230

Figure 2.5 Subcloning plasmid pNJ-230. (A) A schematic diagram of the subcloning of a 237 bp BstX1 (Bx) - Sph1 (Sp) fragment of plasmid pNJs400 19A into pUC19 to produce plasmid pNJ-230, as described in section 2.2.3.2. The α2(I) procollagen promoter, the 5' portion of exon 1 and the pUC19 polylinker are shown as clear, shaded and hatched boxes respectively. The restriction sites Hind III (H), Eco R1 (E), Pvu II (P) and Bam H1 (B) are indicated. **(B)** Nucleotide sequence of a 268 bp Eco R1 - Hind III fragment of plasmid pNJ-230 containing the pUC19 polylinker (lowercase letters), the -179 bp proximal promoter (uppercase letters) and the 5'-portion of the first exon (underlined).



B

-351 -340 -330 -320

aagcttgcat gcCTGCAGAG CACTCCGACG TGTCCCATAG TGTTTCCAAA

Hind III Sph 1 Pst 1

-310 -300 -290 -280 -270

CTTGGAAAGG GCGGGGGAGG GCGGGAGGAT GCGGAGGGCG GAGGTATGCA

-260 -250 -240 -230 -220

GACAACGAGT CAGAGTTTCC CCTTGAAAGC CTCAAAAGTG TCCACGTCCT

-210 -200 -190 -183

CAAAAAGAAT GGAACCAATT TAAGAAGCCA Ggatccccgg gtaccgagct

Bam H1

cgaattc

ECo R1

Plasmid pNJ-170

Figure 2.6 Subcloning plasmid pNJ-170. **(A)** A schematic diagram of the subcloning of a 169 bp Pst 1 (P) - Bst X1 (Bx) fragment from plasmid pNJs400 19B into pUC19 to produce plasmid pNJ-170, as described in section 2.2.3.3. The α2(I) procollagen promoter, the 5' portion of exon 1 and the pUC19 polylinker are shown as clear, shaded and hatched boxes respectively. The restriction sites Sph 1 (Sp), Hind III (H), Eco R1 (E) and Bam H1 (B) are indicated with arrows. **(B)** Nucleotide sequence of a 207 bp Hind III - Eco R1 fragment of plasmid pNJ-170 containing the pUC19 polylinker (lowercase letters) and a 169 bp fragment of the promoter from nucleotides -351 to -183 (uppercase letters).

SDS-polyacrylamide gels (figure 2.7) prior to use in EMSA. Similar distribution of polypeptides was observed in crude nuclear extracts, indicating that minimal degradation of the proteins had occurred during the extraction procedure.

The 114 bp Eco R1 - Hind III fragment of pNJ-59 (fragment E), the 203 bp Eco R1 - Hind III fragment of pNJ-170 (fragment A), the 78 bp Sma 1 fragment of pNJ-230 (fragment B), the 225 bp Hha 1 - Eco R1 fragment of pNJs400 19B (fragment C) and the 190 bp Sma 1 fragment of pNJs400 19B (fragment D) were radioactively labelled as described in Materials and Methods (7.4.1) and used as probes in the EMSA (Figure 2.8).

Similar DNA-protein complexes were observed when SVWI-38 and CT-1 nuclear extracts were allowed to bind to DNA fragments A and B as shown in figure 2.8. Occasionally, an extra complex was observed when CT-1 nuclear extracts were assayed with fragment B (figure 2.8).

Two major DNA-protein complexes, designated complexes I and III, were identified in CT-1 nuclear extracts complexed with fragments D (-107 to +58) or E (-110 to -46), while an additional complex, complex II, was observed in SVWI-38 nuclear extracts (figure 2.8). This SVWI-38 specific complex was occasionally observed in longer exposures of CT-1 autoradiograms, suggesting that low levels of this protein are present in collagen producing cells. Complex III was also present at lower levels in SVWI-38 than in CT-1 cells.

When fragment C (-140 to +58) was used with SVWI-38 and CT-1 nuclear extracts, an extra complex was still observed in SVWI-38 nuclear extracts (figure 2.8). This fragment was tested in order to rule out the possibility that the AP-2 consensus sequence which was destroyed upon digestion with Sma 1 might be involved in formation of other complexes. The remaining complexes were similar in SVWI-38 and CT-1 extracts. It is therefore unlikely that this AP-2 site is responsible for the formation of any extra complexes.

The transient transfection experiments strongly suggested that a *trans*-acting factor(s), which binds to nucleotides

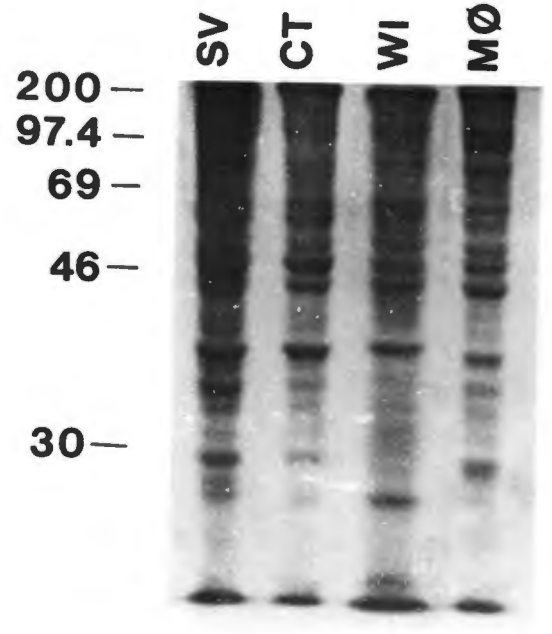


Fig 2.7 SDS-polyacrylamide gel electrophoresis of crude nuclear extracts. Nuclear extracts of SVWI-38 (SV), CT-1 (CT), WI-38 (WI) and p388D_1, (Mφ) cell lines were prepared as described in section 8.3. 40 ug of nuclear extracts were resolved on 10% SDS-polyacrylamide gels and stained with Coomassie brilliant blue, as described in section 7.5.

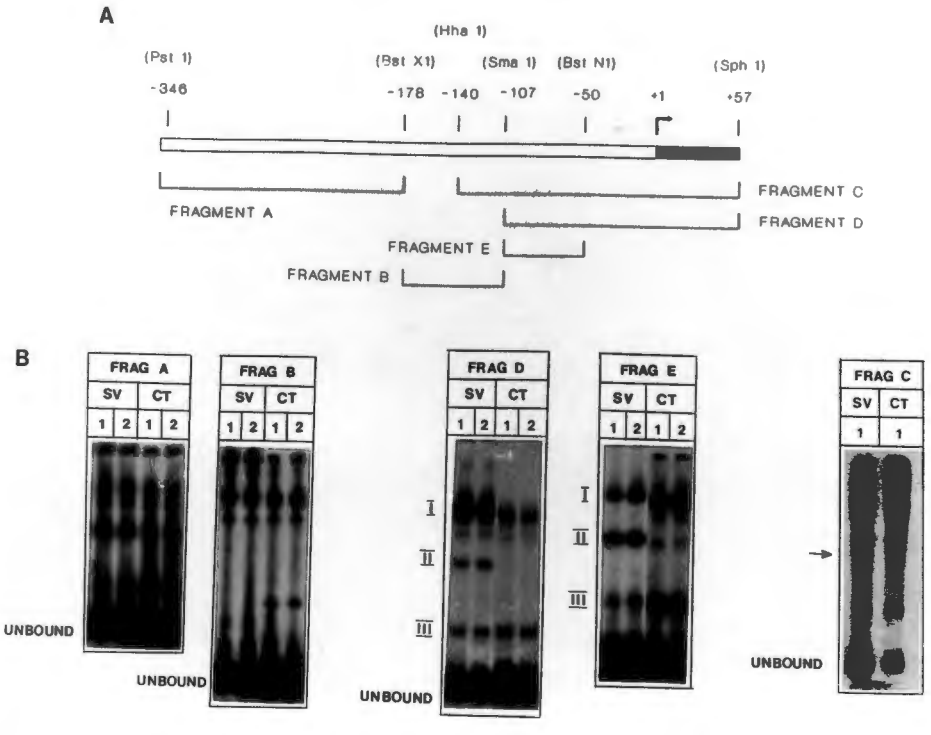


Fig 2.8 Dissection of the α2(I) promoter for DNA-binding activity. **(A)** A schematic diagram of the proximal promoter (clear box) and the 5' portion of exon 1 (shaded box) of the α2(I) procollagen gene. DNA fragments A, B, C, D and E were generated by digesting the promoter with the appropriate restriction enzymes as described in section 2.2.4. **(B)** DNA-protein complexes, were analysed in duplicate by electrophoretic mobility shift assays (lanes 1 and 2) by incubating the indicated end-labelled DNA fragment with 3-4 ug nuclear extract and analysis of the complexes on 5% non-denaturing polyacrylamide gels as described in section 7.4.2. The dried gels were exposed to X-ray film for 16 hrs. The DNA complexes formed with fragments D (nucleotides -107 to +57) and E (nucleotides -107 to -46) are labelled as I, II and III, with complex II being largely SVWI-38 specific. The autoradiograph for fragment E was overexposed in order to show the trace amounts of complex II in CT-1 cells. The additional complex identified in SVWI-38 extracts with fragment C (nucleotides -140 to +57) is indicated by the arrow. The additional complex observed in CT-1 nuclear extracts with fragment B was not consistent.

-107 to +58 of the proximal promoter is responsible for inactivation of the α2(I) procollagen gene in SVWI-38 fibroblasts (Parker *et. al.*, 1992). The results of the EMSAs clearly showed the presence of an extra DNA-protein complex (complex II) in the nuclear extracts of SVWI-38 fibroblasts, which could bind to an inhibitory element between nucleotides -110 to -46 of this gene. It is therefore possible that complex II, together with other factors, is responsible for inhibition of the α2(I) procollagen gene in SVWI-38 fibroblasts. The low levels of complex II in the collagen producing CT-1 cell line suggests that this complex probably plays an important general inhibitory role in the regulation of this gene.

In order to determine whether there was any correlation between the levels of complex II in nuclear extracts and the ability of a cell line to synthesise type I collagen, the presence of the three complexes in nuclear extracts of WI-38 fibroblasts and a non-collagen producing mouse tumour macrophage cell line, $p388D_1$, was assayed. Complexes I and III were present in the extracts of collagen-producing WI-38 cells, while only complexes I and II were identified in $P388D_1$ extracts (figure 2.9). This result also strongly supports the proposal that complex II is a negative *trans*-acting factor which may be responsible for the inactivation or modulation of the α2(I) procollagen gene.

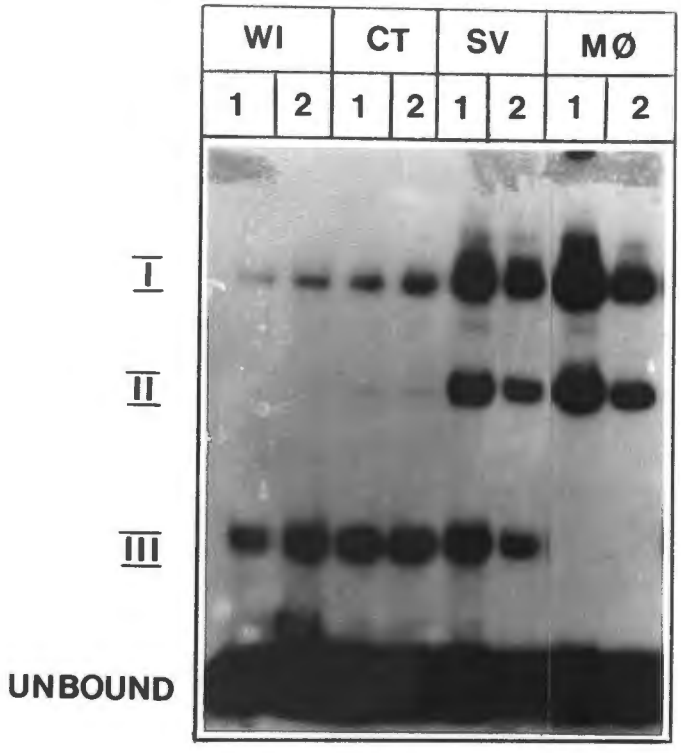


Fig 2.9 The identification of complexes I, II and II in WI-38 (WI), CT-1 (CT), SVWI-38 (SV) and a mouse tumour macrophage, p388D$_1$, (Mφ) cell lines. 3-4 ug of crude nuclear extracts were incubated with 4 ug poly dIdC.poly dIdC and 1 ng (10^4 cpm) of end-labelled fragment E (nucleotides -107 to -46) of the α2(I) procollagen promoter as described in section 7.4. Duplicate samples (lanes 1 and 2) were electrophoresed on 5% non-denaturing polyacrylamide gels at 4^{O} C in 0.5 X TBE. The dried gels were exposed to X-ray film for 16 hrs. The position of complexes I, II and III are indicated.

2.3 DISCUSSION

Although only the -351 to +300 bp region of the SVWI-38 α2(I) procollagen gene was sequenced and shown to be completely normal, it is unlikely that mutations in this gene or its promoter are responsible for its inactivation in SVWI-38 fibroblasts. This hypothesis is supported by the work of Boast *et. al.* (1990), Niederreither *et. al.* (1992) and Goldberg, H. *et. al.* (1992) who have demonstrated that the minimal regulatory elements sufficient for tissue and cell-specific expression of the human and mouse α2(I) procollagen genes lie within the -351 bp proximal promoter. Secondly, 86% sequence homology exists between the mouse and human -351 bp proximal promoters of this gene (Dickson *et. al.*, 1985). This high sequence homology between the two promoters was lost further upstream (M.I. Parker, unpublished data), suggesting that the proximal region of the promoter is probably functionally more important. Finally, transient transfection experiments showed that the α2(I) procollagen gene promoter from SVWI-38 fibroblasts is functional in normal collagen producing cell lines (Smith, 1989).

The results shown in figures 2.8 and 2.9 suggested that the levels of complex II is indirectly proportional to the amount of type I collagen produced. Also, cells that do not synthesise α2(I) procollagen chains appeared to have no detectable amounts (macrophages) or lower levels (SVWI-38) of complex III than collagen producing cells. These observations suggested that the relative ratios of complexes II and III, and not merely the absence or presence of complex II, may be important in determining whether the α2(I) procollagen gene is expressed or not. Barnhart *et. al.* (1988) have shown that the relative activities of three *trans*-acting factors, α-CP1, α-CP2 and α-IRP changes when the human α-globin gene is expressed.

Three potential DNA sequences recognised by sequence-specific transcription factors were identified in this 65 bp region (nucleotides -110 to -46) of the α2(I) procollagen promoter. An inverted CCAAT box is situated between nucleotides -84 to -80 and an inverted HC3 sequence between nucleotides -82 and -77 (Augereau and Chambon, 1986) and a potential AP-2

consensus sequence between nucleotides -104 and -95 (Imagawa *et. al.*, 1987). The importance of these sequences in the regulation of the gene will be discussed in chapter 4.

Although Hatamochi *et. al.* (1988) and Maity *et. al.* (1988) have identified only one complex, a CCAAT binding factor (CBF), in mouse NIH-3T3 nuclear extracts which binds to the -108 to +54 sequences of the mouse α2(I) procollagen promoter, our studies demonstrated the formation of both complexes I and II when mouse macrophage or fibroblast nuclear extracts were used in EMSA (V. Leaner, unpublished data). This finding appears to be inconsistent with the hypothesis that complex II proteins are a transcriptional repressor, since mouse fibroblasts do synthesis α2(I) collagen. The significance of this finding will be discussed in chapter 4.

Goldberg, H. *et. al.* (1992) have postulated that a CTF/NF-1 binding element situated around -300 ($^{5'}TCGN_5GCCAA^{3'}$) in the mouse α2(I) procollagen promoter may play a role in the cell specific expression of this gene. They have shown that collagen producing murine cells have higher levels of NF-1 activity than cells which do not produce collagen. This inverted NF-1 like element is also present in the human promoter ($^{5'}CCGN_5TCCAA^{3'}$) between nucleotides -311 and -299, but there it contains base mismatches, making it less similar than the mouse NF-1 site to the published NF-1 recognition sequence ($^{5'}TGGN_{6-7}GCCAA^{3'}$) (Jones *et. al.*, 1987). It is not known how this diversity in the sequence of the human NF-1 site affects the role of NF-1 in the regulation of the human α2(I) procollagen gene. In contrast, it is possible that the following may modulate the cell-specific expression of the human gene: (1) binding of a negative *trans*-acting factor to the proximal/minimal promoter, (2) the relative ratios of complex II and III, and (3) other transcription factors, like NF-1, which bind further upstream.

CHAPTER 3

THE PARTIAL PURIFICATION OF FIBROBLAST DNA-BINDING PROTEINS

3.1 INTRODUCTION

Eukaryotic gene expression is mediated in part by sequence-specific DNA-binding proteins that bind to promoters and enhancers (Maniatis *et. al.*, 1987; Mitchell and Tjian, 1989). The identification and characterisation of these proteins is an important step towards understanding how they regulate transcription. In order to characterise these low-abundance transcription factors, it is necessary to purify these proteins from nuclear extracts. A combination of sequence-specific DNA-affinity chromatography and conventional chromatographic techniques has made it possible to purify these proteins to homogeneity (Kadonaga and Tjian, 1986).

The purification of transcription factors, like the CCAAT binding factor (CBF) (Hatamochi *et. al.*, 1988), has made it possible to determine the biochemical properties of these proteins. The polypeptide composition, molecular weight, binding specificity and other parameters have been determined for a large number of transcription factors (reviewed in Jones *et. al.*, 1988). Partial protein sequence analysis of a number of purified factors has been successfully used to clone the genes for these proteins. Examples of cloned factors include Sp1 (Kadonaga *et. al., 1987;* Kadonaga *et. al.*, 1988) and TFIIE (Ohkuma *et. al.*, 1991; Hideki *et. al.*, 1991). The structural features of both these factors have been determined from the nucleotide and predicted amino-acid sequence of the cDNAs that encode these proteins.

Before DNA-binding proteins can be successfully purified by DNA-affinity chromatography, the proteins should be partially purified using other conventional chromatographic techniques

to remove nucleases and other activities that might affect the yield and activity. Ion exchange chromatography such as Heparin-agarose (Davison *et. al.*, 1979) and phosphocellulose (Dynan and Tjian, 1983) or gel filtration on Sephacryl S-300 (Cohen *et. al.*, 1986) are a few of the many techniques which have been routinely used in the partial purification of transcription factors.

The partial purification of complex I, II and III proteins from either SVWI-38 or CT-1 cells is described in this chapter.

3.2 RESULTS

3.2.1 Differential salt extraction of nuclear proteins

SVWI-38 nuclei were differentially extracted with either 0.1, 0.2, 0.3 or 0.4 M NaCl as described in Materials and Methods (7.3.3), to determine whether it was possible to differentially solubilise complex I, II and III proteins during the preparation of nuclear extracts. The electrophoretic mobility shift assay (EMSA) (section 7.4.2) was used to test for the presence of DNA-binding activity in the different fractions with fragment E (see figure 2.8) of the α2(I) procollagen promoter. Trace amounts of complex I protein was extracted in Dignam buffer C containing 0.1 M NaCl (figure 3.1). Complex II and III proteins started to be co-extracted with complex I proteins in buffer containing 0.2 M NaCl, but all three complexes were still co-extracted in buffer containing 0.4 M NaCl. Although complex I was extracted with a slightly lower salt concentration, the yields would be very low and it would therefore not be useful to separate the complexes this way during the preparation of nuclear extracts.

3.2.2 Heparin-agarose chromatography

Heparin-agarose cation-exchange chromatography (Davison *et. al.*, 1979) was used to separate and partially purify the DNA-binding complexes (complexes I, II and III) from SVWI-38 and CT-1 extracts. The heparin-agarose columns were packed and prepared as described in section 7.3.4. The method of Dignam *et. al.* (1983) was used to prepare milligram amounts (±30 mg) of crude nuclear extracts from SVWI-38 and CT-1 fibroblasts, as described in Materials and Methods (7.3.1).

In order to determine the optimal salt concentrations required for binding of the three complexes to heparin-agarose and their subsequent elution from the matrix, 3 mg of crude SVWI-38 nuclear extract was applied to a 1 ml column. The column was washed with 7 ml of 0.1 M CB (see section 7.7 for composition) to remove the bulk of the protein. EMSA showed no detectable amounts of complex I, II or III DNA

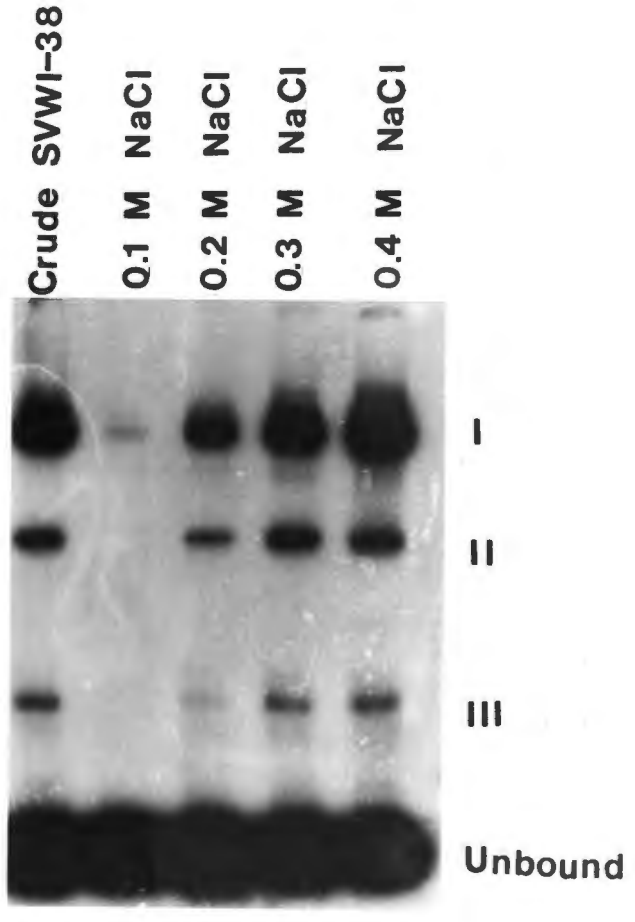


Fig 3.1 Differential salt extraction of nuclear proteins. SVWI-38 fibroblast nuclei were sequentially extracted with buffers containing 0.1 M, 0.2 M, 0.3 M and 0.4 M NaCl as described in Materials and Methods (7.3.3). 3-4 ug of crude extract of the differentially solubilised fractions were incubated with 4 ug poly dIdC.poly dIdC and 1 ng (10^4 cpm) of end-labelled fragment E (see figure 2.8) as described in section 7.4. The samples were electrophoresed on 5% non-denaturing polyacrylamide gels at 4^{o} C using 0.5X TBE and the dried gels exposed to X-ray film for 16 hrs. The position of complexes I, II and III are indicated.

binding activity in the 0.1 M CB wash (figure 3.2). Complex I proteins eluted from the column with 0.2 M KCl, while the 0.3 M KCl fraction contained complex I, II and III proteins.

In order to achieve separation of complex I proteins from complex III proteins, CT-1 protein was eluted from the heparin-agarose column with buffer CB containing 0.25 M KCl. The remaining bound protein was eluted stepwise with 0.4 M and 0.5 M KCl in CB and regenerated with 1.0 M KCl in CB. The 0.1 M, 0.25 M and 0.4 M heparin-agarose fractions were concentrated using centricon-10 microconcentrators, the salt concentration adjusted to 0.1 M KCl and assayed by EMSA using fragment E as a probe. Complex I and III proteins co-eluted in 0.25 M KCl (figure 3.3). This batch of CT-1 extract also contained trace amounts of complex II which eluted from the column together with complexes I and III. Complex II band appears to be strong because of overexposure of the gel.

These preliminary experiments (figures 3.2 and 3.3) suggested that it would be possible to separate complex I from complexes II and III using this technique if the bound protein was initially eluted from the column with 0.2 M KCl and perhaps washed long enough to remove all complex I proteins (figure 3.2).

Complex I, II and III proteins from SVWI-38 extracts and complex I and III proteins from CT-1 extracts were partially separated and purified as described in section 7.3.4. As shown in figures 3.4 and 3.5 this approach proved to be satisfactory for the partial purification and separation of complex I proteins from complex II and III proteins.

As illustrated in tables 3.1 and 3.2, 60 to 65% of the nuclear extract did not bind to heparin-agarose in 0.1 M KCl. No detectable amounts of complex I, II and III proteins were present in this fraction (figures 3.4 and 3.5). About 20% of the loaded protein eluted from the column in the 0.2 M KCl buffer and contained complex I proteins. The 0.4 M KCl buffer was used to elute complex II and III proteins from the column and contained about 10 to 15% of the loaded protein. Assuming 100% recovery of proteins from the column, one would

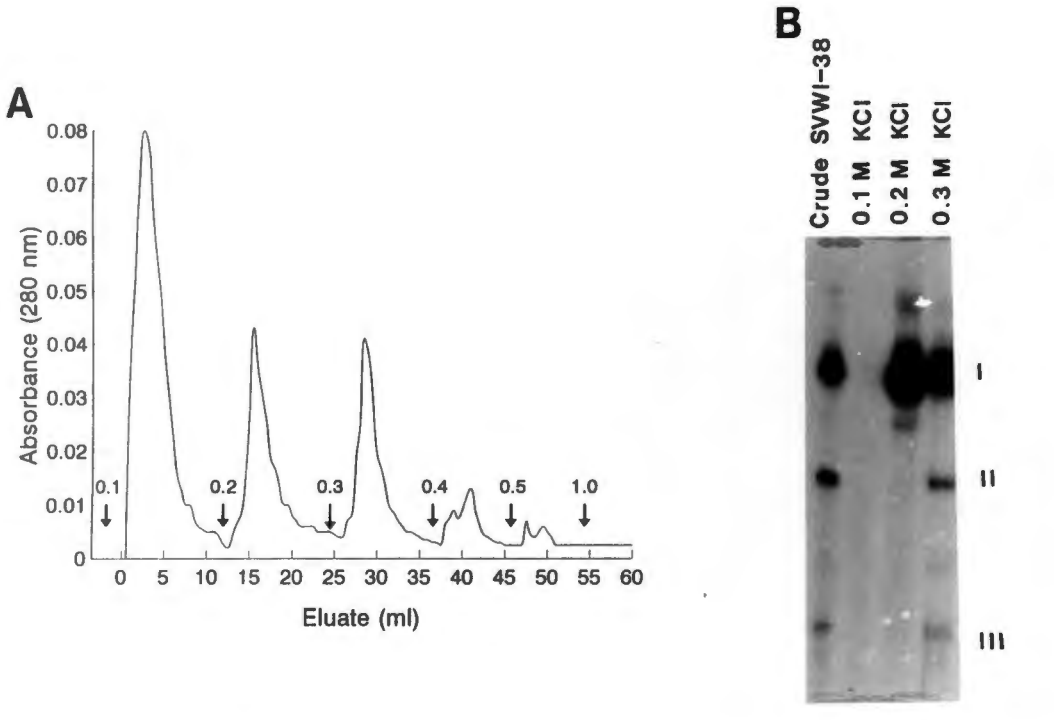


Fig 3.2 Fractionation of SVWI-38 extract on heparin-agarose. **(A)** 3 mg of nuclear extract was applied to a 1 ml column in buffer CB containing 0.1 M KCl. The bound proteins were eluted stepwise with buffer CB containing 0.2 M, 0.3 M, 0.4 M, 0.5 M or 1.0 M KCl as described in section 3.2.2. **(B)** 3-4 ug crude SVWI-38 and concentrated 0.1 M, 0.2 M, 0.3 M KCl heparin-agarose fractions were incubated with 4 ug poly dIdC.poly dIdC and 1 ng (10^4 cpm) of end-labelled fragment E as described in section 7.4. The samples were electrophoresed on 5% non-denaturing polyacrylamide gels at 4^{o} C using 0.5X TBE and the dried gels exposed to X-ray film for 16 hrs. The position of complexes I, II and III are indicated.

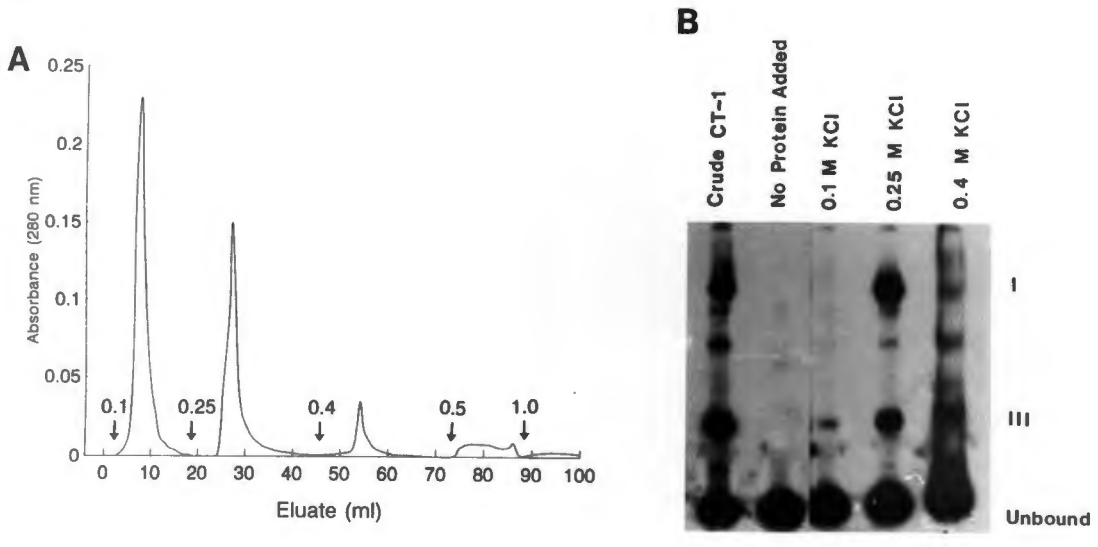


Fig 3.3 Elution of CT-1 extract from heparin-agarose. **(A)** Crude nuclear extract was applied to a 1 ml column in buffer CB containing 0.1 M KCl. The bound protein was eluted stepwise with buffer CB containing 0.25 M, 0.4 M, 0.5 M and 1.0 M KCl as described in section 3.2.2. **(B)** 3-4 ug crude CT-1 and concentrated 0.1 M, 0.25 M, 0.4 M KCl heparin-agarose fractions were incubated with 4 ug poly dIdC.poly dIdC and 1 ng (10^4 cpm) of end-labelled fragment E as described in section 7.4. The samples were electrophoresed on 5% non-denaturing polyacrylamide gels at 4^{O} C using 0.5X TBE and the dried gels were exposed to X-ray film for 16 hrs. The position of complexes I and III are indicated.

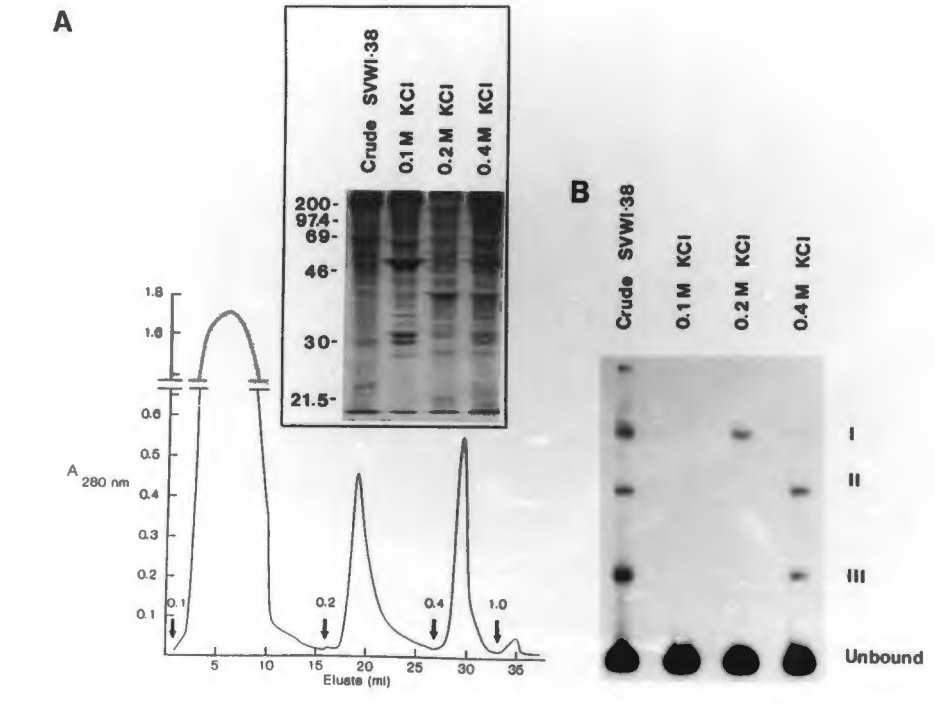


Fig 3.4 Partial purification of SVWI-38 extracts by heparin-agarose chromatography. **(A)** Crude nuclear extracts were applied to the column in buffer CB containing 0.1 M KCl. The bound protein was eluted from the column with buffer CB containing 0.2 M, 0.4 M or 1.0 M KCl as described in section 7.3.4. **(B)** The fractions were concentrated using centricon-10 microconcentrators and the final salt concentration adjusted to 0.1 M KCl. The electrophoretic mobility shift assay was used to assay the various heparin-agarose fractions for DNA-binding activity as described in section 7.4.2. The positions of complexes I, II and III are indicated. **(INSET)** 20 ug of crude nuclear extract and heparin-agarose fractions were resolved on 10% SDS-polyacrylamide gels and stained with Coomassie brilliant blue as described in section 7.5.

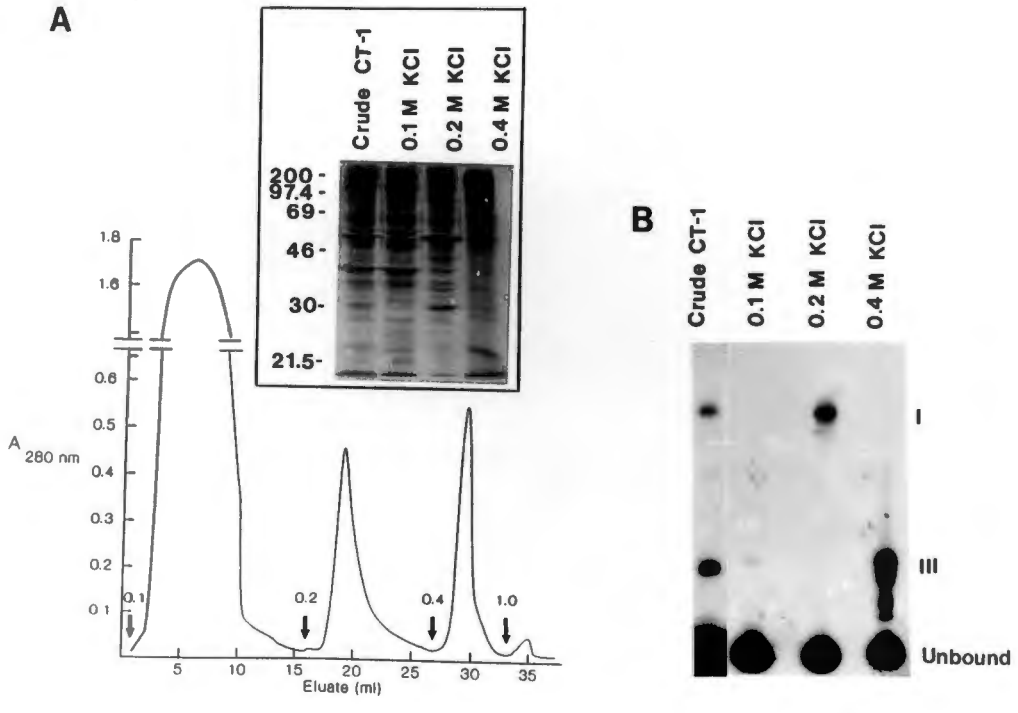


Fig 3.5 Partial purification of CT-1 extracts by heparin-agarose chromatography. **(A)** Crude nuclear extracts were applied to the column in buffer CB containing 0.1 M KCl. The bound protein was eluted from the column with buffer CB containing 0.2 M, 0.4 M or 1.0 M KCl as described in section 7.3.4. **(B)** The fractions were concentrated using centricon-10 microconcentrators and the final salt concentration adjusted to 0.1 M KCl. The electrophoretic mobility shift assay was used to assay the various heparin-agarose fractions for DNA-binding activity as described in section 7.4.2. The positions of complexes I and III are indicated. **(INSET)** 20 ug of crude nuclear extract and heparin-agarose fractions were resolved on 10% SDS-polyacrylamide gels and stained with Coomassie brilliant blue as described in section 7.5.

TABLE 3.1 Purification summary of SVWI-38 nuclear proteins. Heparin-agarose chromatography was used to partially purify crude nuclear extracts as described in section 7.3.4. The fractions were concentrated and the salt concentration adjusted to 0.1 KCl. Data shown was obtained from one of the numerous fractionations.

STAGE OF PURIFICATION	FRACTION	VOLUME (ml)	PROTEIN CONC. (ug/ul)	TOTAL PROTEIN (mg)	PROT. YIELD
1	Crude SVWI-38 nuclear extract	2.0	5.90	11.8	100%
2a	Heparin-agarose 0.2 M KCl fraction	4.8	0.56	2.7	23%
2b	Heparin-agarose 0.4 M KCl fraction	3.4	0.49	1.7	14%
3a	Centricon conc. 0.2 M KCl fraction	0.6	3.63	2.2	19%
3b	Centricon conc. 0.4 M KCl fraction	0.3	3.62	1.1	9%

TABLE 3.2 Purification summary of CT-1 nuclear proteins. Heparin-agarose chromatography was used to partially purify crude nuclear extracts as described for SVWI-38 cells in table 3.1. Data shown is representative of one of the numerous fractionations.

STAGE OF PURIFICATION	FRACTION	VOLUME (ml)	PROTEIN CONC. (ug/ul)	TOTAL PROTEIN (mg)	PROT. YIELD
1	Crude CT-1 nuclear extract	1.5	4.54	6.8	100%
2a	Heparin-agarose 0.2 M KCl fraction	4.4	0.32	1.4	21%
2b	Heparin-agarose 0.4 M KCl fraction	3.5	0.26	0.9	13%
3a	Centricon conc. 0.2 M KCl fraction	0.4	1.74	0.7	10%
3b	Centricon conc. 0.4 M KCl fraction	0.3	2.41	0.7	10%

have expected to obtain about a 5-fold purification of complex I and a 7 to 10-fold purification of complex II and III proteins from crude nuclear extracts. No more than a 2, 3 and 5-fold purification of complex I, II and III proteins, respectively, was obtained. Since up to 50% of protein was lost while concentrating the fractions on centricon-10 microconcentrators, it is most likely that the low yields obtained is due to this concentration step.

The protein integrity of the various fractions were checked on 10% SDS-polyacrylamide gels (insets in figures 3.4 and 3.5) as described in section 7.5. As shown in the insets the 0.2 M and the 0.4 M fractions still contained large amounts of contaminating proteins. These contaminating proteins would have to be removed by other chromatographic techniques to obtain a homogeneous preparation of these complexes.

3.3 DISCUSSION

Heparin-agarose chromatography proved to be an ideal technique for the partial purification of complex I, II and III proteins from SVWI-38 and CT-1 nuclear extracts. The three complexes bound to the column in buffer containing 0.1 M KCl. During this step about 60 to 65% of the loaded protein was removed from these complexes. Complex I proteins eluted from the column in buffer containing 0.2 M KCl, while complex II and III proteins co-eluted from the column in buffer containing 0.4 M KCl. Hatamochi *et. al.* (1988) have eluted a CCAAT-binding factor (CBF) which binds to the equivalent region of the mouse α2(I) procollagen promoter from a heparin-agarose column using buffer containing 0.35 M KCl. It is possible that one of the three complexes is similar to the mouse CBF.

Since the protein concentration of the eluted fractions were low, these fractions were concentrated using centricon-10 microconcentrators. Unfortunately, up to 50% of the protein was lost during this step, resulting in substantial loss of the DNA-binding activities. Secondly, the loss is DNA-binding activity could also be due to degradation caused by the long time period required to concentrate the fractions. This poor recovery of DNA-binding activity may be avoided by eliminating the concentration step by applying larger amounts of protein to the column Secondly, decreasing the flow rate can also minimise the dilution of separated proteins. Alternatively, other concentration procedures such as ammonium sulphate precipitation and dialysis, could be used.

Some of the biochemical properties of the three complexes using the partially purified heparin-agarose fractions will be covered in chapters 4 and 5. Detailed biochemical characteristics can only be determined once the proteins have been purified to homogeneity by sequence-specific DNA-affinity chromatography. Partial amino acid sequencing of the purified factors will also help in the cloning of their genes.

CHAPTER 4

OVERLAPPING TRANSCRIPTION-FACTOR BINDING-SITES WITHIN THE PROXIMAL HUMAN α2(I) PROCOLLAGEN PROMOTER

4.1 INTRODUCTION

The promoters and enhancers of eukaryotic genes are made up of multiple positive, negative, inducible and/or tissue specific transcriptional regulatory elements. Different combinations of these *cis*-elements are present and may be scattered both upstream and downstream of the transcription start site. The number, type, orientation and distance between regulatory *cis*-elements are unique for each of the thousands of genes in a cell. Sequence-specific DNA-binding proteins, which function as either activators or repressors, bind to these recognition sequences and interact with one another, other nuclear proteins and RNA polymerases to regulate gene expression (reviewed by Maniatis *et. al.*, 1987; Mitchell and Tjian, 1989).

Many RNA polymerase II-dependent promoters contain an AT-rich region centered about 25-30 base pairs upstream of the initiation site (Goldberg, 1979). This conserved element, the TATA box, the pentanucleotide CCAAT sequence and many others (section 1.3.6.1), have also been identified within the proximal promoters of the human and mouse α1(I) and α2(I) collagen genes (Dickson *et. al.*, 1985; Rossouw *et. al.*, 1987; Maity *et. al.*, 1988). Examples where gene expression is regulated by the competition of positive and negative transcription factors for distinct (Karsenty and de Crombrugghe, 1989), overlapping (Liu *et. al.*, 1989; Barberis *et. al.*, 1987) or the same (Harada *et. al.*, 1989) regulatory elements have been characterised. Karsenty and de Crombrugghe (1989) have shown that the binding of an inhibitory factor (IF2) to its consensus sequence is inhibited by the binding of a transcriptional activator (CBF)

to a CCAAT box containing element. The IF2 and CBF elements are distinct and situated within a 55 bp segment of the mouse α1(I) procollagen promoter.

The identification and characterisation of *cis*-acting transcription elements in the human α2(I) procollagen promoter has helped in elucidating the mechanism(s) responsible for regulating the expression of this gene. The development of techniques such as the electrophoretic mobility shift assay (Fried and Crothers, 1981), methylation interference assay (Siebenlist and Gilbert, 1980) and a multitude of other techniques (Ausubel *et. al.*, 1987; Hennighausen and Lubon, 1987) has made it possible to analyse the sequence-specific binding of proteins to DNA.

Electrophoretic mobility shift assays and modifications of this technique, such as competition binding assays (Carthew *et. al.*, 1985; Singh *et. al.*, 1986), have been widely used for the identification of DNA-protein complexes. Since the methylation of guanine and adenine residues in DNA-binding sites can interfere with binding of transcription factors, methylation interference assays were initially used to identify which purines in the T7 promoter prevented the binding of RNA polymerase (Siebenlist and Gilbert, 1980). Subsequently, many investigators have used this technique, coupled with the EMSA, to study the interaction of sequence-specific DNA-binding proteins with their binding elements in a variety of promoters and enhancers.

The identification and characterisation of two distinct but overlapping *cis*-acting elements involving the human α2(I) procollagen CCAAT box and the interaction of the candidate SVWI-38-specific transcriptional repressor with the proximal DNA-element is described.

4.2 RESULTS

4.2.1 Analysis of *cis*-elements in the proximal human α2(I) procollagen promoter

Three potential consensus sequences recognised by sequence-specific transcription factors were identified in the 65 bp region between -110 and -46 in the human α2(I) procollagen promoter. An inverted CCAAT box is situated between -80 to -84 and an inverted HC3 (Augereau and Chambon, 1986) element between -82 and -77 (figure 4.1). An AP-2 (Imagawa *et. al.*, 1987) consensus sequence occurs between nucleotides -104 and -95. Further inspection of this short region of the promoter revealed the following potential DNA-elements which could also be involved in DNA-protein complex formation with fibroblast extracts: (1) an AP-2-like sequence containing a single base mismatch is located between -98 to -89, (2) GGAGG and inverted GGAGG boxes within a region of inverted dyad symmetry centered at -82, (3) a TGGAGGC repeat situated between -79 and -61 and (4) a short CCCTAGGG palindrome between -59 to -52 (figure 4.1).

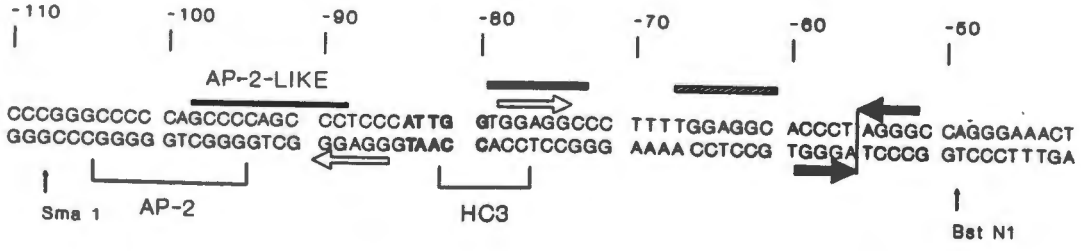


Fig 4.1 Potential *cis*-acting elements within the proximal human α2(I) procollagen promoter. The Sma 1 - Bst N1 promoter fragment contains an inverted CCAAT box (shown in bold), an AP-2 consensus sequence, HC3 motif and an AP-2-like sequence. The fragment also contains a GGAGG and inverted GGAGG boxes within a region of inverted dyad symmetry (open arrows), a direct TGGAGGC repeat (hatched boxes) and a short CCCTAGGG palindrome (shaded arrows).

4.2.2 Identification of DNA-elements involved in DNA-protein complex formation

Nuclear extracts from SVWI-38 and CT-1 fibroblasts formed two complexes with fragment E of the human α2(I) procollagen promoter (figure 2.8). In addition to complexes I and III, an additional complex (complex II) was identified in SVWI-38 extracts. Since sequence-specific DNA-binding proteins could potentially bind to a number of elements within this region of the promoter, competition binding experiments were performed to determine which of these elements were involved in complex I, II and III formation.

The ability of increasing molar excesses of unlabelled double-stranded oligonucleotides A, B, C or D (section 7.4.5 and figure 4.2) to compete with labelled fragment E for complex formation with SVWI-38 or CT-1 extracts was tested by EMSA as described in section 7.4.2.

Oligonucleotide C, corresponding to the 5'-flanking sequences containing the AP-2 and AP-2-like consensus sequences failed to compete out the formation of any of the complexes (figures 4.2 and 4.3). This finding is supported by previous studies (Smith, 1989). Similarly, no competition was observed with oligonucleotide B, containing the inverted CCAAT box and the 5'-flanking sequences (figures 4.2 and 4.3). These results eliminated the possibility that AP-2 or an AP-2-like protein was responsible for complex I, II or III formation with fragment E. Figures 4.2 and 4.3 also show a similar lack of detectable competition with oligonucleotide D. Therefore the TGGAGGC repeat and the CCCTAGGG palindrome, which are present in this oligonucleotide, are not involved in DNA-protein complex formation. This finding is also supported by the finding that a double-stranded 30 bp oligonucleotide, extending from position -77 to -48, was unable to compete out complexes I, II or III when SVWI-38 extracts were assayed (Smith, 1989). This lack of competition is maintained even at a 400-fold molar excess of this 30 bp oligonucleotide.

Oligonucleotide A, containing the inverted CCAAT box together with 5'- and 3'-flanking sequences, competed very effectively

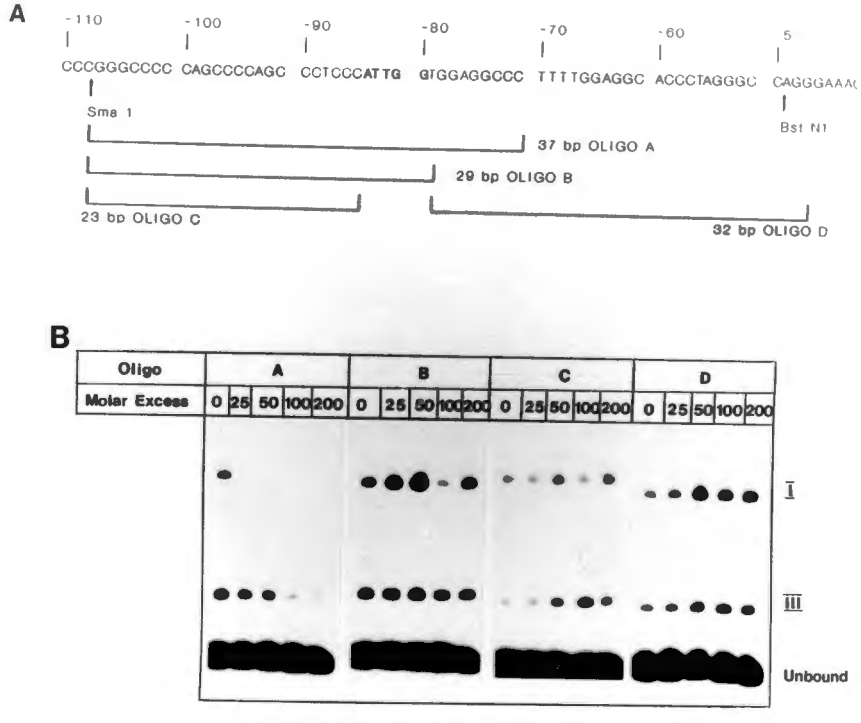


Fig 4.2 CT-1 competition assays. **(A)** Double-stranded oligonucleotides (oligo) A, B, C and D, corresponding to the indicated regions of fragment E (nucleotides -110 to -41) of the human α2(I) procollagen promoter, were prepared as described in section 7.4.5. **(B)** Increasing molar excesses of double-stranded oligonucleotides were incubated with CT-1 nuclear extracts for 10 min prior to the addition of labelled probe (fragment E) as described in section 7.4.2. The DNA-protein complexes were electrophoresed on 5% non-denaturing polyacrylamide gels at 4^{O} C in 0.5X TBE, the gels dried and exposed to X-ray film for 16 hrs. The position of complexes I and III are indicated.

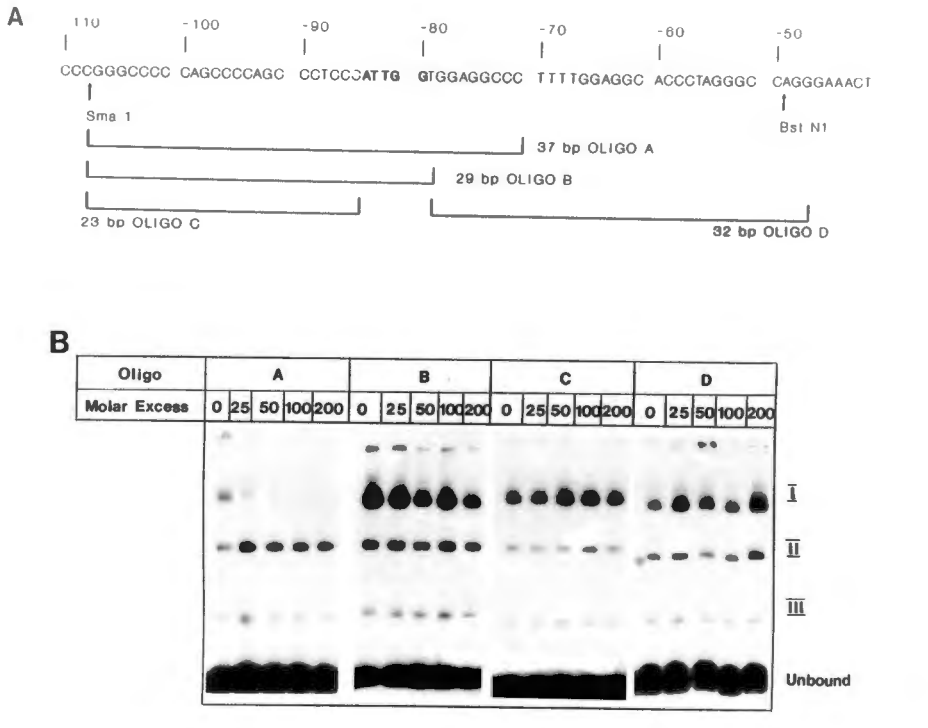


Fig 4.3 SVWI-38 competition assays. **(A)** Double-stranded oligonucleotides (oligo) A, B, C and D, corresponding to the indicated regions of fragment E (nucleotides -110 to -41) of the human α2(I) procollagen promoter, were prepared as described in section 7.4.5. **(B)** Increasing molar excesses of double-stranded oligonucleotides were incubated with SVWI-38 nuclear extracts for 10 min prior to the addition of labelled probe (fragment E) as described in section 7.4.2. The DNA-protein complexes were electrophoresed on 5% non-denaturing polyacrylamide gels at 4^{O} C in 0.5X TBE, the gels dried and exposed to X-ray film for 16 hrs. The position of complexes I, II and III are indicated.

for complex I formation in both CT-1 and SVWI-38 extracts. This competition was observed with a low molar excess (25-fold) of the oligonucleotide, suggesting that complex I proteins bound to this oligonucleotide with a high affinity. There was also weaker competition for complex III (figures 4.2 and 4.4). With SVWI-38 extracts on the other hand, there was an initial increase in the intensities of complexes II and III followed by a gradual decrease in the intensity of complex III (figures 4.3 and 4.5). The initial increase in the intensities of complexes II and III suggested that all three complexes did not form on the same DNA molecule. The preferential inhibition of complex I formation with the competing oligonucleotide probably caused more probe to be available for complex II and III formation, resulting in an initial increase in their intensities. At higher molar excesses of oligonucleotide A, after the competitor had bound to all the available complex I proteins, the vast excess free oligonucleotide competed weakly with complex III proteins, causing a reduction in its intensity.

If the above explanation is valid, then complex III would be competed out more effectively by oligonucleotide A when limiting amounts of protein was used in the binding reactions. As shown in figures 4.4b and 4.5b there is a marked decrease in the intensity of complex III when less SVWI-38 or CT-1 protein was used. There was also a reduction in the intensity of complex II when lower SVWI-38 protein concentrations were assayed (figure 4.5b).

Competition of the 0.2 M SVWI-38 heparin-agarose fractions containing partially purified complex I proteins with oligonucleotide A resulted in a drastic reduction in the intensity of complex I at a low molar excess (25-fold) of oligonucleotide (figure 4.6). Oligonucleotide A competed out complexes II and III very weakly when the 0.4 M heparin-agarose fractions were similarly assayed. These competition experiments suggested that oligonucleotide A contains the minimal elements required for complex I and some of the sequences for complex II and III formation. The inverted CCAAT box, together with 5'- and 3'-flanking sequences, are probably important parts of the recognition sequence.

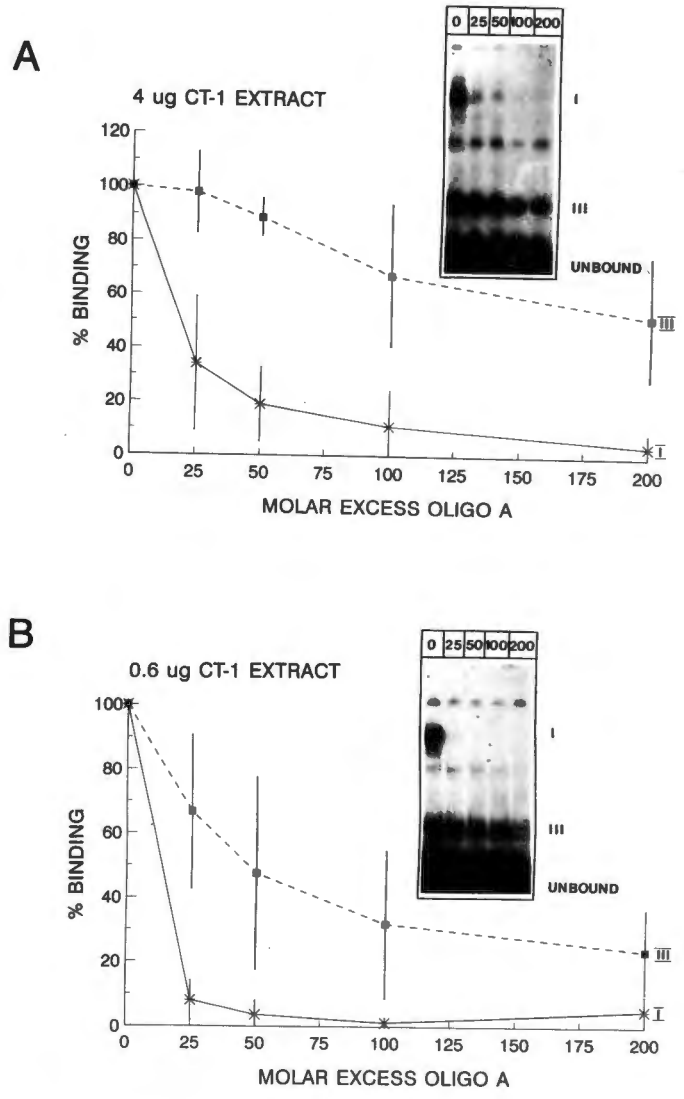


Fig 4.4 The competition kinetics of CT-1 complexes I and III with oligonucleotide A. **(A)** Increasing molar excess of double-stranded oligonucleotide A was incubated with 4 ug CT-1 nuclear extracts and labelled probe (fragment E) (section 7.4.2). The position of complexes I and III are indicated in the inset. The autoradiograms were scanned and the average percentage binding and standard deviation (n=4) of complexes I and III at each molar excess of oligonucleotide A assayed was plotted. Complex I is represented by (▪---▪), while (X--X) represents complex III. **(B)** As for (A) except that 0.6 ug of CT-1 nuclear extracts were assayed. The percentage binding and standard deviation were calculated from three separate experiments.

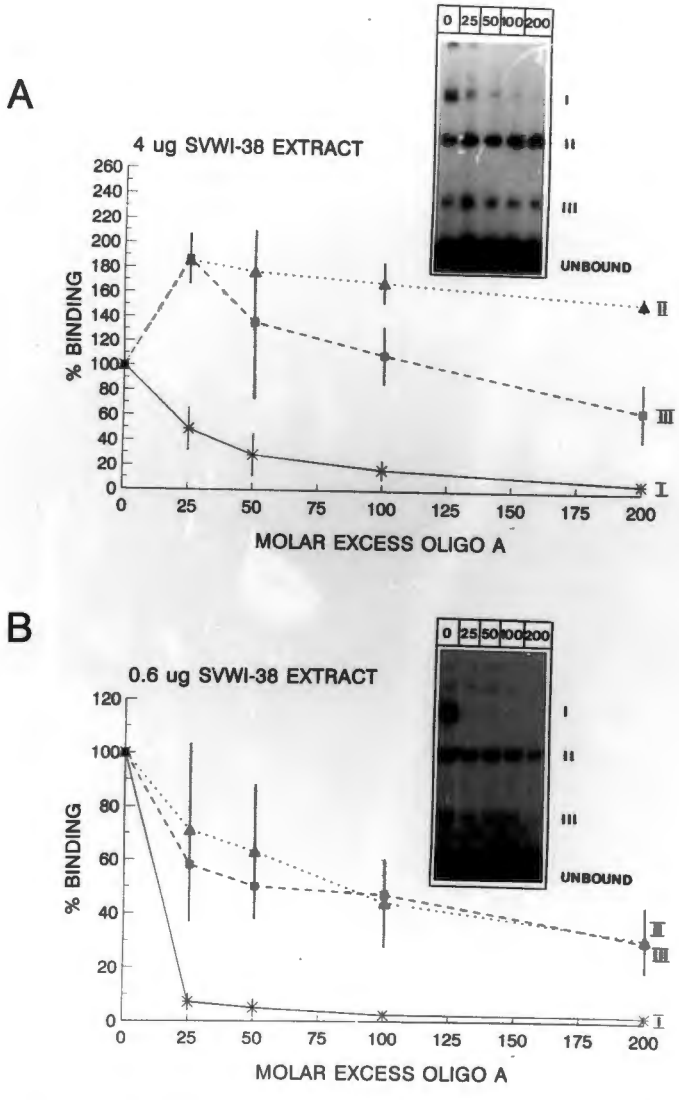


Fig 4.5 The competition kinetics of SVWI-38 complexes I, II and III with oligonucleotide A. **(A)** Increasing molar excess of double-stranded oligonucleotide A was incubated with 4 ug SVWI-38 nuclear extracts and labelled probe (fragment E) (section 7.4.2). The position of complexes I, II and III are indicated in the inset. The autoradiograms were scanned and the average percentage binding and standard deviation (n = 3 or 4) of complexes I, II and III at each molar excess of oligonucleotide A assayed was plotted. Complex I is represented by (▪--▪), (▲···▲) represent complex II while, (X--X) represents complex III. **(B)** As for (A) except that 0.6 ug of SVWI-38 nuclear extracts were assayed. The percentage binding and standard deviation were calculated from three separate experiments.

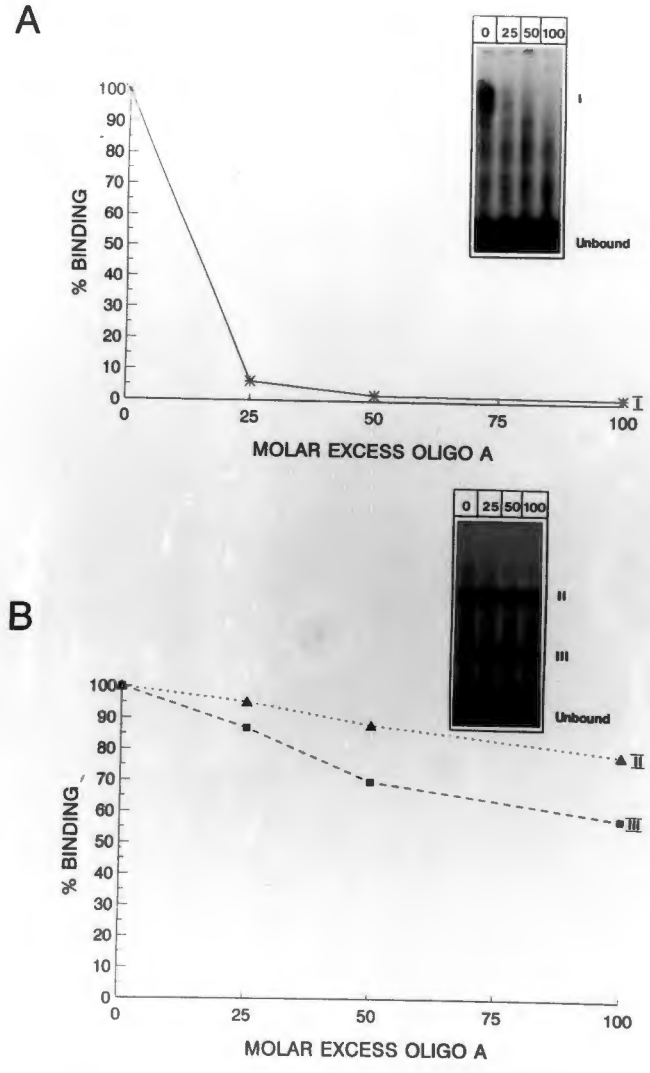


Fig 4.6 The competition kinetics of partially purified complexes I, II and III with oligonucleotide A. **(A)** An increasing molar excess of double-stranded oligonucleotide A was incubated with 4 ug partially purified complex I proteins (0.2 M SVWI-38 heparin-agarose fractions) and labelled probe (fragment E) (section 7.4.2). The position of complexes I is indicated in the inset. The autoradiograms were scanned and the percentage binding at each molar excess of oligonucleotide A assayed was plotted. **(B)** As for (A) except that 4 ug of partially purified complex II and III proteins (0.4 M SVWI-38 heparin-agarose fractions) were assayed. Complex II is represented by (▲···▲), while (X--X) represent complex III.

To confirm the competition experiment results, electrophoretic mobility shift assays were done with SVWI-38 or CT-1 nuclear extracts and radioactively labelled oligonucleotides as described in section 7.4.2. In addition to the four oligonucleotides (A, B, C and D) used in the competition experiments, a 21-mer oligonucleotide E (-74 to -94) and a 47-mer oligonucleotide F (-107 to -61) were used in the binding studies (section 7.4.5). As shown in figures 4.7 and 4.8 none of the complexes formed with oligonucleotide C, whereas trace amounts of complex I was visible after longer exposures (2 days) of the autoradiograms of oligonucleotide B complexes. Since this oligonucleotide contains the inverted CCAAT box and 5'-flanking sequences, this result suggested that the recognition element for complex I probably also extended into the 3'-flanking region.

In the case of oligonucleotide D, trace amounts of complex III (figure 4.7) was visible in longer exposures (3 days) of the autoradiograms. These results implied that the 3'-flanking sequences are required for complex III formation. Under these conditions no detectable amounts of complex III was identified when SVWI-38 extracts were assayed with this oligonucleotide (figure 4.8). Since previous experiments suggested that the concentrations of complex III proteins is lower in SVWI-38 extracts than in CT-1 extracts (chapter 2), these results were not surprising.

As expected from the competition data, all three complexes formed on oligonucleotides A, E and F, with complex I being more predominant than II and III. All three of these oligonucleotides contain the inverted CCAAT box and varying lengths of the 5'- and 3'-flanking sequences. The smallest oligonucleotide, oligonucleotide E, containing 21 bp of the α2(I) procollagen promoter (-94 to -74), most probably contains the minimal elements required for complex I, II and III formation. When compared with the binding results of oligonucleotides A or F, the intensities of complexes II and III were lower when extracts were assayed with oligonucleotide E (figure 4.8). This observation implies that additional sequences present in oligonucleotides A and F are required for complex II and III formation. An inverted CCAAT box, an inverted GGAGG repeat flanking the CCAAT box

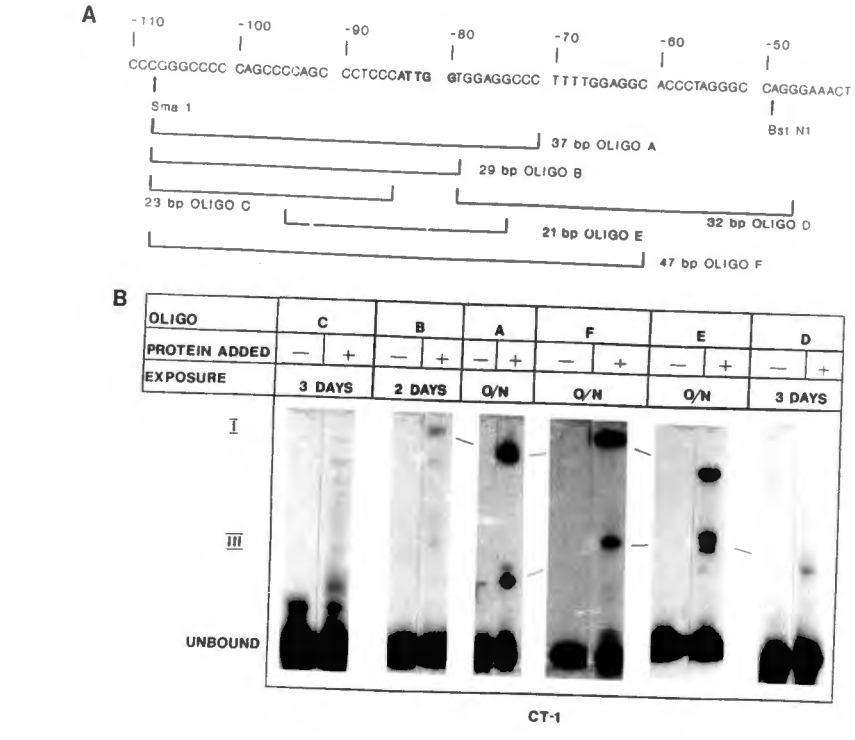


Fig 4.7 Binding assays between labelled oligonucleotides and CT-1 nuclear extracts. **(A)** Double-stranded oligonucleotides (oligo) A, B, C, D, E and F, corresponding to the regions indicated were labelled as described in sections 7.4.5 and 7.4.1. **(B)** 4 ug of CT-1 nuclear extracts (lanes marked +) were incubated with labelled probe (oligo) and the DNA-protein complexes electrophoresed, together with free probe (lanes marked -), on 5% non-denaturing polyacrylamide gels as described in section 7.4.2. The dried gels were exposed to X-ray film for the indicated period of time (exposure). The position of complexes I and III are indicated.

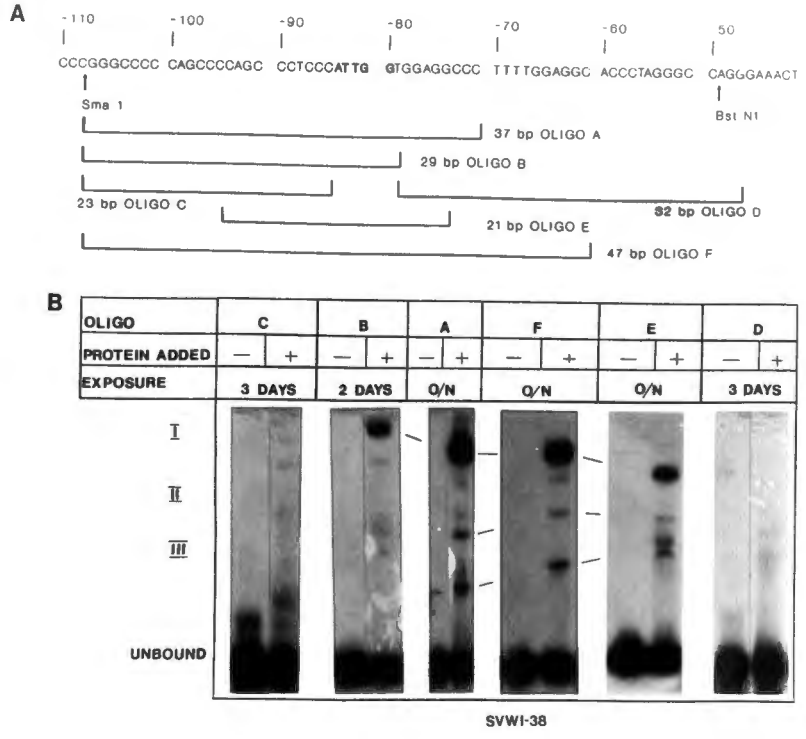


Fig 4.8 Binding assays between SVWI-38 nuclear extracts and labelled oligonucleotides. **(A)** Double-stranded oligonucleotides (oligo) A, B, C, D, E and F, corresponding to the regions indicated were labelled as described in sections 7.4.5 and 7.4.1. **(B)** 4 ug of SVWI-38 nuclear extracts (lanes marked +) were incubated with labelled probe (oligo) and the DNA-protein complexes electrophoresed, together with free probe (lanes marked -), on 5% non-denaturing polyacrylamide gels as described in section 7.4.2. The dried gels were exposed to X-ray film for the length of time indicated (exposure). The position of complexes I, II and III are indicated.

and an inverted HC3 motif are all present in this short fragment of the collagen promoter. It is possible that some or all three of these motifs are important for complex I, II and III formation.

The binding experiments also showed that the complex I and III recognition elements are distinct. The element responsible for complex I formation appears to be upstream to the element responsible for complex III formation (figure 4.7). These experiments were not sensitive enough to determine the relative position of the complex II *cis*-element.

4.2.3 Mutually exclusive DNA-protein complex formation

The following observations suggested that there is mutually exclusive formation of complexes I, II and III with *cis*-elements in the human α2(I) procollagen promoter:- (1) The initial increase in the intensities of complexes II and III when oligonucleotide A was used in competition studies (figure 4.5). (2) The minimal sequences required for complex I, II and III formation are located within a short 21 bp region of the α2(I) procollagen promoter. To examine the possibility that the formation of one of the DNA-protein complexes prevented the simultaneous binding of the remaining transcription factors, mutually exclusive binding assays were done (see section 7.4.3).

Increasing amounts of partially purified CT-1 (figure 4.9) or SVWI-38 (figure 4.10) complex I proteins pre-bound to fragment E, resulted in decreased formation of complexes II and III upon the addition of partially purified complex II and III proteins (0.4 M KCl heparin-agarose fractions). Similarly, saturating fragment E with partially purified CT-1 complex III (figure 4.9) or SVWI-38 complex II and III proteins (figure 4.10a) was accompanied by a decrease in the formation of complex I. These results showed that complex I, II or III proteins do not bind simultaneously to their respective *cis*-regulatory elements. When partially purified complex I, II or III proteins were added to bind to fragment E simultaneously all three complexes formed

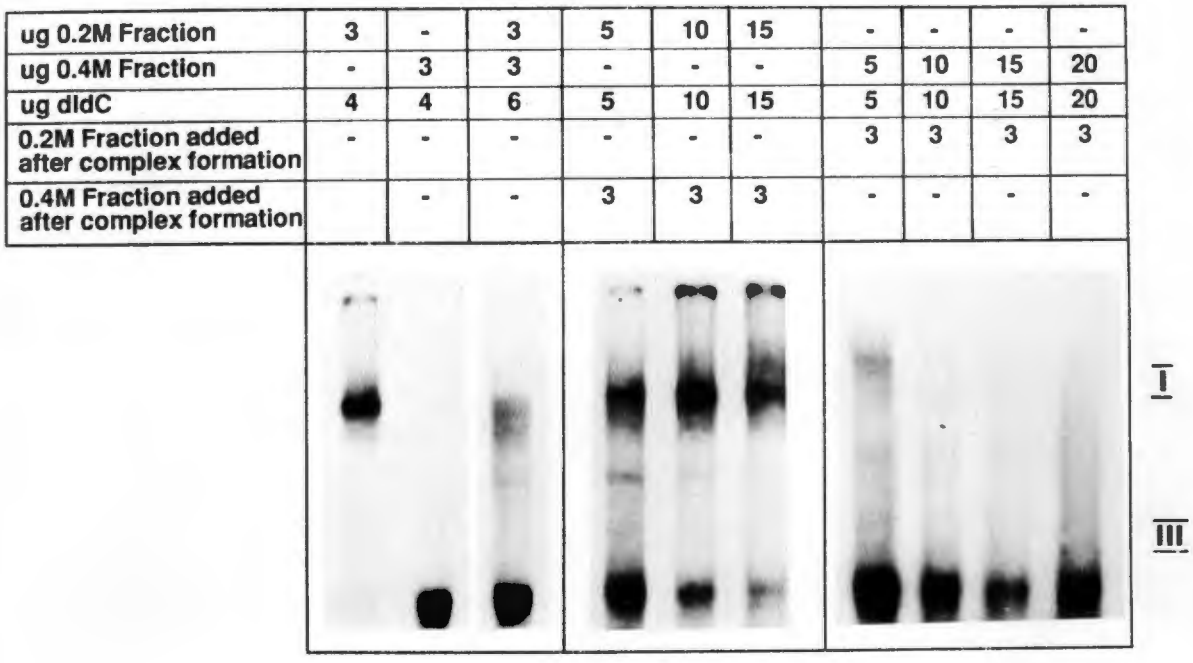
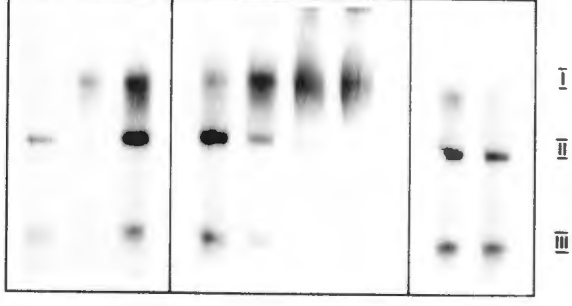


Fig 4.9 Mutually exclusive binding assays with partially purified complex I and III proteins. Fragment E (-110 to -46) was incubated with increasing amounts of partially purified complex I (0.2 M CT-1 heparin-agarose fraction) or complex III (0.4 M CT-1 heparin-agarose fraction) proteins for 20 min on ice as described in section 7.4.3. Three ug of partially purified complex III (0.4 M CT-1 heparin-agarose fraction) or complex I proteins (0.2 M CT-1 heparin-agarose fraction) were added, incubated for a further 20 min on ice, the DNA-protein complexes were electrophoresed on 5% non-denaturing polyacrylamide gels, the gels dried and exposed to X-ray film for 16 hrs (section 7.4.2). Binding of complex I or III proteins on their own or mixing equal amounts of the two fractions are shown in the first three lanes. The position of complexes I and III are indicated.

A

ug 0.2M Fraction	-	3	3	5	10	15	20	-	-
ug 0.4M Fraction	3	-	3	-	-	-	-	5	10
ug dIdC	4	4	6	5	10	15	20	5	10
0.2M Fraction added after complex formation	-	-	-	-	-	-	-	3	3
0.4M Fraction added after complex formation		-	-	3	3	3	3	-	-



B

ug 0.4M CT-1 Fraction	3	-	5	10
ug 0.4M SVWI-38 Fraction	-	3	-	-
ug dIdC	4	4	5	10
0.4MSVWI-38 Fraction added after complex formation	-	-	3	3

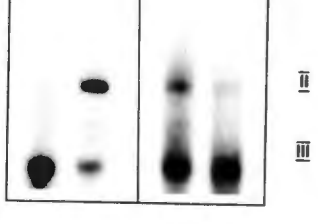


Fig 4.10 Mutually exclusive binding assays with partially purified complex I, II and III proteins. **(A)** Fragment E (-110 to -46) was incubated with increasing amounts of partially purified complex I (0.2 M SVWI-38 heparin-agarose fractions) or complex II and III (0.4 M SVWI-38 heparin-agarose fractions) proteins prior to the addition of 3 ug of partially purified complex II and III proteins or complex I proteins as described in the legend of figure 4.9. The position of complexes I, II and III are indicated. **(B)** Increasing amounts of partially purified complex III proteins (0.4 M CT-1 heparin-agarose fraction) were pre-bound to fragment E prior to the addition of 3 ug of partially purified complex II and III proteins (0.4 M SVWI-38 fraction) and assayed as in (A). The CT-1 and SVWI-38 0.4 M heparin-agarose fractions were also assayed without adding other protein fractions. The position of complexes II and III are indicated.

presumably each on separate DNA molecules (figures 4.9 and 4.10a).

Figure 4.10b shows that prior binding of increasing amounts of partially purified complex III proteins to fragment E abolished the formation of complex II. This result suggested that complex III proteins also do not bind simultaneously with complex II proteins to their regulatory elements either.

These results support the hypothesis that complex I, II and III formation occur in a mutually exclusive manner.

4.2.4 Methylation interference analysis

As already discussed, complex I, II and III proteins interact with sequences located within a 21 bp segment extending from nucleotides -94 to -74. This 21 bp fragment contains an inverted CCAAT box, an inverted GGAGG repeat and an inverted HC3 motif. Since the competition and gel retardation assays failed to show which of these three DNA-elements were responsible for DNA-protein complex formation, methylation interference assays were performed (section 7.4.4). This assay was used to identify the relative importance of individual purine residues in complex formation.

Methylation of all the guanine and most of the adenine bases on both strands between nucleotides -92 and -72 interfered to some degree with complex I formation when CT-1 (figures 4.11 and 4.12a) or SVWI-38 (figures 4.13 and 4.14a) extracts were assayed. In both CT-1 (figure 4.12a) and SVWI-38 (figure 4.14a) complex I proteins bound to a 21 bp recognition element, $^{5'}$GCCCTCCCATTGGTGGAGGCC$^{3'}$, extending from -92 to -72 (figure 4.15). This tripartite recognition element contains a GGAGG box and its inverted counterpart (CCTCC) flanking the CCAAT box. Since purines in both the GGAGG boxes and the CCAAT box are involved in DNA-protein contacts, this recognition element will be referred to as a GGAGG/CCAAT binding element (G/CBE), while complex I proteins will be called a GGAGG/CCAAT box factor (G/CBF). Although purine residues within the HC3 motif, which is also present in the G/CBE, are involved in DNA-protein contacts, the methylation patterns of the G/CBE and HC3 motif are different (Augereau

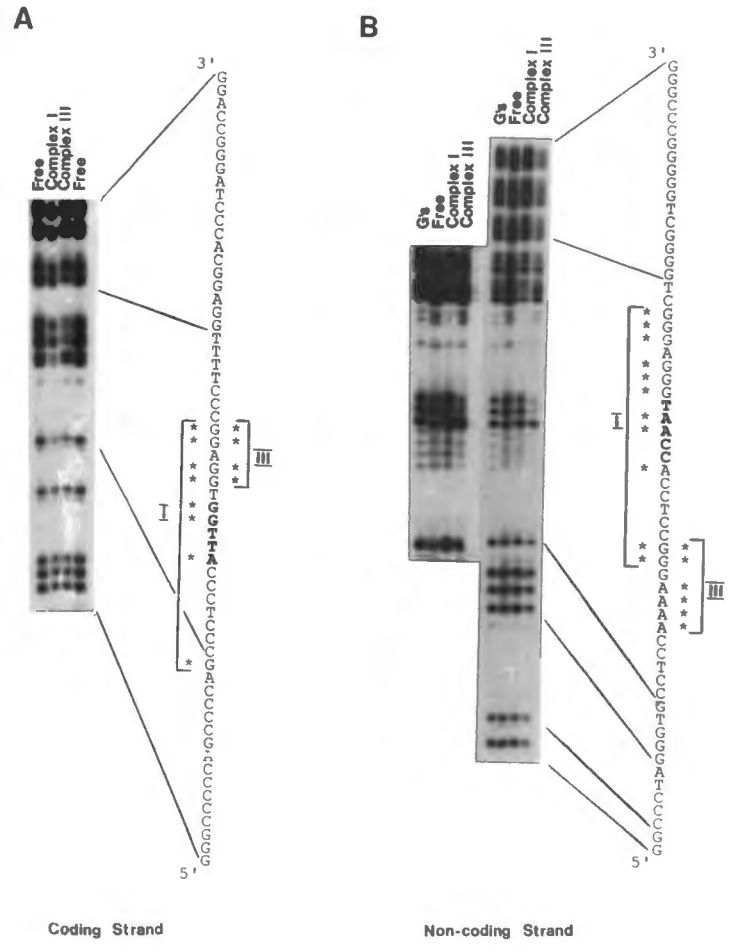


Fig 4.11 Methylation interference analysis of complex I and III formation. DNA was end-labelled with γ-^{32}P-ATP, partially methylated with DMS and incubated with crude CT-1 nuclear extract as described in section 7.4.4. DNA-protein complexes were resolved on 5% non-denaturing polyacrylamide gels from which the free and complexed DNA was extracted, cleaved with piperidine and resolved on 12% urea/polyacrylamide sequencing gels as described in Materials and Methods (section 7.4.4). The dried gels were exposed to X-ray film for at least 16 hrs. **(A)** The DNA sequence of the coding strand between nucleotides -107 and -47. The guanine residues which are important for complex I and III formation are shown with asterisks. The CCAAT box is in bold. **(B)** The DNA sequence of the coding strand between nucleotides -107 and -50. The guanine residues which are important for complex formation are indicated as in A. The autoradiogram on the left is a longer exposure in order to shown the protected adenines. The lane marked with G's represents DNA cleaved with DMS to indicate the guanines in the sequence..

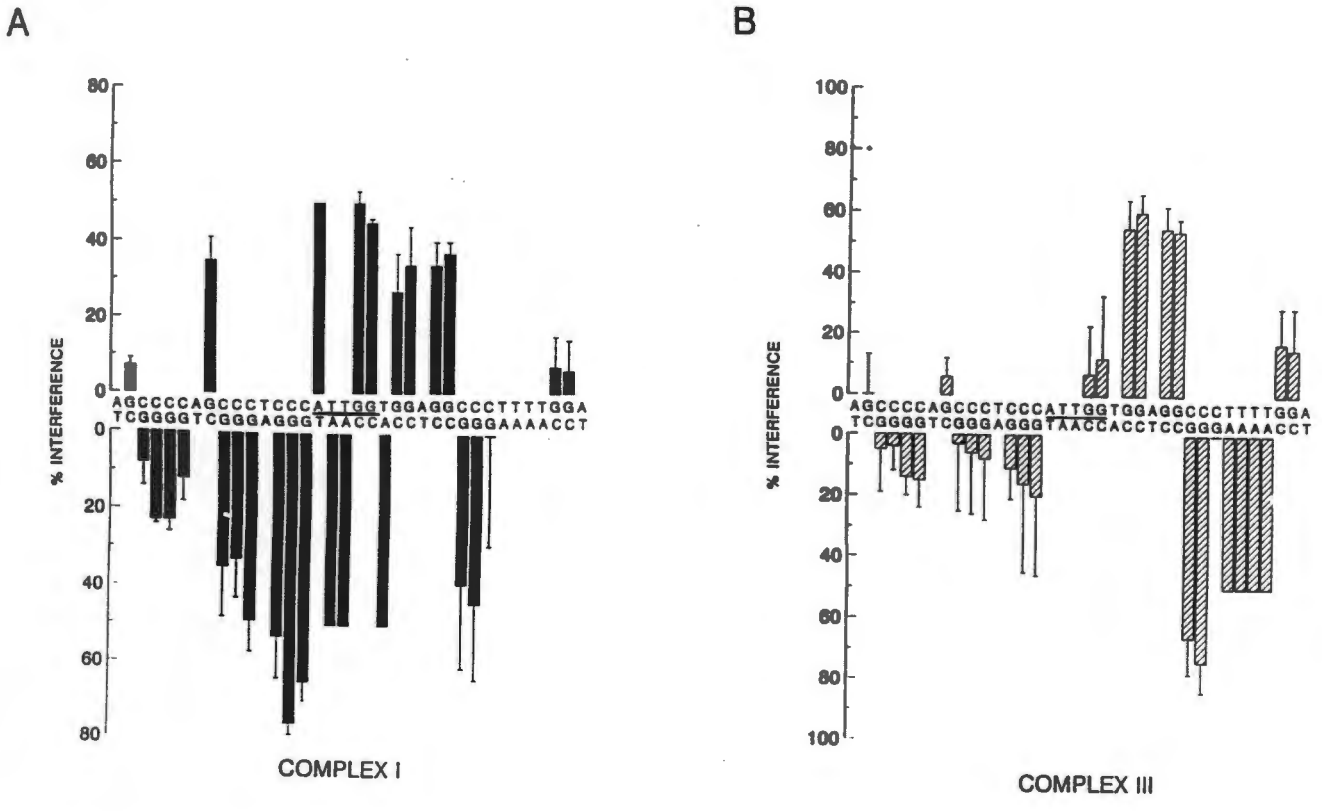


Fig 4.12 Histogram of the CT-1 Methylation interference data. Methylation interference analysis were performed as described in the legend to figure 4.11. The autoradiograms were scanned and the average percentage interference and standard deviations for each guanine residue was determined. Both the coding (top) and non-coding (bottom) strands are shown. Interference by methylation of an adenine residue is indicated as 50% interference. The inverted CCAAT box is underlined. **(A)** Methylation pattern of complex I (n = 4). **(B)** Methylation pattern of complex III (n = 4).

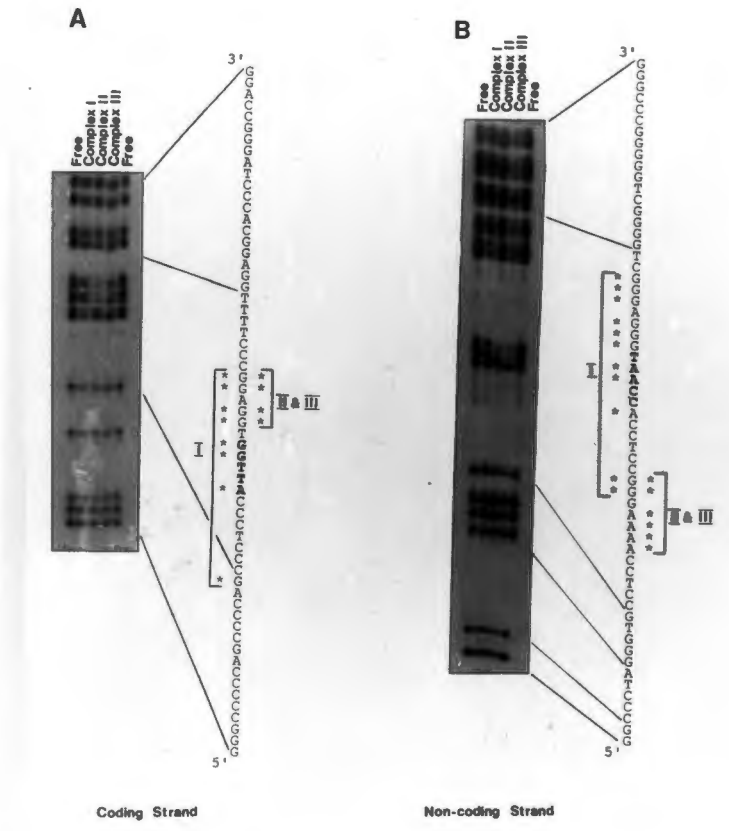


Fig 4.13 Methylation interference analysis of SVWI-38 complex I, II and III formation. DNA was end-labelled with γ-^{32}P-ATP, partially methylated with DMS and incubated with crude SVWI-38 nuclear extract as described in section 7.4.4. DNA-protein complexes were resolved on 5% non-denaturing polyacrylamide gels from which the free and complexed DNA was extracted, cleaved with piperidine and resolved on 12% urea/polyacrylamide sequencing gels as described in Materials and Methods (section 7.4.4). The dried gels were exposed to X-ray film for at least 16 hrs. **(A)** The DNA sequence of the coding strand between nucleotides -107 and -47. The guanine residues which are important for complex I, II and III formation are shown with asterisks. The CCAAT box is in bold. **(B)** The DNA sequence of the coding strand between nucleotides -107 and -50. The guanine residues which are important for complex formation are indicated as in A.

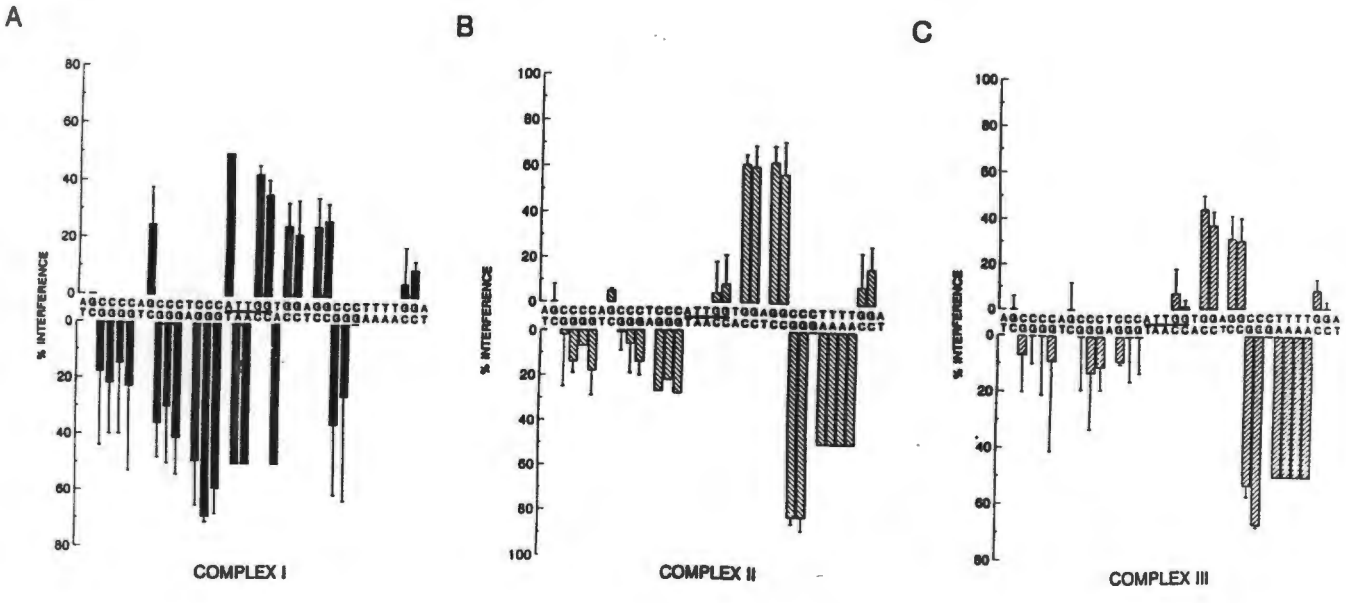


Fig 4.14 Histogram of the SVWI-38 Methylation interference data. Methylation interference analysis were performed as described in the legend to figure 4.13. The autoradiograms were scanned and the average percentage interference for each purine residue was determined as described in the legend of figure 4.12. Both the coding (top) and non-coding (bottom) strands are shown. Interference by methylation of an adenine residue is indicated as 50% interference. The inverted CCAAT box is underlined. **(A)** Methylation pattern of complex I (n = 3). **(B)** Methylation pattern of complex II (n = 3). **(C)** Methylation pattern of complex III (n = 3).

and Chambon, 1986). It is, therefore, unlikely that complex I is related to the unknown protein from HeLa extracts which binds to the HC3 motif in the mouse immunoglobulin heavy-chain enhancer. Interestingly, the results presented in figures 4.12a and 4.14a also indicates that methylation of the non-coding strand interfered more strongly with complex formation than methylation of the coding strand. Although purines in both the upstream and downstream GGAGG boxes were involved in DNA-protein contacts, the upstream inverted GGAGG and CCAAT boxes appeared to be more important for complex I formation.

Methylation of the four guanine bases in the downstream GGAGG box reduced the formation of complexes II and III as did methylation of the two guanines at positions -72 and -73 and the four A's between nucleotides -70 and -67 on the non-coding strand. The nucleotides at positions -72 and -73 appeared to form the core of this recognition sequence. The guanine at -71 does not appear to be involved in DNA-protein contacts (figures 4.11 and 4.13). As illustrated in figures 4.12b, 4.14b and 4.14c and summarised in figure 4.15, the similarity of the complex II and III interference patterns suggested that they bind to the same DNA-element. Since complex II and III proteins are most probably a transcriptional repressor and an activator respectively (section 2.3), it is proposed that their recognition element be called a collagen modulating element (CME). The 12 bp CME situated between nucleotides -78 to -67, $^{5'}$**GGAGG**CCCTTTT$^{3'}$, contains the downstream GGAGG box and 3'-flanking sequences. The data in figures 4.12b, 4.14b and 4.14c also suggested that purines upstream of this GGAGG box were weekly involved in complex II and III DNA-protein contacts. Since this GGAGG box and two of the 3'-flanking nucleotides ($^{5'}$**GGAGG**CC$^{3'}$) also form part of the G/CBE, the human α2(I) procollagen promoter contains two distinct but overlapping *cis*-elements situated around the CCAAT box.

Since the N3 of adenine and N7 of guanine are methylated by dimethyl sulphate in the minor and major grooves respectively (Siebenlist and Gilbert, 1980), complex I, II and III proteins bound in both the major and minor grooves of DNA.

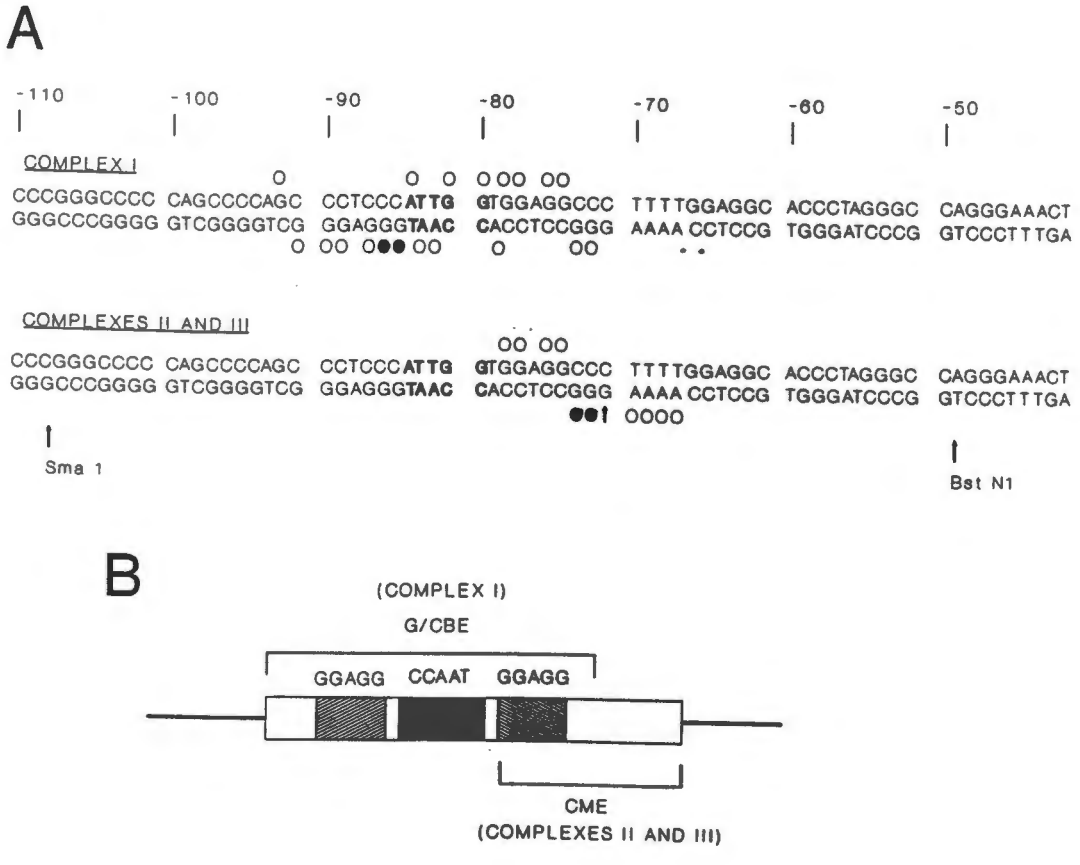


Fig 4.15 Summary of the methylation interference data. (A) The 70 bp fragment of the human α2(I) procollagen promoter (nucleotides -110 to -41) showing the inverted CCAAT box (bold) and the purines that interfere with complex I (top sequence) and complexes II and III (bottom sequence) formation; open circles show 25 to 70% interference and solid circles >70% interference. Methylation of the guanine indicated with an arrow enhanced complex formation. **(B)** A schematic diagram of the G/CBE and CME (open box) within the human α2(I) procollagen promoter. The inverted CCAAT box is shown as a solid box, while the upstream inverted and downstream GGAGG boxes are shown as hatched boxes.

Complex I and II proteins were also identified in a non-collagen producing mouse macrophage tumour cell line, $p388D_1$ (figure 2.9). Figure 4.17 shows that these murine proteins are related to or identical to the human complex I and II proteins because they bound in an identical manner to the G/CBE and CME of the human α2(I) procollagen promoter. Hatamochi *et. al.* (1986, 1987) have shown that only CBF in NIH-3T3 nuclear extracts binds to the equivalent region of the mouse α2(I) procollagen promoter. This and other studies (V. Leaner, unpublished data), on the other hand, identified proteins responsible for complex I and II formation with the highly homologous human promoter in murine macrophage, fibroblast and myeloma nuclear extracts.

The presence of complex I proteins in a number of mouse and human cell lines together with the homology between the murine CBF recognition element and the human G/CBE (figure 4.16), strongly suggested that the proteins responsible for complex I (G/CBF) formation maybe identical or similar to the CBF in NIH-3T3 cells (Hatamochi *et. al.*, 1986; Maity *et. al.*, 1988) (see chapter 5). CME on the other hand, contains a crucial three-base mismatch between the human and mouse promoters (figure 4.16) and therefore probably will not form complex II, although these proteins are present in many murine cell lines.

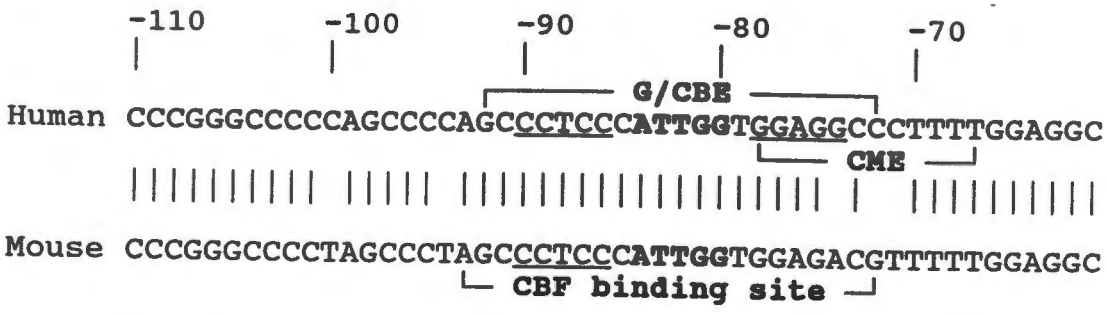


Fig 4.16 Comparison of the G/CBE with the mouse CBF recognition element. Sequence alignment of a 50 bp region (nucleotides -110 to -61) of the human and mouse α2(I) procollagen promoters. The inverted CCAAT box is shown in bold and the inverted and direct GGAGG boxes are underlined. This region of the human promoter contains the G/CBE and CME, while the equivalent region of the mouse promoter only contains a binding site for CBF (or G/CBF).

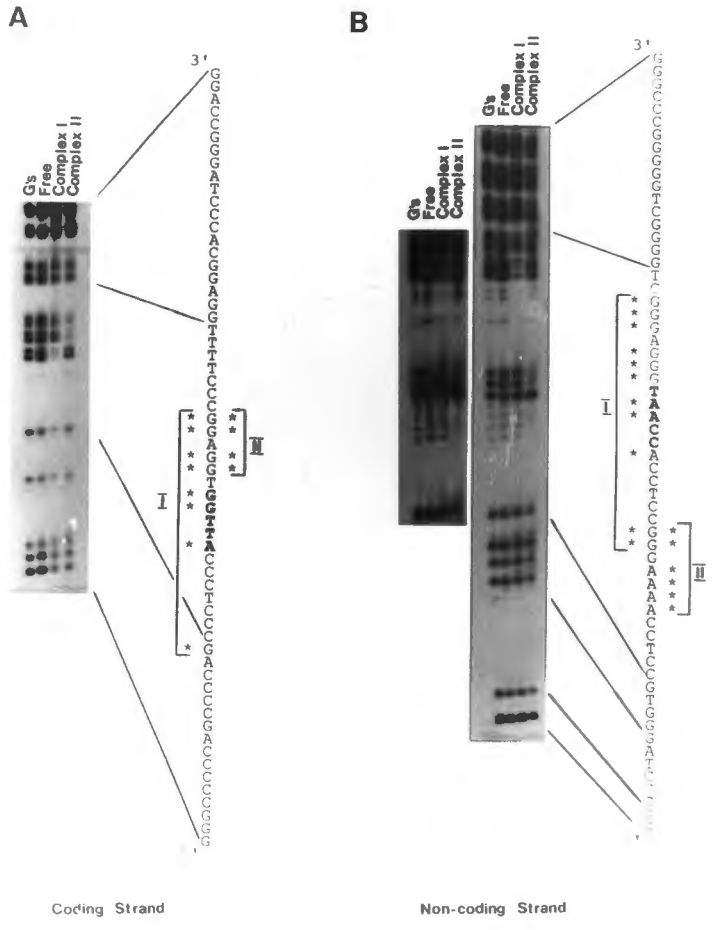


Fig 4.17 Methylation interference analysis of p388D$_1$ complex I and II formation. DNA was end-labelled with γ-^{32}P-ATP, partially methylated with DMS and incubated with crude mouse tumour macrophage, p388D$_1$, nuclear extracts as described in section 7.4.4. DNA-protein complexes were resolved on 5% non-denaturing polyacrylamide gels from which the free and complexed DNA was extracted, cleaved with piperidine and resolved on 12% urea/polyacrylamide sequencing gels as described in Materials and Methods (section 7.4.4). The dried gels were exposed to X-ray film for at least 16 hrs. The lanes marked with G's represent DNA cleaved DMS to indicate the guanines in the sequence. **(A)** The DNA sequence of the coding strand between nucleotides -107 and -47. The guanine residues which are important for complex I and II formation are shown with asterisks. The CCAAT box is in bold. **(B)** The DNA sequence of the coding strand between nucleotides -107 and -50. The guanine residues which are important for complex formation are indicated as in (A). The autoradiogram on the left is a longer exposure in order to show the protected adenines.

4.2.5 Mutation analysis

As summarised in figure 4.16, the methylation interference assays identified two distinct overlapping elements in the human α2(I) procollagen promoter. Methylation of purines in these elements interfered to varying extents with complex formation. Double-stranded oligonucleotides containing mutated GGAGG and CCAAT boxes were used in EMSA to test the relative importance of these motifs in complex I, II and III formation (sections 7.4.2 and 7.4.5).

Mutating the inverted CCAAT box from ATTGG to CTTGC resulted in a significant reduction in complex I (81 ± 12%) and III (37 ± 7%) in CT-1 (figure 4.18) and SVWI-38 (figure 4.19) nuclear extracts. Similar mutations in the mouse α2(I) procollagen CCAAT box also abolished CBF binding (Karsenty *et. al.*, 1988). The slight reduction in complex III formation supported the observation from the methylation interference assays that sequences 5' of the CME, which includes the CCAAT box, were weakly involved in DNA-protein contacts (figures 4.12b and 4.14c). Binding of complex II proteins to the MUT-CCAAT oligonucleotide was virtually unaffected when compared to binding with the wild-type oligonucleotide (oligonucleotide F) (98 ± 22%).

The upstream and downstream GGAGG boxes were mutated to TTATT in the oligonucleotides MUT-US and MUT-CME respectively. As expected from the methylation interference data, formation of complex II was reduced (68 ± 13%) and complex III was abolished (98%) on MUT-CME (figure 4.19). This was accompanied by a drastic increase in the binding (147 ± 44%) of complex I proteins (G/CBF) to this oligonucleotide. This implied that the downstream GGAGG box is not vital for complex I formation. This result confirmed the methylation interference data which suggested that this motif was not as important as the upstream motifs for G/CBF binding.

Complex I formation was virtually abolished when the upstream inverted GGAGG box was mutated to TTATT (oligonucleotide MUT-US, 5 ± 4% binding). Since this GGAGG box does not form part of the CME, there was no reduction in the formation (106 ± 15%) of complex III on oligonucleotide MUT-US when

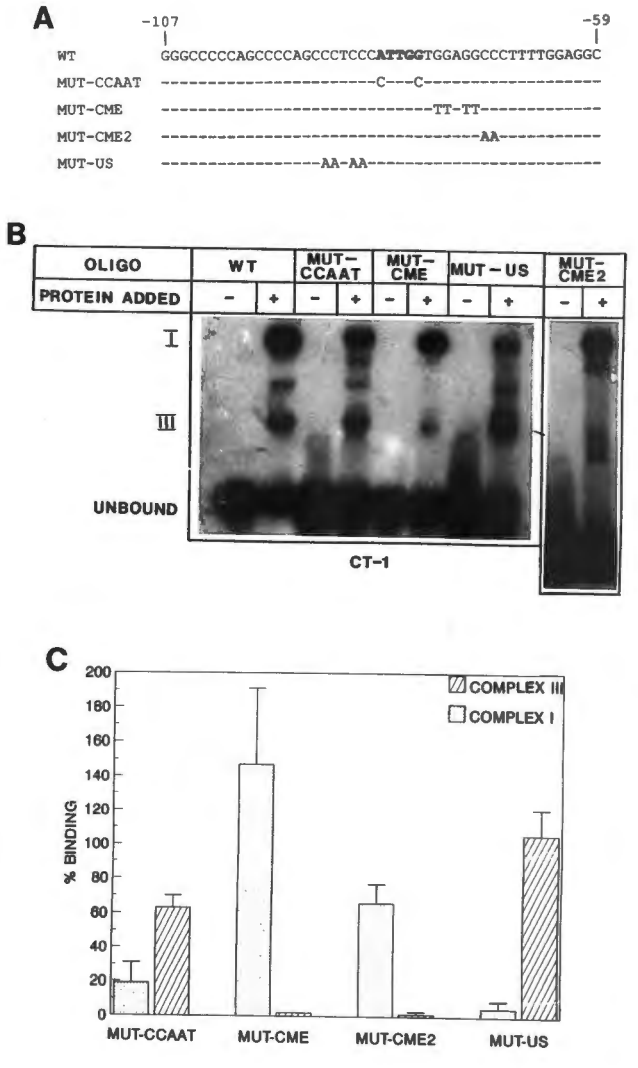


Fig 4.18 Mutation analysis of CT-1 complex formation. **(A)** Double-stranded oligonucleotides, containing mutated GGAGG boxes (MUT-CME, MUT-US), CCAAT box (MUT-CCAAT) and the cytosine dinucleotides at -72 and -73 (MUT-CME2), were used in EMSA. **(B)** 3-4 ug of crude CT-1 nuclear extract was incubated with 1 ng (10^4 cpm) labelled double-stranded oligonucleotide (+) or without added protein (-) and electrophoresed on 5% non-denaturing polyacrylamide gels as described in section 7.4.2. The gels were fixed in 10% acetic acid for at least 20 min, dried and exposed to X-ray film for at least 16 hrs. The positions of complexes I and III are indicated. Complex I is over-exposed in the figure. **(C)** The autoradiograms were scanned and the average relative percentage binding and standard deviation for each complex was plotted.

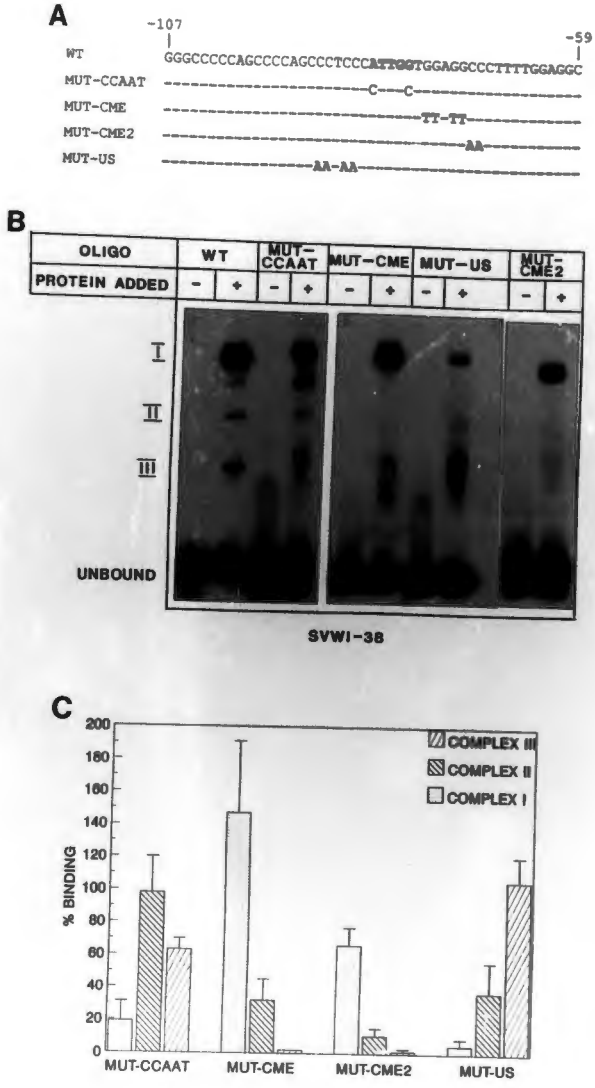


Fig 4.19 Mutation analysis of SVWI-38 complex formation. **(A)** Double-stranded oligonucleotides were prepared and radioactively labelled as described in the legend of figure 4.18. **(B)** 3-4 ug of crude SVWI-38 nuclear extracts were incubated with 1 ng (10^4 cpm) labelled probe (+) or without added protein (-) and electrophoresed on 5% non-denaturing polyacrylamide gels as described in section 7.4.2. The gels were fixed in 10% acetic acid for at least 20 min, dried and exposed to X-ray film for at least 16 hrs. The positions of complexes I, II and III are indicated. Complex I is over-exposed in the figure. **(C)** The autoradiograms were scanned and the average relative percentage binding and standard deviation for each complex was plotted.

CT-1 or SVWI-38 extracts were assayed. There was, however, a puzzling reduction (63 ± 19%) in the formation of complex II on this mutated oligonucleotide (figure 4.19). This implied that certain sequences upstream of the CME are also required for complex II formation (figure 4.14b).

Since the two cytosine residues at positions -72 and -73 form the core of the CME, they were mutated to AA in oligonucleotide MUT-CME2 and used as a probe in EMSA. Figures 4.18 and 4.19 shows that the formation of complexes II and III (11 ± 5% and 1.5 ± 1.5% binding respectively) was virtually abolished with this oligonucleotide. Since these two residues form part of the G/CBE, it was not surprising that a 34 ± 11% reduction complex I formation was observed.

The formation of complex I was abolished or drastically reduced when the single cytosine spacer nucleotide between the inverted upstream GGAGG and CCAAT boxes was removed (figures 4.20 and 4.21). This mutated oligonucleotide was also unable to compete out SVWI-38 or CT-1 complex I formation in competition binding assays (data not shown).

To summarise, mutations in the upstream GGAGG box, the CCAAT box and the cytosines at positions -72 and -73 resulted in reduction in the binding of G/CBF to the G/CBE (complex I). The integrity of these three motifs in the G/CBE was essential for G/CBF binding. The single cytosine nucleotide separating the inverted upstream GGAGG and CCAAT boxes also appeared to be important for complex I formation, while no reduction of complex I formation was observed when the downstream GGAGG box was mutated. It is possible that the single cytosine deletion is so deleterious because is destroys the spatial arragement between the GGAGG and CCAAT boxes that are important for complex formation. Mutations in the downstream GGAGG box and nucleotides -72 and -73 resulted in significant reduction in complex II and III formation on the CME. These mutation results also suggested that sequences upstream of the CME are involved in the formation of complexes on the collagen promoter.

4.2.6 Delineation of G/CBE and CME

Competition assays showed that oligonucleotide D was unable to compete out complex II and III (figures 4.2 and 4.3) formation in SVWI-38 and CT-1 extracts. Secondly, only trace

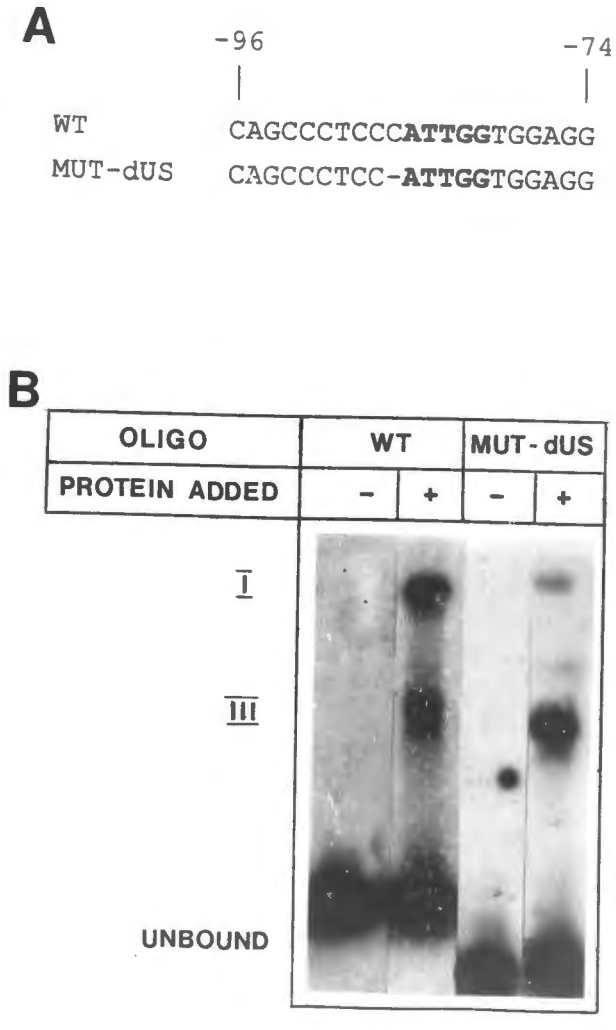


Fig 4.20 Deletion of a single cytosine residue in the G/CBE abolishes complex I formation in CT-1 nuclear extracts. **(A)** Double-stranded mutated (MUT-dUS) and wild-type (WT) oligonucleotides containing the G/CBE used in this study. The single cytosine residue between the upstream GGAGG and CCAAT boxes was deleted in MUT-dUS. **(B)** 3-4 ug of crude CT-1 nuclear extracts were incubated with 1 ng (10^4 cpm) of labelled probe (+) or without added proteins (-) and electrophoresed on 5% non-denaturing polyacrylamide gels as described in section 7.4.2. The gels were fixed in 10% acetic acid for at least 20 min, dried and exposed to X-ray film for 16 hrs. The positions of complexes I and III are indicated.

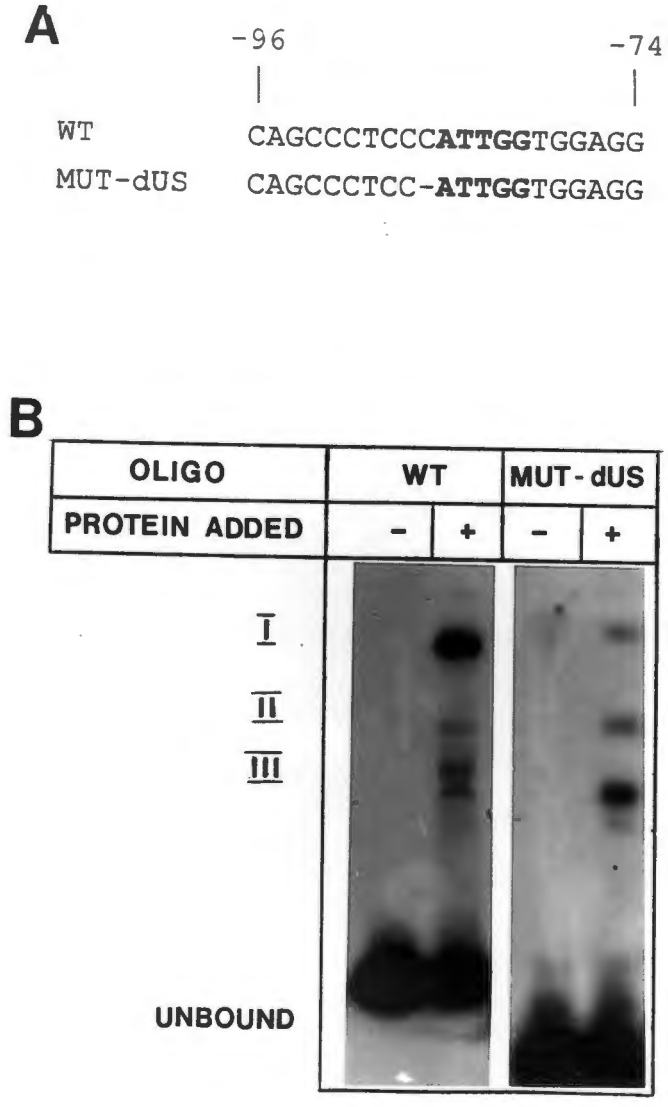


Fig 4.21 Deletion of a single cytosine residue in the G/CBE abolishes complex I formation in SVWI-38 nuclear extracts. **(A)** Double-stranded mutated (MUT-dUS) and wild-type (WT) oligonucleotides containing the G/CBE used in this study. The single cytosine residue between the upstream GGAGG and CCAAT boxes was deleted in MUT-dUS. **(B)** 3-4 ug of crude SVWI-38 nuclear extracts were incubated with 1 ng (10^4 cpm) of labelled probe (+) or without added protein (-) and electrophoresed on 5% non-denaturing polyacrylamide gels as described in section 7.4.2. The gels were fixed in 10% acetic acid for at least 20 min, dried and exposed to X-ray film for 16 hrs. The positions of complexes I, II and III are indicated.

amounts of complex III proteins bound to oligonucleotide D (figure 4.7). Since the methylation interference data showed that this oligonucleotide contains the minimal DNA-element (-78 to -66) for complex II and III formation, these *trans*-acting factors are expected to bind to this oligonucleotide (figure 4.15). In order to test the possibility that non-specific 5'-flanking sequences were sufficient for complex II and III formation, the double-stranded oligonucleotide D was cloned into the Sma 1 site of pUC19 to produce plasmid pNJ-D. This insert contains a 33 bp fragment between nucleotides -78 to -46 containing the downstream GGAGG box and 3'-flanking sequences (figures 4.22c and 4.23). The insert was released with Eco R1 and Hind III so that pUC19 sequences were attached to the probe fragments used in EMSA. Figures 4.24 and 4.25 show that complexes II and III formed with such a probe containing non-specific pUC19 polylinker sequences.

Several lines of evidence suggested that the addition of non-specific 3'-flanking sequences to oligonucleotide B may increase the binding of complex I proteins (G/CBF) to this oligonucleotide:- (1) The oligonucleotide contains the most important motifs in the G/CBE for complex I formation, namely the inverted distal GGAGG and CCAAT boxes; (2) mutations in the downstream GGAGG box did not cause a reduction in complex I formation (figures 4.18 and 4.19); (3) trace amounts of complex I was detected on longer exposures of autoradiograms when oligonucleotide B was used as a probe in electrophoretic mobility shift assays (figures 4.7 and 4.8). The double-stranded oligonucleotide was therefore cloned into the Sma 1 site of pUC19 to produce plasmid pNJ-B (figures 4.22b and 4.23). Complex I formation was observed when a Hind III - Eco R1 fragment of plasmid pNJ-B was used as a probe with CT-1 (figure 4.24) or SVWI-38 (figure 4.25) nuclear extracts in EMSA. The adjacent pUC19 sequences are not homologous to the deleted promoter sequences, and therefore could not be crucial in the formation of complex I.

Plasmid pNJ-F contains a 50 bp fragment extending from nucleotides -110 to -60 of the collagen promoter (figures 4.22a and 4.23) which includes the G/CBE and CME responsible for complex I, II and III formation. As expected complexes I, II and III in SVWI-38 nuclear extracts (figure 4.25) and

A

aagcttgcat gcctgcaggt cgactctaga ggatccCCCG GGCCCCCAGC

Hind III — Bam H1

CCCAGCCCTC CCATTGGTGG AGGCCCTTTT GGAGGCgggt accgagctcg

Kpn 1

aattc

Eco R1

Plasmid pNJ-F

B

aagcttgcat gcctgcaggt cgactctaga ggatccccGC CCCCAGCCCC

Hind III — Bam H1

AGCCCTCCCA TTGGTGGgta ccgagctcga attc

Kpn 1 — Eco R1

Plasmid pNJ-B

C

gaattcgagc tcggtacccG GAGGCCCTTT TGGAGGCACC CTAGGGCCAG

Eco R1 — Kpn 1

GGgggatcct ctagagtcga cctgcaggca tgcaagctt

Bam H1 — Hind III

Plasmid pNJ-D

Fig 4.22 Nucleotide sequences of promoter inserts in plasmids pNJ-F (A), pNJ-B (B) and pNJ-D (C). The three plasmids were cloned by ligating phosphorylated double-stranded oligonucleotides F, B and D (figure 4.7) into the Sma 1 site of pUC19 as described in section 7.1.2.3. The polylinker of pUC19 is in lowercase letters while the oligonucleotide sequences are shown as uppercase letters. Various restriction sites are also indicated.

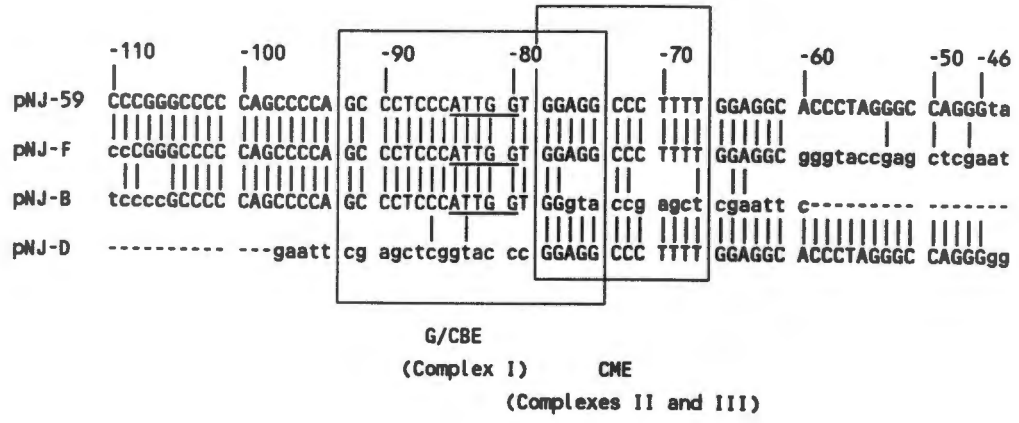


Fig 4.23 Sequence comparison of plasmid pNJ-59 with plasmids pNJ-F, pNJ-B and pNJ-D. Plasmid pNJ-59 (section 2.2.3.1), containing a 65 bp fragment (nucleotides -110 to -46) of the human α2(I) procollagen promoter is compared with the corresponding regions of plasmids pNJ-F, pNJ-B and pNJ-D. The G/CBE and CME are boxed while the inverted CCAAT box is underlined. Sequences originating from the pUC19 polylinker are shown as lowercase letters. Alignment with the pNJ-59 sequence are indicated by the vertical line.

complexes I and III in CT-1 nuclear extracts (figure 4.24) formed with this fragment. An additional complex, which migrates slightly faster than complex I, was observed when this fragment was used as a probe. The identity of this complex will be discussed in chapter 5.

These results supported the hypothesis that there are distinct adjacent DNA-binding elements in this region of the human α2(I) procollagen promoter. The upstream GGAGG and inverted CCAAT boxes are the most important components of the G/CBE because mutations in the downstream GGAGG box or replacing it with non-homologous sequences does not abolish binding of complex I proteins (G/CBF) to the G/CBE. Finally, non-homologous 5'-flanking sequences are sufficient for complex II and III proteins to bind the CME. In addition, the elements responsible for complex I and II/III formation have been separated.

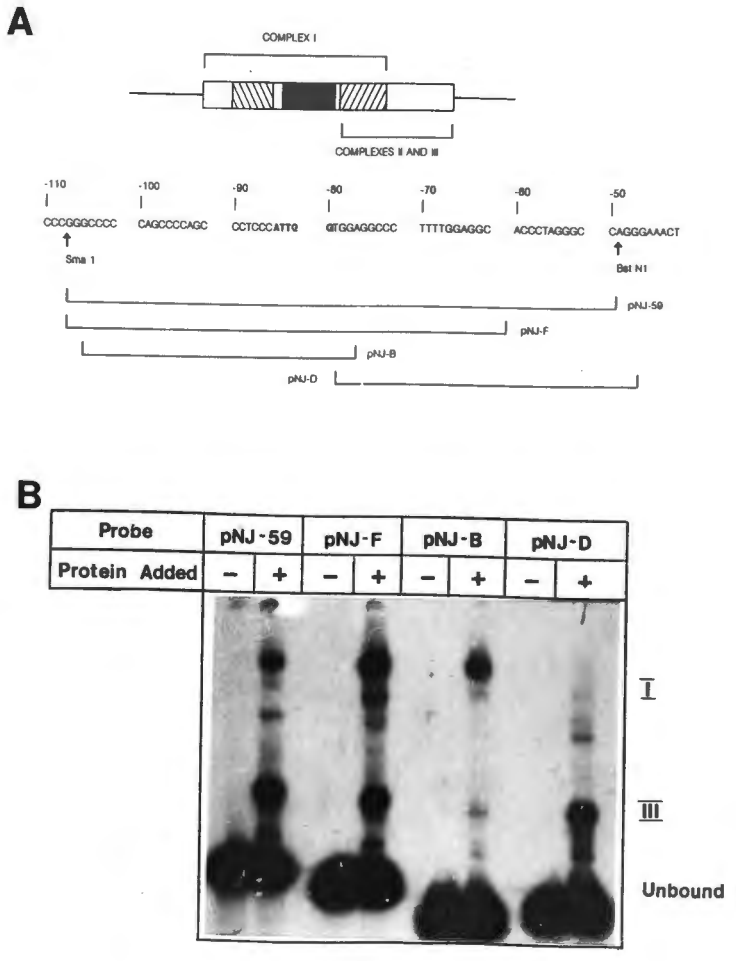


Fig 4.24 Delineation of G/CBE and CME. **(A)** A schematic diagram of the G/CBE (complex I) and CME (complex II and III). The inverted CCAAT box is shown in bold, while the GGAGG boxes are hatched. The nucleotide sequence of a 60 bp fragment extending from position -110 to -41 of the human α2(I) procollagen promoter is also shown. The cloned inserts in plasmids pNJ-59, pNJ-F, pNJ-B and pNJ-D are indicated. **(B)** Hind III - Eco R1 fragments of the four plasmids were prepared and radioactively labelled as described in sections 7.1.1.1 and 7.4.1. 3-4 ug of crude CT-1 nuclear extracts were incubated with 1 ng (10^4 cpm) labelled probe (+) or without added protein (-) and electrophoresed on 5% non-denaturing polyacrylamide gels as described in section 7.4.2. The gels were dried and exposed to X-ray film for 16 hrs. The positions of complexes I and III are indicated. The identity of the extra band which migrated slightly faster than complex I will be discussed in chapter 5.

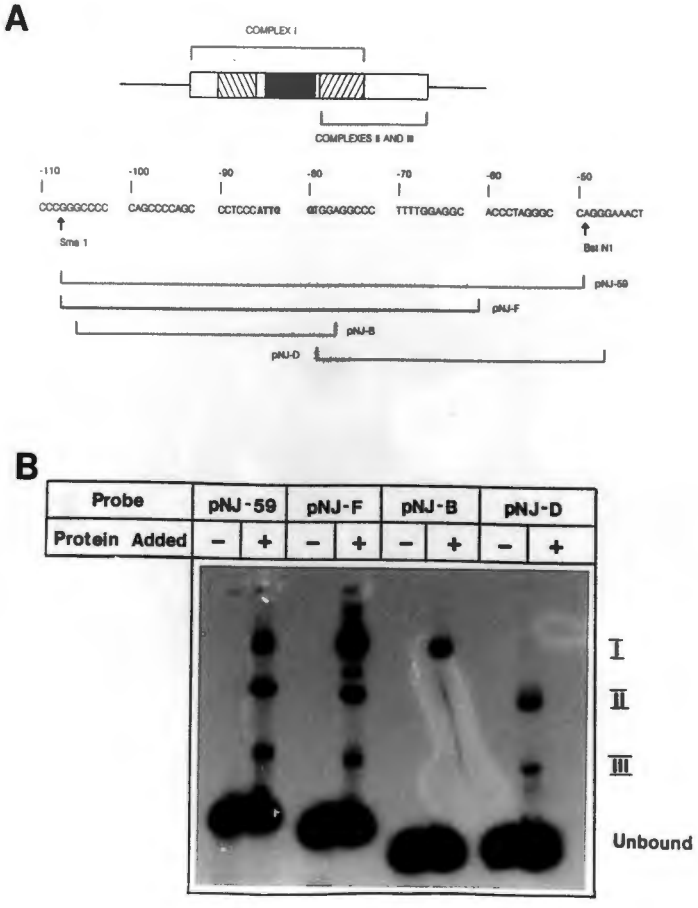


Fig 4.25 Delineation of G/CBE and CME. **(A)** A schematic diagram of the G/CBE (complex I) and CME (complex II and III). The inverted CCAAT box is shown in bold, while the GGAGG boxes are hatched. The nucleotide sequence of a 60 bp fragment extending from position -110 to -41 of the human α2(I) procollagen promoter is also shown. The cloned inserts in plasmids pNJ-59, pNJ-F, pNJ-B and pNJ-D are indicated. **(B)** Hind III - Eco R1 fragments of the four plasmids were prepared and radioactively labelled as described in sections 7.1.1.1 and 7.4.1. 3-4 ug of crude SVWI-38 nuclear extracts were incubated with 1 ng (10^4 cpm) labelled probe (+) or without added protein (-) and electrophoresed on 5% non-denaturing polyacrylamide gels as described in section 7.4.2. The gels were dried and exposed to X-ray film for 16 hrs. The positions of complexes I, II and III are indicated. The identity of the extra band which migrated slightly faster than complex I will be discussed in chapter 5.

4.3 DISCUSSION

DNA-protein complexes which form on a 60 bp fragment of the human α2(I) procollagen promoter have been identified in nuclear extracts prepared from human fibroblasts. In a SV40-transformed WI-38 cell line in which no α2(I) procollagen chains are produced, an additional complex (complex II) was identified. Several lines of evidence suggested that the SVWI-38 specific complex II is a transcriptional repressor, which is partially or totally responsible for the inactivation of the α2(I) procollagen gene in this cell line (Chapter 2). Competition, electrophoretic mobility shift and methylation interference assays were used to identify two distinct but overlapping DNA-recognition elements within this region of the collagen promoter. Complex I, II and III proteins were shown to bind to these overlapping *cis*-elements.

Complex I proteins complexed with a 21 bp element extending from position -92 to -72, named the G/CBE. The methylation interference assays, mutation analysis and binding studies suggested that the inverted upstream GGAGG and the CCAAT boxes were the most important motifs for complex I formation. Deletion of a single cytosine which separates these two motifs also abolished or drastically reduced the formation of complex I. Although the methylation interference patterns showed that the downstream GGAGG box formed part of the recognition element, mutation or replacement of this motif did not significantly reduce complex I formation.

Several distinct factors which bind to *cis*-elements containing a CCAAT box have been identified and this complex may well be a member of this family of CCAAT box binding factors. The complex I protein(s) has therefore been called a GGAGG/CCAAT box factor (G/CBF) which could be different from, related to or identical to previously identified CCAAT box binding factors (section 1.4). The specificity of a particular CCAAT box binding protein for a *cis*-element containing a CCAAT motif is probably determined by the flanking sequences. Comparison of the human G/CBE with the mouse CBF recognition element in the α2(I) procollagen promoters showed that these elements are highly homologous

(figure 4.16) (Hatamochi *et. al.*, 1986; Maity *et. al.*,1988). Secondly, Framson and Bornstein (1993) have identified a NF-Y binding site in the human thrombospondin 1 (TSP-1) promoter which closely resembles the G/CBE in the human α2(I) procollagen promoter (figure 4.26). The TSP-1 NF-Y binding site contains an upstream inverted GGAGG-like box with a single base mismatch, an inverted CCAAT box and a downstream GGAGG box. The functionally important single cytosine spacer between the upstream GGAGG and CCAAT motifs is conserved in both promoters. These observations strongly suggest that G/CBF is identical or related to CBF (figure 4.16) and NF-Y and is therefore a member of the hetero-multimeric family of CCAAT box binding proteins (section 1.4.3).

```
    AACATTTTTCTGATTGGTTAAAAGTT    Eα
      ||    ||  ||||||   | |
AGCCCCAGCCCTCCCATTGGT-GGAGGCCC    α2(I) Procollagen
  ||      || ||||||||  |||||
GCCCGGCCGCCGCCCATTGGCCGGAGGAAT    Thrombospondin 1
```

Fig 4.26 Comparison of the G/CBE with NF-Y sites. The G/CBE of the human α2(I) procollagen promoter was aligned with the NF-Y sites of the thrombospondin 1 promoter (Framson and Bornstein, 1993) and the Eα promoter (MHC class II gene) (Dorn *et. al.*, 1987). The inverted CCAAT boxes are in bold, the GGAGG boxes are underlined, while the GGAGG-like box is doubly underlined.

CBF (Karsenty *et. al.*, 1988) and NF-Y (Dorn *et. al.*, 1987), however, bind to their respective *cis*-elements only if they contain an intact CCAAT motif. Since mutation of the CCAAT box in the G/CBE did not totally abolish G/CBF binding to this element, it is possible that G/CBF is not identical to CBF and NF-Y as suggested from their respective recognition elements. Since it was not possible to accurately determine the relationship of G/CBF with other CCAAT box binding factors from these experiments, the characterisation (chapter 5) and eventual purification of this factor will throw some light on this matter.

The methylation interference patterns showed that complex II and III proteins bound to the same 12 bp *cis*-element between -78 and -67 (the CME). The CME ($^{5'}$**GGAGG**CCCTTTT$^{3'}$) contains

the downstream GGAGG box, which loosely forms part of the G/CBE, and 3'-flanking sequences. Mutations of the GGAGG box and the cytosine dinucleotide at positions -72 and -73 resulted in a drastic reduction or abolition of complex II and III formation. A reverse NF-kB-like consensus sequence, $^{5'}$GDRRADYCCC$^{3'}$, (Lenardo and Baltimore, 1989) containing a single base mismatch is located between nucleotides -80 and -71. The core sequence of the CME is also located in this region of the promoter, suggesting that the CME could be a NF-kB-like recognition element.

Another possibility is that the protein component of complex II was SV40-large T antigen. Firstly, Butel and Jarvis (1986) have shown that SV40 large T antigen is located in the nucleus of SV40-transformed cells, and secondly, the CME overlaps a large T antigen-like recognition sequence present in the α2(I) procollagen promoter (Ghosh *et. al.*, 1981). Smith (1989) has shown that large T antigen is not responsible for complex II formation by competition gel shift analysis and by supershifting using an anti-large T antibody. The relationship between complexes II and III proteins and other DNA-binding proteins, like NF-kB, will be discussed in chapter 5.

The reconstitution assays demonstrated that complex I, II and III proteins bound to their respective sites in a mutually exclusive manner. A possible mechanism for the repression of α2(I) procollagen gene expression by complex II proteins could be via competition with complex III proteins for binding to the CME. Binding of complex II to the CME also prevented G/CBF from binding to its recognition element. Barberis *et. al.* (1987) have identified a candidate repressor, designated the CCAAT displacement factor (CDF), which does not bind simultaneously with the CCAAT binding factor to their respective overlapping elements in the sea urchin histone H2B-1 promoter. CBF inhibits the binding of inhibitory factor 2 (IF2) to its recognition sequence in the mouse α1(I) procollagen promoter (Karsenty and de Chrombrugghe, 1989). Recently, Liu *et. al.* (1993) have demonstrated that the chicken ovalbumin transcription factor (COUP-TF) acts as a competitive repressor of the mouse lactoferrin gene. This gene, which is activated in response

to oestrogen stimulation, contains an overlapping COUP-TF and oestrogen receptor binding elements in its promoter.

These results also indicate that different *trans*-acting factors are involved in the regulation of the human and mouse α2(I) procollagen gene. The CME is not present in the mouse promoter (figure 4.16) and it is also not clear if the mouse CBF is related to G/CBF. It is therefore possible that there are different species-dependent mechanisms responsible for regulating the expression of this gene. Certain recognition elements in the human, hamster and chicken TK promoters are not present in the mouse promoter (Arcot *et. al.*, 1989; Lieberman *et. al.*, 1988). The chicken and human α2(VI) collagen promoters are also structurally different (Koller *et. al.*, 1991; Saitta *et. al.*, 1992), suggesting that species-dependent mechanisms may be operative in the regulation of this gene.

In summary, two distinct overlapping regulatory elements, the GGAGG/CCAAT box element (G/CBE) and collagen modulating element (CME), have been identified in the -80 bp region of the human α2(I) procollagen promoter. Collagen producing cell lines contain only one type of CME binding complex, whereas SVWI-38 cells contain two. Evidence indicates that the second SVWI-38 complex acts as a repressor of collagen synthesis (V. Leaner, unpublished data).

CHAPTER 5

CHARACTERIZATION OF THE PROMOTER BINDING PROTEINS

5.1 INTRODUCTION

Based on their structure, sequence-specific DNA-binding proteins can be divided into two broad categories. The first category includes transcription factors comprising a single polypeptide containing both DNA-binding and regulatory domains. Factors such as Sp1 (Kadonaga *et. al.*, 1987), Oct1 (Sturm *et. al.*, 1988) and Oct2 (Clerc *et. al.*, 1988) can bind to their recognition sequences as monomers and/or homomultimers. Alternatively, two or more distinct polypeptides may interact with each other to produce a functionally active factor, making up the second category of transcription factors. The fos-jun heterodimer, which binds to the AP-1 consensus sequence is a well known example of this hetero-multimeric category of transcription factors (Chiu *et. al.*, 1988).

Before attempting to purify a candidate novel sequence-specific transcription factor or clone its gene, it is advisable to first determine if it is homologous or related to previously characterised factors. This is essential because; (1) a large number of transcription factors have been purified and characterised to date; (2) structurally different factors have been shown to bind to the same or related promoter elements (section 1.4) and (3) the same protein can bind to apparently unrelated elements (Pfeifer *et. al.*, 1987a). A variety of techniques, including analytical gel filtration (Siegel and Monty, 1966), density gradient centrifugation (Martin and Ames, 1961), UV-crosslinking (Chodosh *et. al.*, 1986), Southwestern blotting (Bowen *et. al.*, 1980) and protein-mediated Southwestern blotting (Matsuno *et. al.*, 1989), have been developed and

used extensively to characterise DNA-binding proteins in crude extracts.

Physico-chemical characteristics include Stokes radii and sedimentation coefficients using gel filtration and sucrose or glycerol gradient centrifugation respectively. An accurate determination of both the molecular weight and frictional ratio can be calculated from these parameters if the partial specific volume of the protein can be reasonably estimated. This approach has been used to determine the native molecular mass of a murine CCAAT box binding factor, α-CP1 (Kim *et. al.*, 1988 and Kim and Sheffery, 1990).

The characterisation of *trans*-acting factors which bind to the GGAGG/CCAAT box (G/CBE) and collagen modulating (CME) elements is described in this chapter. The techniques described above were used to determine the molecular weights and structural complexities of these proteins in crude or partially purified SVWI-38 or CT-1 nuclear extracts. The results were compared with published data on the transcription factors in order to determine whether the G/CBE or CME binding proteins were related to any of the previously characterised factors.

5.2 RESULTS

5.2.1 NATIVE MOLECULAR WEIGHT DETERMINATION

Analytical gel filtration (Siegel and Monty, 1966) on Sephacryl S-300 and glycerol gradient sedimentation (Martin and Ames, 1961) were used to determine the Stokes radii and sedimentation coefficients of complex I, II and III proteins in crude nuclear extracts under non-denaturing conditions.

5.2.1.1 ANALYTICAL GEL FILTRATION

Sephacryl S-300 gel filtration columns were packed and calibrated and used for nuclear extract fractionation as described in Materials and Methods (7.6.1). The electrophoretic mobility shift assay (EMSA) was used to identify DNA-binding activity in the various fractions using fragment E (see section 2.2.4) as a probe (7.4.2).

In the absence of a detergent such as nonident P-40 (NP-40), no separation of the complex I, II and III proteins was achieved (figure 5.1). This was probably due to aggregation of proteins, since the inclusion of 0.1% NP-40 to the samples resulted in good separation of the complexes (compare figure 5.1 with figures 5.2a and 5.3b). It was not necessary to add detergent to the elution buffer, addition to the samples prior to loading sufficed (Kim and Sheffery, 1990).

As expected the larger complex I proteins eluted first followed by complex II proteins and finally the protein component(s) of complex III (figure 5.3). Identical elution volumes were obtained for the proteins involved in the common SVWI-38 and CT-1 complexes (figures 5.2 and 5.3). The trace amounts of complex II proteins, which were present in some batches of CT-1 extracts also eluted from the column with elution volumes similar to SVWI-38 complex II proteins.

The Kav of each complex was calculated from the elution (Ve), void (Vo) and total (Vt) volumes as summarised in table 5.1. The Stokes radii were determined using a semi-logarithmic plot of Stokes radius versus Kav (figure 5.8a). Under non-

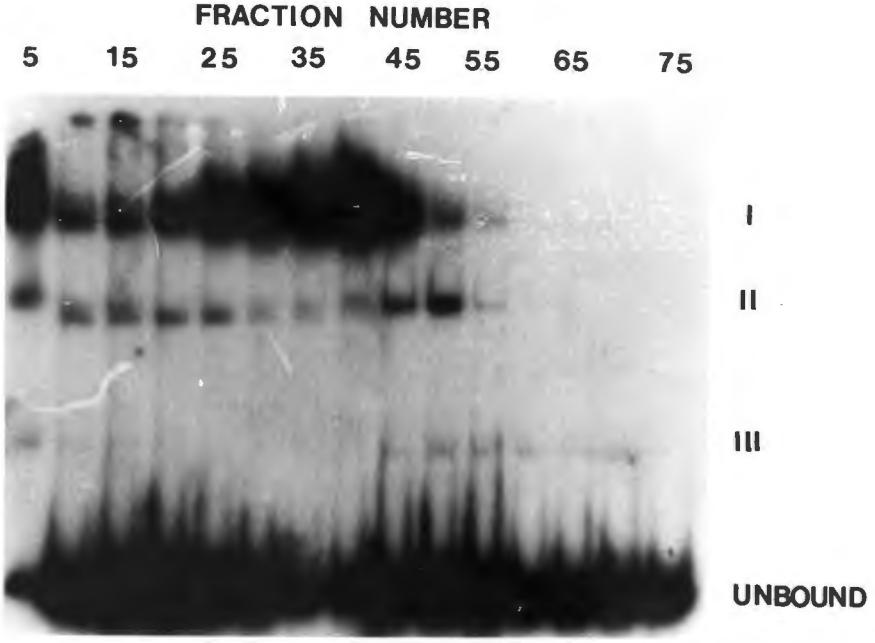


Fig 5.1 Fractionation of nuclear extracts by Sephacryl S-300 gel filtration. Crude SVWI-38 nuclear extracts were fractionated on a 1.6 cm X 100 cm column of Sephacryl S-300 in buffer CB containing 30% glycerol and 0.1 M KCl (section 7.6.1). The DNA-binding activity in 10_4 ul of every 5th fraction was assayed using 1 ng (10^4 cpm) of end-labelled fragment E as a probe (7.4.2). The samples were electrophoresed on 5% non-denaturing polyacrylamide gels at 4^{O} C using 0.5 X TBE. The dried gels were exposed to X-ray film for 16 hrs. The positions of complexes I, II and III are indicated.

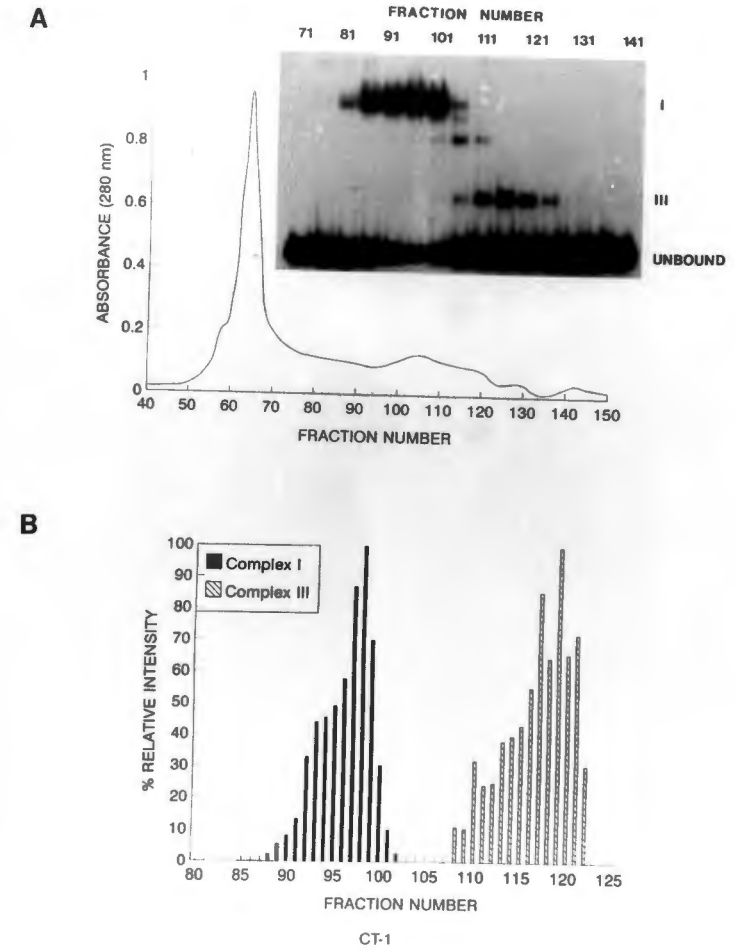


Fig 5.2 Analytical gel filtration of CT-1 extracts on Sephacryl S-300 in the presence of NP-40. (A) Crude nuclear extracts were applied to the column in buffer CB containing 0.1 M KCl, 30% glycerol and 0.1% NP-40 and fractionated with CB containing 0.1 M KCl at a flow rate of 3 ml/cm^2/h. One ml fractions were collected and their DNA-binding activity was assayed as described in figure 5.1. . The DNA-binding activity of every 5th fraction is shown in the inset where the positions of complexes I and III are indicated. The trace amounts of complex II activity in this batch of CT-1 extract is clearly visible. **(B)** Histograms of the relative DNA-binding activities in each fraction. The autoradiograms were scanned and the results for complexes I and III expressed as % relative intensity of each fraction.

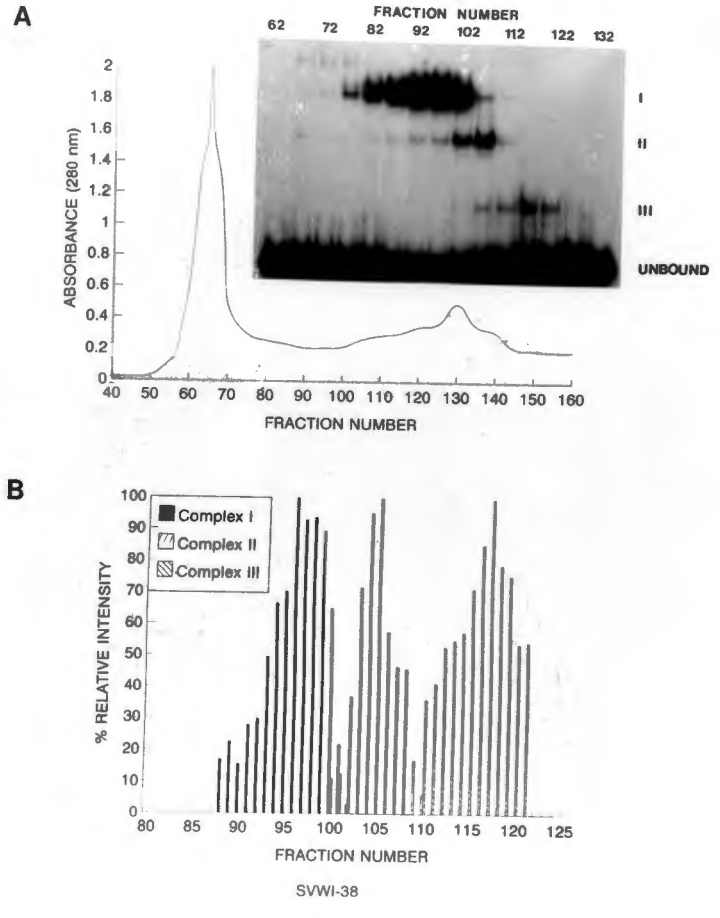


Fig 5.3 Analytical gel filtration of SVWI-38 extracts on Sephacryl S-300 in the presence of NP-40. (A) Crude nuclear extracts were applied to the column in buffer CB containing 0.1 M KCl, 30% glycerol and 0.1% NP-40 and processed as described in the legend to figure 5.1. **(B)** Histograms of the relative DNA activities in the collected fractions. The autoradiograms were scanned and the results for complexes I, II and III expressed as % relative intensity of each fraction.

TABLE 5.1 Analytical gel filtration of SVWI-38 and CT-1 extracts. A Sephacryl S-300 column was packed and calibrated as described in section 7.6.1. The Stokes radii (R_s), molecular weights (MW), elution volumes (V_e), calculated K_{av} and the standard deviations of n determinations are tabulated for the protein standards. The void (V_o) and total (V_t) volumes were determined using blue dextran 2000 and thymidine respectively. Crude nuclear extracts were fractionated as described in section 7.6.1. The electrophoretic mobility shift assay was used to quantitate the DNA-binding activity in each fraction using fragment E (nucleotides -110 to -46). The K_{av} of each complex was calculated from their V_e, the V_o and V_t and the Stokes radii determined from the semi-logarithmic plot of Stokes radius versus K_{av}.

STANDARDS/SAMPLES	Ve (ml)	K_{av}	R_s (nm)	MW (KDa)	n
STANDARDS					
Thyroglobin	76.8 ± 0.5	0.113 ± 0.004	8.5	669	2
Ferritin	87.5 ± 1.5	0.211 ± 0.014	6.1	440	3
Catalase	99.4 ± 1.0	0.320 ± 0.009	5.22	232	3
Aldolase	103.4 ± 0.3	0.357 ± 0.003	4.81	158	3
BSA	113.7 ± 1.3	0.451 ± 0.012	3.55	68	2
Ovalbumin	123.0 ± 0.1	0.537 ± 0.001	3.05	43	2
Blue Dextran (Vo)	64.5 ± 0.8	-	-	-	10
Thymidine (Vt)	173.5 ± 0.2	-	-	-	2
SAMPLES					
Complex I	99.5 ± 0.3	0.321 ± 0.003	5.22 ± 0.04	232	4
Complex II	108.6 ± 0.3	0.405 ± 0.003	4.12 ± 0.03	107	3
Complex III	120.7 ± 0.1	0.516 ± 0.001	3.15 ± 0.01	50	2

denaturing conditions, complex I and III proteins from SVWI-38 and CT-1 extracts eluted from the column with Stokes radii of 5.22 ± 0.04 nm and 3.15 ± 0.01 nm respectively, while the SVWI-38 complex II factor had a Stokes radius of 4.12 ± 0.03 nm. The Stokes radii of complexes I, II and III were comparable to spherical proteins of 232, 107 and 50 KDa respectively. The protein distribution of selected SVWI-38 fractions involved in either complex I, II or III formation (labelled pools I, II or III respectively in figure 5.4) were checked on 10% SDS-polyacrylamide gels after they had been concentrated by acetone precipitation as described in section 7.5. As shown in figure 5.4 these fractions still contained large amounts of contaminating polypeptides.

5.2.1.2 GLYCEROL GRADIENT SEDIMENTATION

The sedimentation coefficients ($S_{20,W}$) of the three complexes were determined on linear 15 to 40% glycerol gradients as described in section 7.6.2. The linearity of the gradients were checked by adding 100 ul 0.4% phenol red to the 40% glycerol solution before pouring the gradient. After centrifugation, fractions were collected (section 7.6.2) and the relative concentration of phenol red in the fractions was measured at 400 nm (Siegel and Monty, 1966). As shown in figure 5.5 the phenol red and therefore the glycerol concentration was found to be linear.

About half a milligram of crude nuclear extracts were centrifuged through the gradients using 0.5 mg ovalbumin as an internal standard as described in Materials and Methods (7.6.2). The DNA-binding activity in the collected fractions were quantitively measured by EMSA with fragment E as a probe (7.4.2). The percentage glycerol in the fractions where complex I, II and III activities peaked was calculated for both CT-1 (figure 5.6) and SVWI-38 (figure 5.7) extracts and are tabulated (table 5.2). Sedimentation coefficients of 4.9 ± 0.4, 3.9 ± 0.1 and 3.1 ± 0.3 were calculated for the protein components of complexes I, II and III respectively (figure 5.8b).

The following equation was used to estimate the molecular weights of the three factors from the sedimentation data

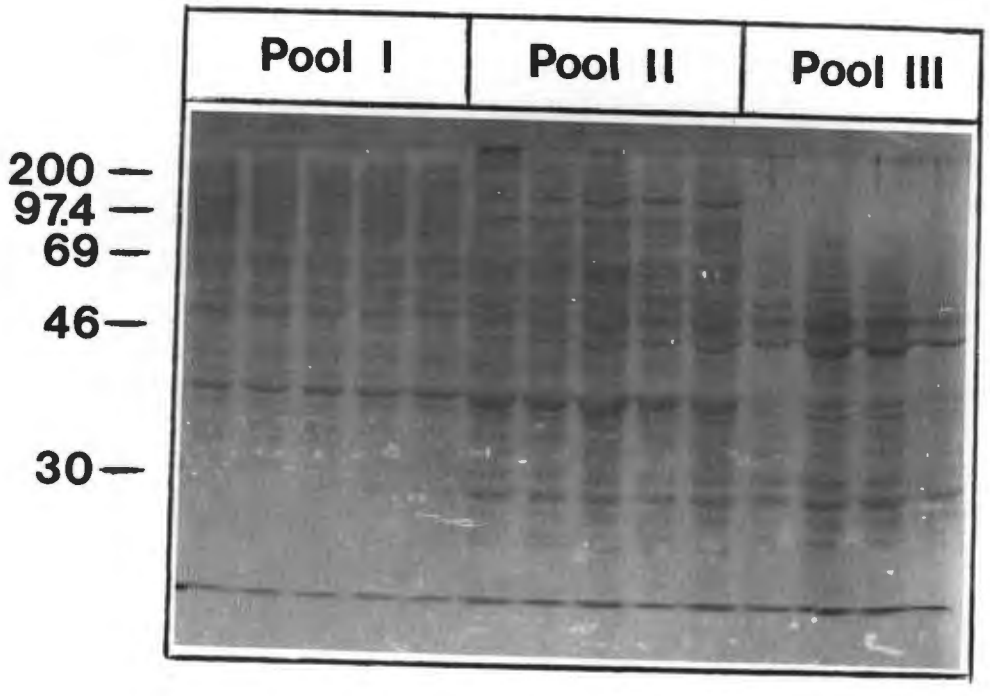


Fig 5.4 SDS-PAGE analysis of SVWI-38 extracts fractionated by Sephacryl S-300 gel filtration. Crude nuclear extracts were fractionated on Sephacryl S-300 columns as described in figure 5.3 and section 7.6.1. Two hundred ul (30 to 40 ug protein) of selected fractions, 92 to 96 (pool I), 105 to 109 (pool II) and 115 to 118 (pool III) containing complex I, II and III DNA-binding activities respectively, were concentrated by acetone precipitation. The proteins were resolved on a 10% SDS-polyacrylamide gel and stained with Coomassie brilliant blue as described in Materials and Methods (section 7.5). The molecular weights are indicated in KDa.

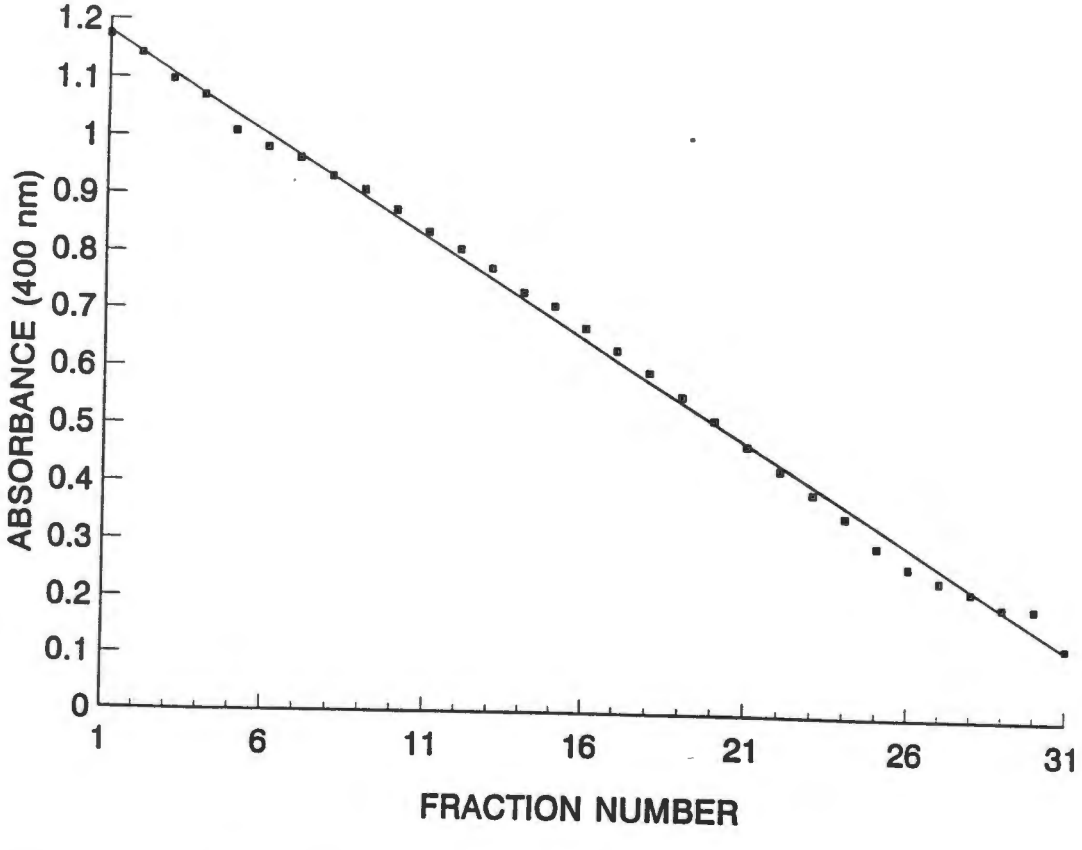


Fig 5.5 Linearity of the glycerol gradients. One hundred ul of 0.4% phenol red was added to the 40% glycerol solution prior to preparing gradients. After centrifugation , the fractions were collected as described in section 7.6.2 and the intensity of phenol red in the fractions measured at 400 nm. The plot of fraction number versus absorbance was linear, indicating the linearity of the gradients.

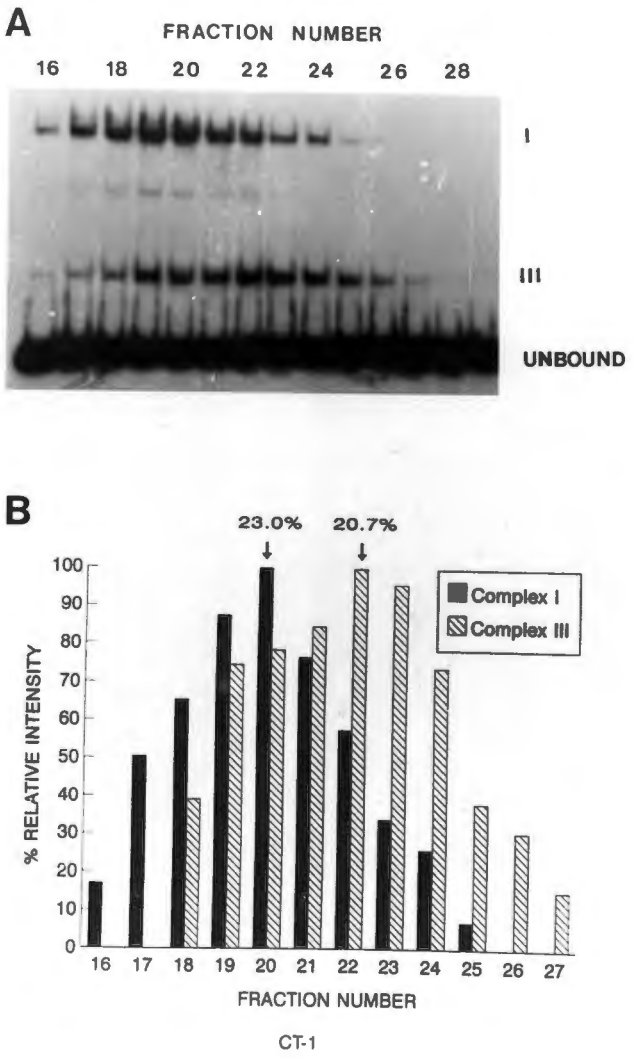


Fig 5.6 Glycerol gradient sedimentation of CT-1 nuclear extracts. **(A)** Half a milligram of crude nuclear extract and 0.5 mg ovalbumin (internal standard) in a final volume of 100 ul was layered onto 15 to 40% linear glycerol gradients and centrifuged for 16 hrs at 53 000 rpm and 4^{o} C in a Beckman SW65 rotor. Three-drop fractions were collected from the bottom of each centrifuge tube and the DNA-binding activity in 10 ul of each fraction was quantitated by the EMSA as described in section 7.4.2. The positions of complexes I and III are indicated. **(B)** Histograms of the relative DNA-binding activities in each fraction. The autoradiograms were scanned and the relative intensity of each fraction was plotted. The percentage glycerol in the fractions where complex I and III DNA-binding activities peaked was calculated as described in section 7.6.2.

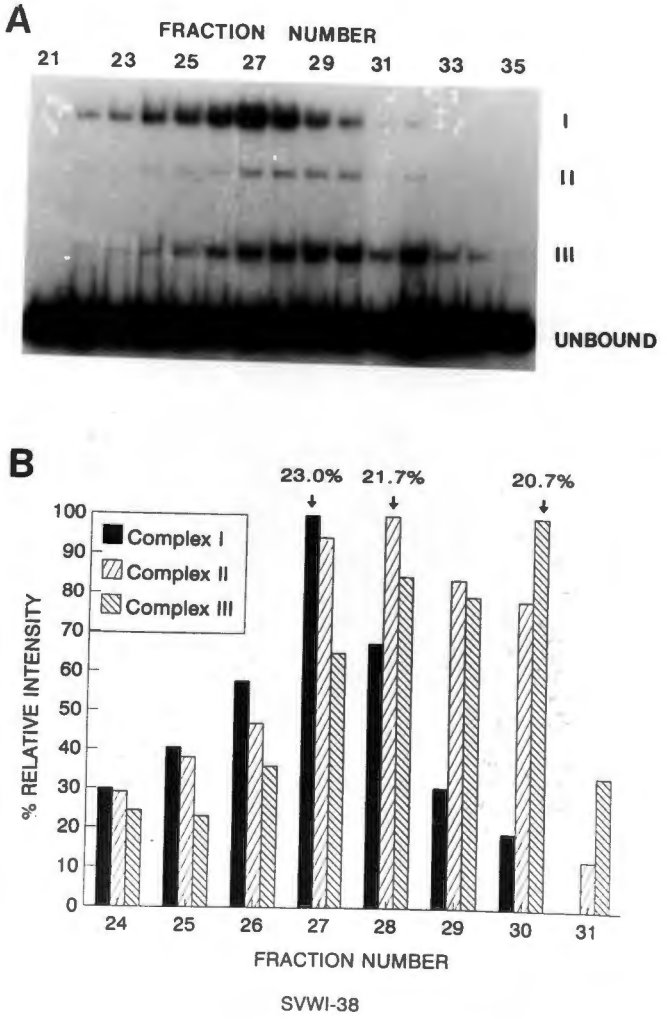


Fig 5.7 Glycerol gradient sedimentation of SVWI-38 nuclear extracts. **(A)** Half a milligram of crude nuclear extract and 0.5 mg ovalbumin (internal standard) in a final volume of 100 ul was layered onto the gradients, centrifuged, fractions collected and assayed as described in the legend to figure 5.6. **(B)** Histograms of the relative DNA-binding activities in each fraction. The autoradiograms were scanned and the relative intensity of each fraction was plotted. The percentage glycerol in the fractions where the DNA-binding activity of the three complexes peaked was calculated as described in section 7.6.2.

Table 5.2 Glycerol gradient sedimentation of SVWI-38 and CT-1 nuclear extracts. Linear 15 to 40% glycerol gradients were prepared and calibrated as described in section 7.6.2. The sedimentation coefficients ($S_{20,W}$) molecular weights (MW), percent glycerol and the standard deviations of n determinations are tabulated for the protein standards. The percentage glycerol in the fractions were the protein standards and samples peaked were calculated as described in section 7.6.2. Crude nuclear extracts and a protein internal standard were centrifuged and fractions collected as described in Materials and Methods (7.6.2). DNA-binding activity in each fraction was quantitated bu EMSA using fragment E as a probe. The molecular weights of the complexes were estimated from the sedimentation data as described in section 5.2.1.2.

STANDARDS/SAMPLES	% GLYCEROL	$S_{20,W}$ (S)	MW (KDa)	n
STANDARDS				
Catalase	30.8 ± 0.4	11.3	232	6
BSA	22.9 ± 0.4	4.4	68	2
Ovalbumin	21.8 ± 0.2	3.7	43	4
Carbonic Anhydrase	20.6 ± 0.0	3.2	29	2
Lysozyme	18.9 ± 0.4	1.9	14	2
SAMPLES				
Complex I	23.0 ± 0.5	4.9 ± 0.4	80	4
Complex II	21.7 ± 0.1	3.9 ± 0.1	57	2
Complex III	20.7 ± 0.4	3.1 ± 0.2	40	4

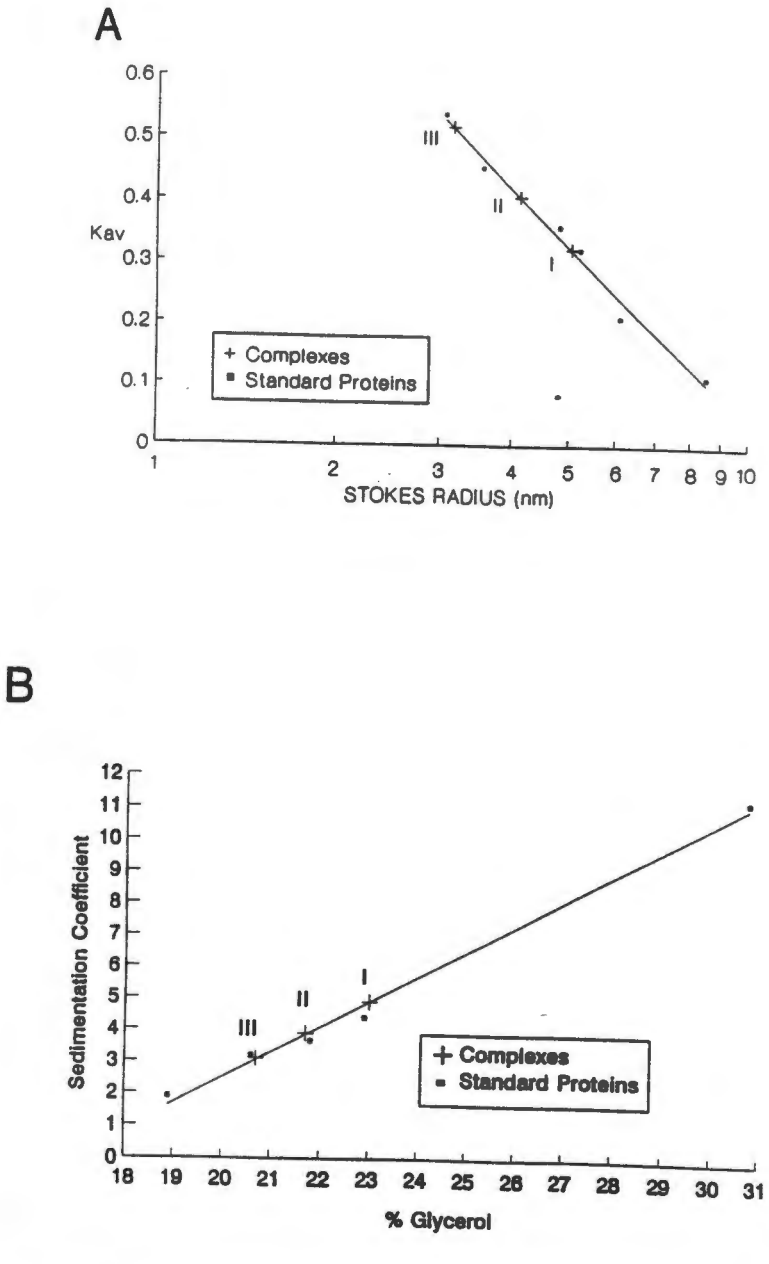


Fig 5.8 Stokes radii and sedimentation coefficient determination. Sephacryl S-300 gel filtration columns and 15 to 40% linear glycerol gradients were calibrated and used for fractionation of nuclear extracts as described in Materials and Methods (7.6.1 and 7.6.2). **(A)** The gel filtration column was calibrated using protein standards and the log of the Stokes radius (thyroglobin, 8.50 nm; ferritin, 6.10 nm; catalase, 5.22 nm; aldolase, 4.81 nm; BSA, 3.55 nm and ovalbumin, 3.05 nm) versus k_{av} for each protein was plotted, where $k_{av}=(Ve-Vo)/(Vt-Vo)$. The Stokes radii of complexes I, II and III were determined from the curve. **(B)** Glycerol gradient calibration curve was prepared by plotting the percent glycerol (see section 7.6.2) in the fractions where the standard proteins sedimented versus sedimentation coefficient (catalase, 11.3 S; BSA, 4.4 S; ovalbumin, 3.7 S; carbonic anhydrase, 3.2 S and lysozyme, 1.9 S). The sedimentation coefficients of the three complexes were determined from the curve.

alone (Martin and Ames, 1961):

$$MW_{sample} = MW_{std}(S_{sample}/S_{std})^{3/2} \qquad \text{(Equation 1)}$$

where MW_{sample} and MW_{std} are the molecular weights of the samples and protein standards and S_{sample} and S_{std} are their respective sedimentation coefficients. When BSA was used as the standard, complex I, II and III proteins had native molecular weights of 80, 57 and 40 KDa respectively.

5.2.1.3 CALCULATION OF NATIVE MOLECULAR WEIGHTS

Since there was a discrepancy when the molecular weights of the protein components in the three complexes were calculated separately from the gel filtration and sedimentation results, the native molecular weights and the frictional ratios of the proteins were determined from the Stokes radii and sedimentation coefficients using the following equations (Siegel and Monty, 1966):

$$MW = 6\pi nNas/(1-vp) \qquad \text{(Equation 2)}$$

$$f/f_o = a/(3vM/4\pi N)^{1/3} \qquad \text{(Equation 3)}$$

where MW is the molecular weight, f/f_o is the frictional ratio, n is the viscosity of water at 20^o C (1.0×10^{-3} kg/m/s) (Resnick *et. al.*, 1993), N is Avogadro's number (6.022×10^{23} mol^{-1}), a is the calculated Stokes radius, s is the calculated sedimentation coefficient (1 Svedberg unit (S) = 10^{-13} sec), v is the partial specific volume and p is the density of H_2O at 20^o C (0.998 g/ml) (Resnick *et. al.*, 1993). If the partial specific volume is assumed to be 0.725 cm^3/g (Kim and Sheffery, 1990), then the native molecular weights of complexes I, II and III under non-denaturing conditions would be 105, 66 and 40 KDa respectively. The frictional ratios of complexes I, II and III were calculated to be 1.68, 1.54 and 1.40 respectively. The ratio of the long to the short axes (axial ratio) of the factors would therefore be approximately 15, 10 and 8 if they were ellipsoid (Freifelder, 1982).

5.2.2 POLYPEPTIDE COMPOSITION OF THE COMPLEXES

In order to determine whether the factors which bound to the G/CBE and CME boxes were monomers, homo- or heteromultimers, UV-crosslinking experiments (Chodosh *et. al.*, 1986), analytical gel filtration under denaturing conditions and Southwestern blotting (Bowen *et. al.*, 1980) methodologies were employed to assay crude or partially purified SVWI-38 or CT-1 nuclear extracts.

5.2.2.1 UV-CROSSLINKING

UV-crosslinking data were used to determine the molecular weights of the polypeptide components which are in close contact with the DNA. Double-stranded, 37-mer oligonucleotide A (see figure 4.2), containing the G/CBE and 5'-portion of the CME, was substituted with bromodeoxyuridine and ^{32}P-deoxycytidine and used as a probe in electrophoretic mobility shift assays as described in Materials and Methods (7.6.3). As shown in figure 5.9, the replacement of thymidine with bromodeoxyuridine did not affect the ability of the oligonucleotide to form complex I, II or III proteins.

DNA-Protein complexes on the bromodeoxyuridine substituted DNA were resolved on non-denaturing polyacrylamide gels, UV-crosslinked in situ (Wu, C. *et. al.*, 1987), visualised by autoradiography, the bands excised, the DNA-protein complexes eluted and resolved on 10% SDS-polyacrylamide gels as described in sections 7.6.3.2 and 7.6.3.5. Complex I comprised proteins with a molecular weight of 120 KDa in both SVWI-38 and CT-1 (figure 5.10a). Occasionally two faint bands of 240 and 50 KDa were present (figure 5.10b). Proteins of 78 and 69 KDa were crosslinked in complex II, while identical protein smears of 30 to 46 KDa were identified in both SVWI-38 and CT-1 complex III.

Since it was not possible to accurately determine the apparent molecular weights of labelled complex III proteins using the above method, Sephacryl S-300 fractions, containing only complex III DNA-binding activity, were UV-crosslinked to Bromodeoxyuridine and ^{32}P-deoxycytidine labelled oligonucleotide F and resolved on SDS-polyacrylamide gels.

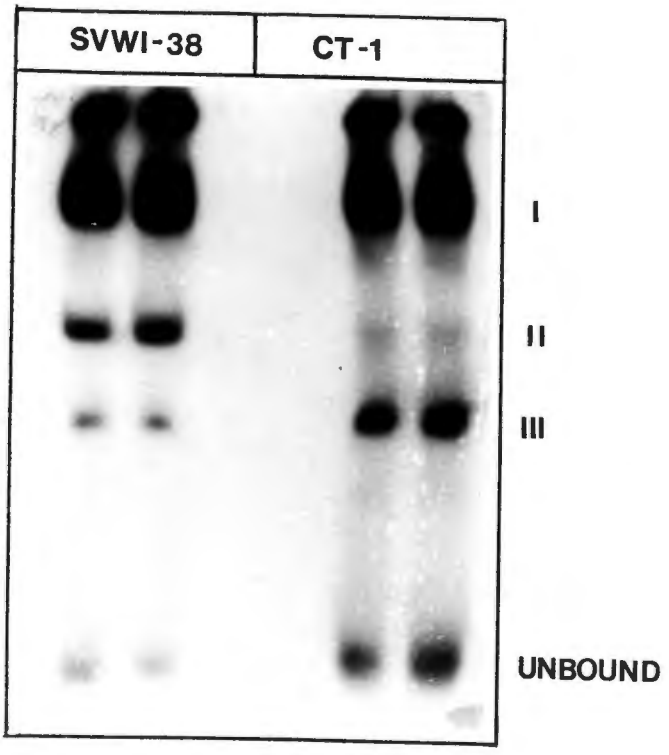


Fig 5.9 DNA-protein complex formation with bromodeoxyuridine substituted DNA. 40 ug of crude SVWI-38 and CT-1 nuclear extracts were incubated with 10 ug of poly dIdC, and 10 ng (10^5 cpm) of Br-dUTP and ^{32}P-dCTP labelled oligonucleotide A (nucleotides -107 to -71) as described in section 7.6.3. Duplicate samples were electrophoresed on 5% non-denaturing polyacrylamide gels at 4^{O} C in 0.5X TBE. The wet gels were exposed to X-ray film at 4^{O} C between 2 to 16 hours. The position of complexes I, II and III are indicated.

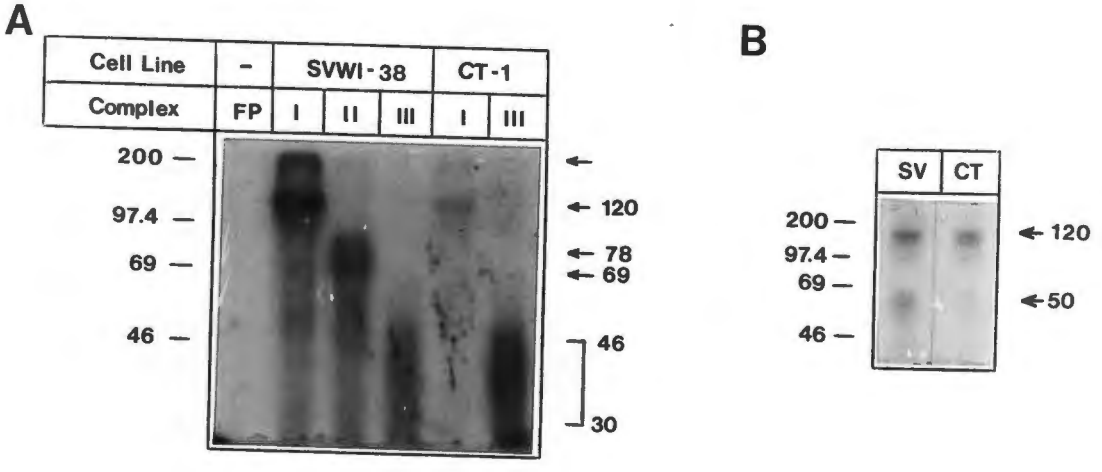


Fig 5.10 Polypeptide composition of DNA-protein complexes. **(A)** SVWI-38 and CT-1 nuclear extracts were incubated with Br-dUTP and ^{32}P-dCTP labelled probe and resolved on 5% non-denaturing polyacrylamide gels as described in the legend of figure 5.9. The DNA-protein complexes were UV-crosslinked *in situ*, visualised by autoradiography, the bands excised, the DNA-protein complexes eluted and resolved on 10% SDS-polyacrylamide gels as described in section 7.6.3.2. The molecular weight markers and crosslinked products are indicated in KDa. The lane labelled FP is the free probe. **(B)** UV-crosslinked SVWI-38 (SV) and CT-1 (CT) complex I proteins resolved on a 10% SDS-polyacrylamide gel.

The 47-mer oligonucleotide F, extending from nucleotides -107 to -61, was chosen for these experiments because it contained all the recognition sequences, the complete CME, 5'- and 3'-flanking sequences. Sephacryl S-300 fractions containing only complex I or II DNA-binding activities were also assayed with this oligonucleotide to confirm the previous results.

In order to determine the optimal time required for UV-irradiation, binding reactions were irradiated for increasing lengths of time (0 to 60 min) as described in section 7.6.3.3. The amounts of incorporated label increased when SVWI-38 complex I was irradiated for longer time intervals (figure 5.11a). As expected, no crosslinked polypeptides were generated when UV-irradiation was omitted. The number and relative intensities of the four labelled species of 234, 145, 89 and 79 KDa did not change with the different irradiation times tested. It was decided to irradiate the binding reactions for 15 min prior to SDS-polyacrylamide gel electrophoresis. The 145 KDa band was always the strongest, while the 234 KDa band was very faint or sometimes not even present. Occasionally an extra band which migrated faster than the 145 KDa complex was detected in both SVWI-38 and CT-1 extracts (figure 5.12a).

On average the bands generated by crosslinking oligonucleotide F to G/CBF were 25 KDa larger than those crosslinked to the shorter oligonucleotide A probe (145 versus 120 KDa and 79 versus 50 KDa). The decrease in the mobility of crosslinked G/CBF-oligo F complex bands was therefore probably due to the additional DNA. In an attempt to minimise this effect the samples were digested with nucleases prior to electrophoresis as described in section 7.6.3.4. Unfortunately, as shown in figure 5.11b, all the bands were abolished when the samples were treated with nucleases.

Since the native molecular weight of G/CBF was calculated to be 105 KDa (section 5.2.1.3), the predominant 120 and 145 KDa bands, produced when partially purified complex I proteins were crosslinked to oligonucleotides A and F respectively, probably contained a single G/CBF. As explained previously, the 50 KDa (oligonucleotide A) and 79 KDa (oligonucleotide F)

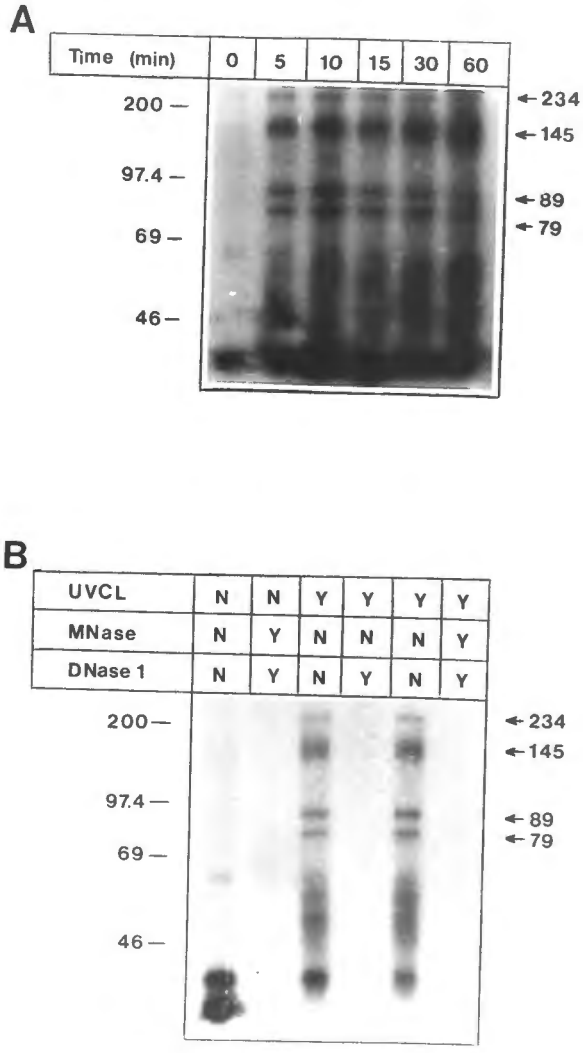


Fig 5.11 Optimisation for UV-crosslinking. **(A)** Partially purified SVWI-38 complex I proteins (Sephacryl S-300 fractions) were crosslinked (UVCL) to oligonucleotide F (nucleotides -107 to -61) as described in section 7.6.3.3. Briefly the binding reactions were irradiated for increasing lengths of time (0 to 60 min), resolved on 8% SDS-polyacrylamide gels, dried and exposed to X-ray film. The molecular weight markers (left) and crosslinked products (right) are indicated in KDa. **(B)** Selected UV-crosslinked (UVCL) samples (30 min) were treated with nucleases (DNase and/or MNase) prior to SDS-PAGE (7.6.3.3). Y and N represent treated and untreated samples respectively. The molecular weight markers and crosslinked products are indicated in KDa.

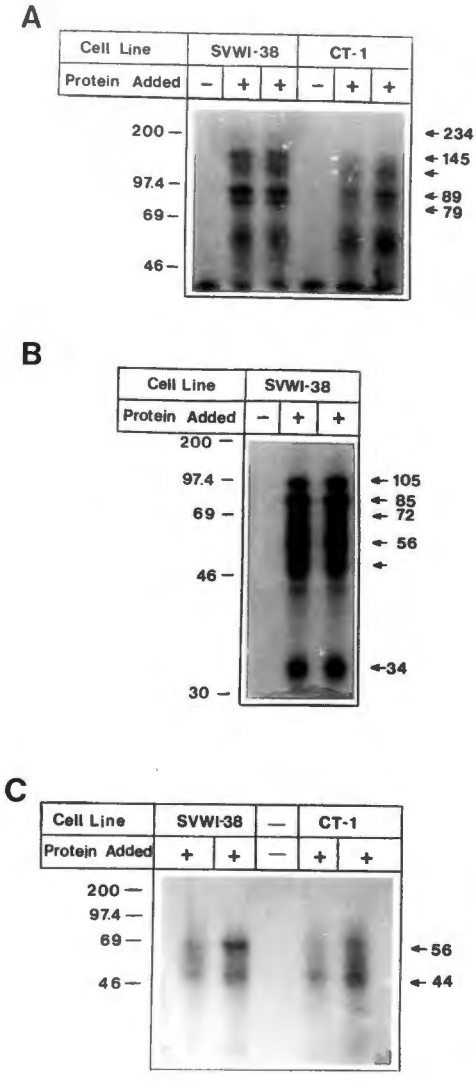


Fig 5.12 UV-crosslinking analysis of complex I, II and III proteins. Partially purified SVWI-38 and CT-1 complex I, II and III proteins (Sephacryl S-300 fractions) were UV-crosslinked to oligonucleotide F (nucleotides -107 to -61) as described in section 7.6.3.3, resolved on SDS-polyacrylamide gels, dried and exposed to X-ray film. The molecular weight markers (left) and crosslinked products (right) are indicated in KDa. **(A)** UV-crosslinked complex I proteins resolved on an 8% gel, **(B)** UV-crosslinked complex II proteins resolved on a 10% gel and **(C)** complex III proteins resolved on a 12% gel.

bands were most likely degradation products of G/CBF. The extra 89 KDa species identified when this factor was crosslinked to oligonucleotide F could either be a degradation product or a species where only some of the components of G/CBF were crosslinked to the probe. The presence of the weak 200 and 234 bands coupled to oligonucleotides A and F respectively, implied that the stoichiometry of the various protein components in complex I could be complex. It is tempting to speculate that there are at least two forms of complex I where the 109 KDa G/CBF (section 5.2.1.3) can bind to its DNA-element either as a monomer or homodimer to produce protein complexes of about 120 and 230 KDa respectively.

Occasionally, an extra complex which migrated slightly faster than complex I, was present when nuclear extracts were assayed by electrophoretic mobility shift assays using oligonucleotide F (figures 4.19, 4.24 and 4.25). Figure 4.19 also shows that this extra complex was always present and much stronger when an oligonucleotide containing a mutated CCAAT box, MUT-CCAAT, was used in the binding studies. When this extra complex was crosslinked to oligonucleotide A and resolved on 10% SDS-polyacrylamide gels, two bands with molecular weights of 120 and 50 KDa were identified (figure 5.13). The 200 KDa band was not detected, suggesting that this complex was structurally a less complex form of complex I. This result supports the hypothesis that the stoichiometry of G/CBF varies to produce at least two forms of the complex I.

Three predominant bands of 105, 85 and 34 KDa were identified when partially purified Sephacryl S-300 fractions containing only complex II DNA-binding activity were crosslinked to oligonucleotide F and resolved on 10% SDS-polyacrylamide gels (figure 5.12b). Minor bands ranging between 46 and 72 KDa where also identified. The 72 and a 56 KDa species were always the sharpest in this region of the gel. Occasionally, an extra band which migrated slightly faster than the 56 KDa species was also present. Assuming the extra 10 bp of oligonucleotide F contributed about 25 KDa to the apparent molecular weights of the bands (in the form of conformation changes), the 105 and 85 KDa bands were probably equivalent

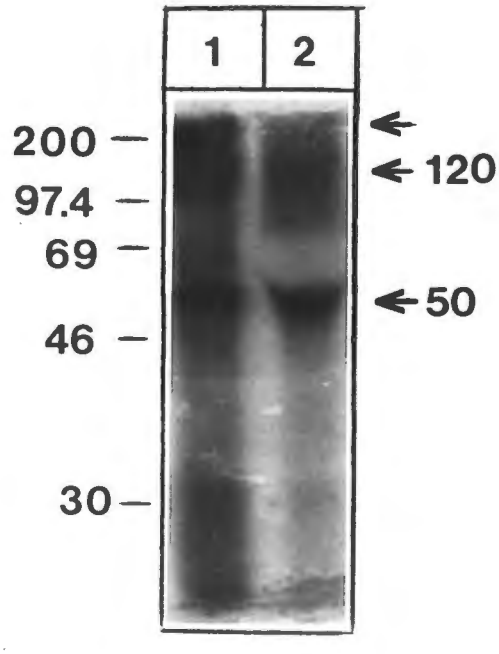


Fig 5.13 Polypeptide composition of complex I and "minor" complex I proteins. Crude SVWI-38 nuclear extracts were UV-crosslinked to oligonucleotide A as described in the legend to figures 5.9 and 5.10. Complex I (lane 1) and a slightly faster migrating complex ("minor" complex I) (lane 2), were excised, the DNA-protein complexes eluted and resolved on 10% SDS-polyacrylamide gels as described in section 7.6.3.2. The molecular weight markers (left) and crosslinked products (right) are indicated in KDa.

to the 78 and 69 KDa bands generated when the shorter oligonucleotide A was used in the assay. Since only two bands were identified when crude SVWI-38 extracts were assayed with oligonucleotide A, the minor bands within the 46 to 72 KDa range were probably due to degradation of complex II proteins during gel filtration chromatography. Since both the 78 and 69 KDa bands (figure 5.10a) are slightly larger than the native molecular weight of 66 KDa for complex II protein(s) as determined from the Stokes radius and sedimentation coefficients (section 5.2.1), this protein could be a monomer, where the smaller band is a degradation product. Alternatively, the presence of smaller bands could be an indication that the factor is a hetero-multimer where the protein components of the 34 and 85 KDa bands could dimerise to form the 105 KDa crosslinked product.

When oligonucleotide F was crosslinked to partially purified complex III proteins, two products with apparent molecular weights of 56 and 44 KDa where identified in both cell lines (figure 5.12c). Both these bands are slightly larger than the native molecular mass of 40 KDa for the protein component of complex III (section 5.2.1). This factor could also be a monomer or a multimer where all the polypeptide components are in direct contact with and probably crosslinked to the DNA. Alternate splicing of a single gene product or degradation during purification could explain the presence of the two bands.

5.2.2.2 DENATURING GEL FILTRATION

Since it wasn't clear from the UV-crosslinking experiments whether complex I, II or III proteins were multimers or monomers, analytical gel filtration experiments were repeated under denaturing conditions to obtain more insight into the structural complexity of these factors. Crude SVWI-38 extracts were fractionated in the presence of 0.8 M KCl in an attempt to disrupt all protein-protein interactions so that each protein component would elute independently (Kim *et. al.*, 1988).

One to 2 mg of crude SVWI-38 extract containing 0.8 M KCl and 0.25% NP-40 in a final volume of 400 ul was loaded onto a

Sephadex G-200 column and fractionated with buffer CB containing 0.8 M KCl and 0.02% sodium azide as described in section 7.6.1. The bulk of the loaded protein eluted in the void volume (figure 5.14a). Five ug BSA was added to every 200 ul fraction collected to minimise protein loss during dialysis. Every 5th fraction was dialysed overnight against buffer CB (section 7.7) containing 0.1 M KCl to remove the salt and restore DNA-binding activity. The inset in figure 5.14a shows that complex II DNA-binding activity eluted first, followed by complex I and finally complex III activities.

The Stokes radius of each complex was calculated as described previously (section 5.2.1.1) and shown in table 5.3 and figure 5.14b. Under denaturing conditions, SVWI-38 complex I, II and III proteins had Stokes radii of about 3.68, 4.24 and 3.17 nm respectively. The Stokes radii of complex II and III proteins did not change significantly after fractionation in the higher salt concentration, suggesting that these factors are monomers. An alternative explanation is that their non-covalent protein-protein interactions were not disrupted under these conditions as have been shown for the protein-protein interaction between the α and ß subunits of α-CP1 (Kim and Sheffery, 1990). The decrease in the Stokes radius of G/CBF (complex I) from 5.22 nm to 3.67 nm confirmed that the factor is multimeric. Under these conditions the various components of the factor sedimented independently and G/CBF activity is detected in those fractions where the components co-elute.

5.2.2.3 SOUTHWESTERN BLOTTING

Both the UV-crosslinking and denaturing gel filtration data suggested that the protein components of complexes II and III could be monomeric. In order to verify this, Southwestern blotting experiments were done as described in Materials and Methods (section 7.6.4).

Crude SVWI-38 and CT-1 nuclear extracts were resolved on 12% SDS-polyacrylamide gels, renatured and transferred to

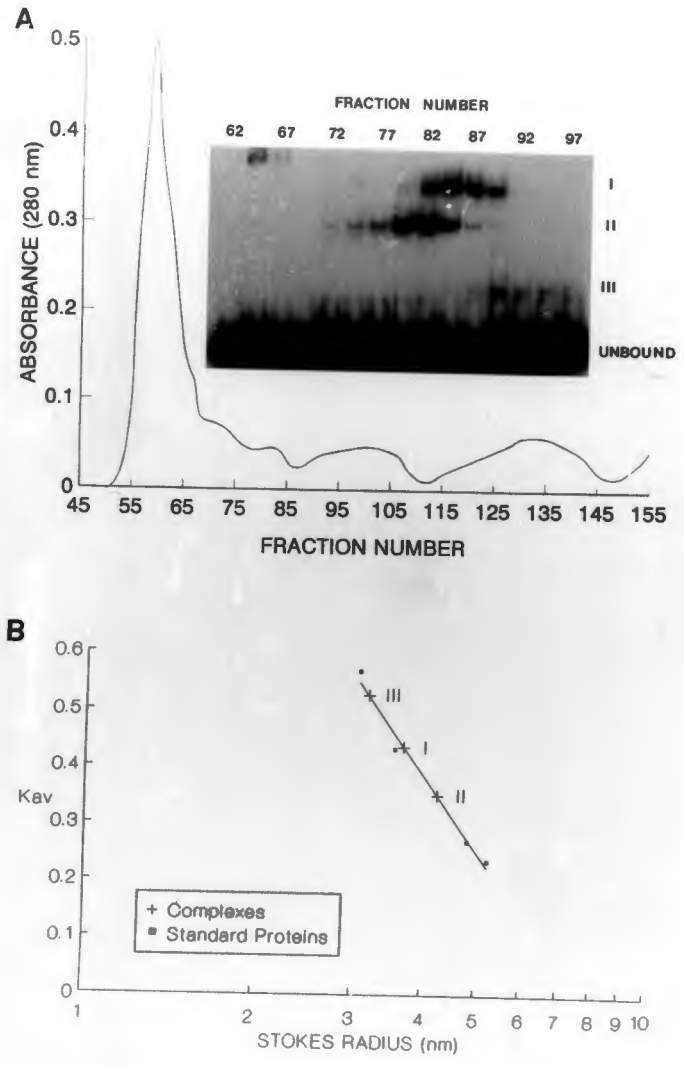


Fig 5.14 Gel filtration of SVWI-38 nuclear extracts on Sephadex G-200 in the presence of 0.8 M KCl. **(A)** One to two mg crude nuclear extracts in a final volume of 400 ul buffer CB containing 0.8 M KCl and 0.25% NP-40 were applied to a 1 cm X 60 cm column and fractionated with CB containing 0.8 M KCl and 0.02% sodium azide at a flow rate of 3 ml/cm^2/h. Two hundred ul fractions were collected and every 5th fraction dialysed against CB containing 0.1 M KCl. The DNA-binding activity in each dialysed fraction was determined by EMSA using fragment E (nucleotides -110 to -46) as a probe (inset). The positions of complexes I, II and III are indicated. **(B)** The gel filtration column was calibrated using protein standards and the log of the Stokes radius (table 5.3) versus k_{av} of each standard protein was plotted. The Stokes radii of complexes I, II and III under denaturing conditions were determined from the curve.

TABLE 5.3 Denaturing analytical gel filtration of SVWI-38 nuclear extracts. Proteins were fractionated on Sephadex G-200 columns as described in section 7.6.1. The Stokes radii (R_S), molecular weights (MW), elution volumes (V_e), calculated K_{av} and the standard deviations of n determinations are tabulated. The void (V_o) and total (V_t) volumes were determined using blue dextran 2000 and thymidine respectively. Crude nuclear extracts were fractionated in the presence of 0.8 M KCl as described in section 7.6.1. Every 5th fraction was dialysed against buffer CB containing 0.1 M KCl to restore DNA-binding activity. The DNA-binding activity in these fractions was quantitated by EMSA using fragment E as a probe. The K_{av} of each complex was calculated from their V_e, the V_o and V_t and the Stokes radii determined from the semi-logarithmic plot of Stokes radius versus K_{av} of the standard proteins.

STANDARDS/SAMPLES	Ve (ml)	K_{av}	R_s (nm)	MW (KDa)	n
STANDARDS					
Catalase	18.76 ± 0.09	0.237 ± 0.003	5.22	232	3
Aldolase	19.60 ± 0.42	0.271 ± 0.018	4.81	158	2
BSA	23.50 ± 0.77	0.432 ± 0.032	3.55	68	4
Ovalbumin	26.80 ± 0.70	0.568 ± 0.029	3.05	43	3
Blue Dextran (Vo)	13.02 ± 0.16	-	-	-	7
Thymidine (Vt)	37.27 ± 0.48	-	-	-	6
SAMPLES					
Complex I	23.60	0.436	3.68	76	2
Complex II	21.53	0.351	4.24	117	1
Complex III	25.77	0.527	3.17	49	2

nitrocellulose membranes. The membranes were blocked in buffer containing either BSA (1% or 5%) or Carnation nonfat milk powder (0.5% or 1%) to determine the optimal blocking conditions. The blocked membranes were probed with ^{32}P-labelled fragment E (figure 2.8) in buffer containing 0.25% of either BSA or Carnation nonfat milk powder. Three bands of molecular weights 120, 46 and 42 KDa were identified when the membranes were blocked with 1 or 5% BSA (figure 5.15), whereas only the 120 KDa polypeptide was obtained when blocking with 0.5% Carnation nonfat milk powder. Filters blocked with buffer containing 1% Carnation nonfat milk powder failed to bind any probe due to the coating of all the sites and possibly introducing steric hindrance. If membranes blocked with 0.5% Carnation nonfat milk powder were washed less stringently and exposed to X-ray film for longer periods of time four polypeptides with molecular weights of 120, 46, 42 and 34 KDa were observed (figure 5.16b).

Membranes were stained with amido black as described in section 7.6.4 to check for the efficiency of protein transfer. As shown in figure 5.16c, transfer occurred evenly across the membranes. Only trace amounts of the large molecular weight polypeptides remained in the gels after electroblotting (data not shown).

Since the 46, 42 and 34 KDa bands were present in both SVWI-38 and CT-1 nuclear extracts and since they are in the size range for complex III proteins, they could be the protein component of this complex. The difference in the molecular weights of the 46 and 34 KDa bands (12 KDa) and the two bands produced in the UV-crosslinking experiments (figure 5.12c) were identical (56 KDa to 44 KDa). These data therefore suggests that complex III proteins are monomers and that the smaller species are probably degradation products which still retain DNA-binding activity or modified forms. The data also suggested that the SVWI-38 specific complex II proteins are hetero-multimers because no SVWI-38 specific bands were identified in these experiments. Alternatively, if complex II binds as a monomer, then the proteins did not renature after electrophoresis. The presence of the very strong 120 KDa band in both CT-1 and SVWI-38 extracts could

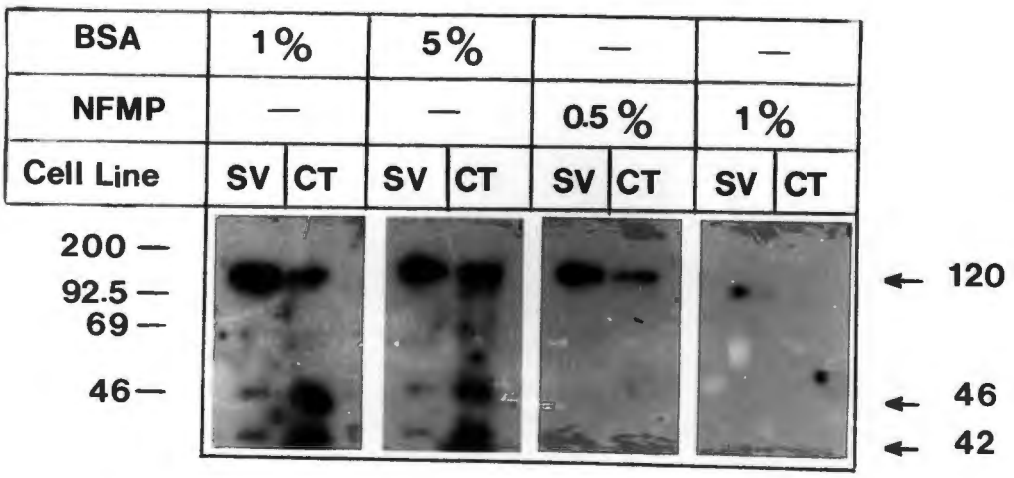


Fig 5.15 Optimising conditions for the Southwestern blotting experiments. Crude SVWI-38 and CT-1 nuclear extracts were resolved on 12% SDS-polyacrylamide gels, the gels renatured and the proteins transferred to nitrocellulose membranes as described in section 7.6.4. The filters were incubated in 50 ml of blocking buffer (section 7.7) containing either 1% or 5% BSA or 0.5% or 1% Carnation nonfat milk powder (NFMP) for at least 2 hours at room temperature with gentle shaking. The blocked membranes were probed with ^{32}P-labelled fragment E (nucleotides -110 to -46) in buffer containing either 0.25% of either BSA or NFMP, washed (section 7.6.4) and exposed to X-ray film for between 16 hrs and 3 days. The molecular weight markers (left) and bands (right) are indicated in KDa.

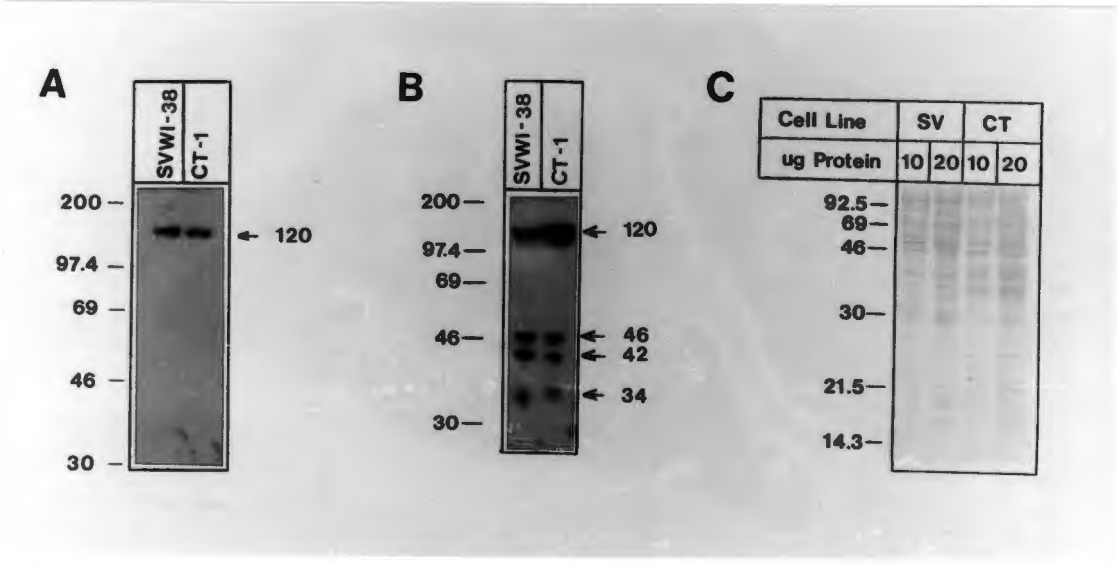


Fig 5.16 Southwestern blotting analysis of SVWI-38 and CT-1 nuclear extracts. **(A)** Crude nuclear extracts were resolved on 12% SDS-polyacrylamide gels, the gels renatured, the proteins transferred to nitrocellulose membranes and used in binding assays using fragment E as a probe as described in Materials and Methods (section 7.6). **(B)** Filters processed as above but were washed less stringently (shorter period of time) and exposed for longer periods of time. **(C)** After the crude nuclear extracts were resolved on 12% SDS-polyacrylamide gels and transferred to nitrocellulose membranes, the membranes were stained with 0.3% amido black in 20% ethanol and 7% acetic acid for 15 min with gentle shaking and destained in warm (50^{O} C) 10% acetic acid.

not be explained and these results should therefore be interpreted with caution (see section 5.3).

5.2.3 Protein-protein interactions are not DNA-dependent.

The denaturing gel filtration data showed that the 109 KDa G/CBF was a multimer. A 120 KDa and a 145 KDa band was observed when this factor was UV-crosslinked to oligonucleotides A and F respectively. These bands were probably generated when a single 109 KDa G/CBF protein-complex was UV-crosslinked to the respective oligonucleotides. Occasionally extra weak bands of 200 and 234 KDa were observed when oligonucleotides A and F respectively were used in the assays. These results suggested that they could be multiple combinations of factors in the G/CBF-DNA complex (different forms of the complex). To test whether DNA could mediate the multimerisation of complex I, II or III proteins, DNA-protein complexes were assayed by gel filtration on Sephacryl S-300 column.

About 40 ug of SVWI-38 or CT-1 nuclear extracts were bound with 20 ng (2 X 10^5 cpm) ^{32}P-labelled oligonucleotide F in a final volume of 80 ul containing 20 ug poly dIdC.poly dIdC as described in section 7.4.2. The amount of probe chosen for these experiments was such that all the probe would be bound to proteins. Half a ml of buffer CB containing 30% glycerol and 0.1% NP-40 was added to the binding reactions, layered onto a calibrated Sephacryl S-300 column and fractionated as described in Materials and Methods (section 7.6.1). Figure 5.17, shows that two overlapping DNA-protein peaks (a and b) eluted from the column with identical elution volumes when comparing SVWI-38- and CT-1-Oligo F complexes. The Kav of each peak (table 5.4) was calculated as described previously (section 5.2.1.1) and the Stokes radii determined from the semi-logarithmic plot of Stokes radius versus Kav (figure 5.17c). Peaks a and b have Stokes radii of 5.89 ± 0.09 and 4.09 ± 0.08 nm which are comparable to spherical proteins of 295 and 107 KDa respectively. The Stokes radius of oligonucleotide F was determined by processing 2 X 10^5 cpm free oligonucleotide over the column (peak c in figure 5.17). The Stokes radius of peak c was 3.69 ± 0.01 nm which is corresponds to a size of a spherical

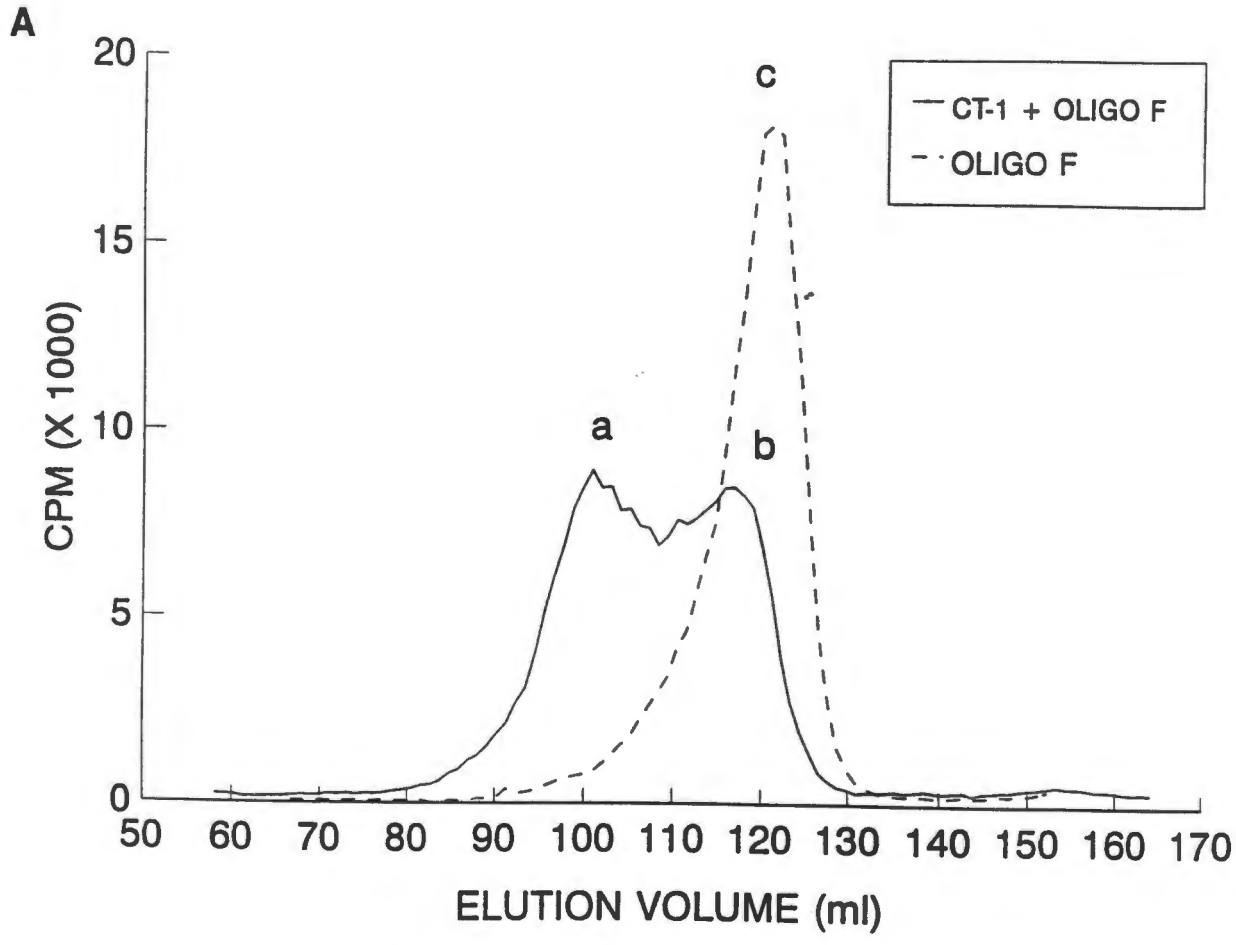
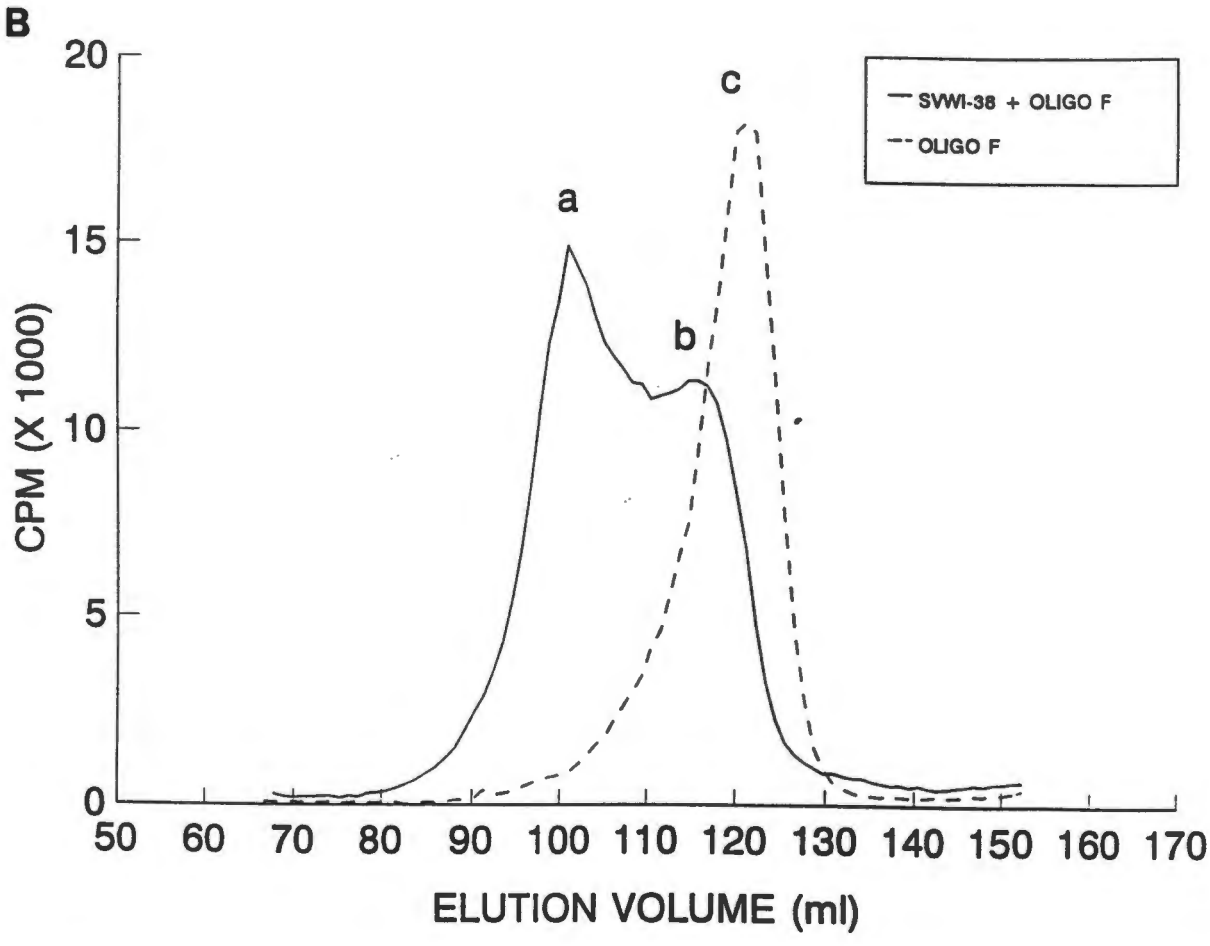


Fig 5.17 Analytical gel filtration of DNA-protein complexes on Sephacryl S-300. **(A)** Crude CT-1 or **(B)** SVWI-38 nuclear extracts were prebound with limiting amounts of oligonucleotide F (oligo F), layered onto a 100 cm X 1.6 cm column of Sephacryl S-300 and fractionated as described in section 7.6.1. The amount of radioactivity in each fraction was determined by scintillation counting. The elution profiles of the oligo F-protein complexes (peaks a + b) and free oligo F (peak c) is shown. **(C)** The Sephacryl S-300 gel filtration columns were calibrated as described in Materials and Methods (7.6.1). A calibration curve was prepared by plotting the log of the Stokes radius (thyroglobin, 8.50 nm; ferritin, 6.10 nm; catalase, 5.22 nm; aldolase, 4.81 nm and BSA, 3.55 nm) versus k_{av}, where $k_{av}=(Ve-Vo)/(Vt-Vo)$. The Stokes radii of the three peaks in (A) and (B) were determined from the curve.


B
SVWI-38 + OLIGO F
OLIGO F
a
b
c
CPM (X 1000)
20
15
10
5
0
50 60 70 80 90 100 110 120 130 140 150 160 170
ELUTION VOLUME (ml)

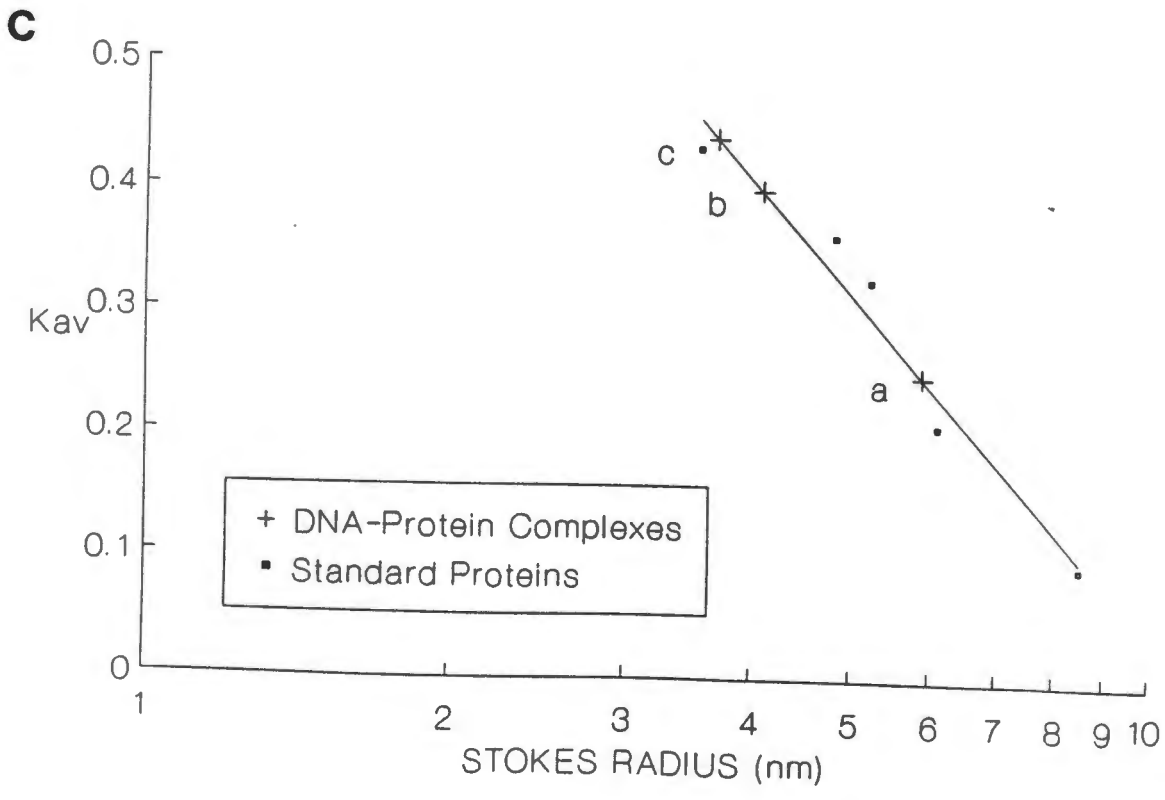

C
0.5
0.4
0.3
0.2
0.1
0
Kav
c
b
a
+ DNA-Protein Complexes
▪ Standard Proteins
1 2 3 4 5 6 7 8 9 10
STOKES RADIUS (nm)

TABLE 5.4 Analytical gel filtration of SVWI-38 and CT-1 DNA-protein complexes. A Sephacryl S-300 column (100 cm X 1.6 cm) was packed and calibrated as described in section 7.6.1. The Stokes radii (Rs), molecular weights (MW), elution volumes (Ve), calculated Kav and standard deviations for n determinations are tabulated. The Vo and Vt were determined using blue dextran 2000 and thymidine respectively. The elution profiles of SVWI-38- and CT-1-protein-oligo F complexes and free oligonucleotide F were determined as described in the legend to figure 5.17. The Kav of peaks a, b and c was calculated from their Ve, the Vo and Vt and the Stokes radii determined from the semi-logarithmic plot of Stokes radius versus Kav.

STANDARDS/SAMPLES	Ve (ml)	K_{av}	R_s (nm)	MW (KDa)	n
STANDARDS					
Thyroglobin	85.1 ± 0.2	0.095 ± 0.002	8.5	669	2
Ferritin	96.9 ± 1.0	0.209 ± 0.010	6.1	440	2
Catalase	109.0	0.326	5.22	232	1
Aldolase	112.6	0.361	4.81	158	1
BSA	119.9 ± 1.3	0.431 ± 0.013	3.55	68	2
Blue Dextran (Vo)	75.2 ± 0.6	-	-	-	2
Thymidine (Vt)	178.9 ± 3.5	-	-	-	2
SAMPLES					
Peak A (Complex I)	101.1 ± 0.6	0.250 ± 0.006	5.89 ± 0.09	295	7
Peak B (Complex III)	116.5 ± 0.6	0.398 ± 0.006	4.09 ± 0.08	107	5
Peak C (Oligo F)	3.69 ± 0.01	75	120.7 ± 0.1	0.439 ± 0.001	2

protein of 75 KDa.

G/CBF eluted from the gel filtration column as a 232 KDa protein in the absence of DNA (section 5.2.1.1), while the protein component of peak a contributed about 220 KDa (295 KDa minus 75 KDa) to the DNA-protein complex. Complex III proteins, on the other hand, eluted from the gel filtration column as a 50 KDa protein in the absence of DNA (section 5.2.1.1), while the protein component of peak a contributed about 32 KDa (107 KDa minus 75 KDa) to the DNA-protein complex. Peaks a and b are therefore probably complexes I and III respectively. These results therefore suggested that the DNA does not mediate multimerisation of complex I or III proteins. The complex II peak was masked by peaks a and b in the SVWI-38 experiment.

5.3 DISCUSSION

The biochemical and binding properties of complex I, II and III proteins in crude and partially purified SVWI-38 and CT-1 nuclear extracts have been investigated in this and a previous study (Smith, 1989). These properties were used in comparative studies to determine whether these factors were related to previously characterised DNA-binding proteins. In this study, analytical gel filtration, glycerol gradient sedimentation, UV-crosslinking and Southwestern blotting studies were used to determine the native molecular weights, Stokes radii, sedimentation coefficients and structural complexity of the these transcription factors.

Under non-denaturing conditions, G/CBF had a Stokes radius and sedimentation coefficient of 5.22 ± 0.04 nm (comparable to a spherical protein of 232 KDa) and 4.9 ± 0.4 S respectively, a calculated native molecular weight of 105 KDa and a frictional ratio of 1.66. The Stokes radius of 5.89 ± 0.09 nm for the G/CBF-oligo F complex on the other hand was comparable to a spherical protein of 299 KDa. The CCAAT box factor, α-CP1, purified from murine erythroleukemia (MEL) cells, has a native molecular weight of 101 KDa, sedimentation coefficient of 4.3 S, Stokes radius of 5.7 nm and frictional ratio of 1.78 (Kim and Sheffery, 1990). The ubiquitous CCAAT box binding factor, NF-Y, also has a similar sedimentation coefficient on glycerol gradients and the estimated molecular weight of the NF-Y/E_{α} oligo complex is 250-300 KDa (Hooft van Huijsduijnen *et. al.*, 1987).

The gel filtration data also showed that G/CBF is a multicomponent CCAAT box binding factor because the various protein components of the factor eluted independently on Sephadex G-200 in the present of 0.8 M NaCl and as an intact complex in the absence of salt on Sephacryl S-300. The gel filtration data also suggested that the various components of the factor interacted with each another in the absence of DNA. The proposed heteromeric structure of G/CBF is consistent with a number of studies which suggest that the related CP-1 (Chodosh *et. al.*, 1988a), α-CP1 (Kim and Sheffery, 1990) and CBF (Hatamochi *et. al.*, 1988) CCAAT box binding factors are heteromers in solution.

Three bands of approximately 200, 120 and 50 KDa were identified in complex I by UV-crosslinking to a 37-mer oligonucleotide A, while four bands of 234, 145, 89 and 79 KDa by UV-crosslinking to the 47-mer oligonucleotide F. Since G/CBF has a calculated molecular mass of 109 KDa, the predominant 120 and 145 KDa bands were probably produced when a single heterotypic complex was UV-crosslinked to oligonucleotides A and F respectively. The 50, 89 and 79 KDa bands were probably degradation products, a finding supported by the data of Hooft van Huijsduijnen *et. al.* (1987). The weak 200 and 234 KDa bands suggested that the stoichiometry of G/CBF in the DNA-protein complex maybe more complex. G/CBF might bind with the oligonucleotides as a dimer to produce the larger complexes. This result is consistent with that of the heterotrimeric factor, CBF, where three polypeptides associate with one another in the absence of DNA to form a functional factor with an apparent molecular mass of about 113 KDa (Maity *et. al.*, 1992). Glutaraldehyde crosslinking with purified CBF, on the other hand, produced three bands with apparent molecular weights of 170, 100 and 80 KDa, suggesting that the structure of CBF may also be more complex (Maity and de Crombrugghe, 1992).

These similarities strongly suggested that G/CBF belongs to the family of the previously identified heteromeric CCAAT box binding factors, NF-Y, CP-1, α-CP1 and CBF. Other studies on the effects of heat, ionic strength and various metal ion chelators on NF-Y (Hoof van Huijsduijnen, 1987) and G/CBF (Smith, 1989) DNA-binding activities supports this hypothesis.

The CME binding proteins (complexes II and III) had Stokes radii of 4.12 and 3.15 nm, sedimentation coefficients of 3.9 and 3.2, native molecular weights of 66 and 41 KDa and frictional ratios of 1.54 and 1.38 respectively, which did not differ when fractionation was performed in the presence or absence of KCl.

Since several bands were identified when complex II proteins were UV-crosslinked to oligonucleotides, it was not clear if this factor was homo- or hetero-multimeric. The Southwestern

blotting data strongly suggested that this factor is hetero-multimeric since no SVWI-38 specific bands were identified. However this observation should be interpreted with caution since it is possible that the protein did not renature correctly after electrophoresis. The denaturing gel filtration results, on the other hand showed that the factor is either monomeric or that the protein-protein interactions between the various components are stable in 0.8 M KCl.

The identification of two crosslinked bands with molecular weights of 56 and 44 KDa and three bands of 46, 42 and 34 KDa when both CT-1 and SVWI-38 nuclear extracts were assayed by the UV-crosslinking and Southwestern blotting experiments respectively, suggested that complex III proteins were monomers. The sizes of the larger bands were similar to the native molecular weight of the protein and the smaller bands could be degradation products or modified forms. Although these results and the denaturing gel filtration data strongly suggested that complex III proteins are monomers, the smaller bands could be a indication that the factor is hetro-multimeric, but it is highly unlikely. Additional experiments would need to do be done to confirm this hypothesis.

It was not clear from the data how complex II and III proteins were related. It is possible that complex III proteins are a specific degradation product of complex II proteins. Alternatively complex II and III formation could be the result of different post-translational modifications of the same factor. They could also contain a common DNA-binding subunit(s) or they maybe two distinct transcription factors which bind to the same *cis*-element on the collagen promoter. On example of the latter is the transcriptional activator, interferon regulatory factor-1 (IRF-1), and repressor (IRF-2) which bind to the same *cis*-elements in the type I interferon (IFN) and IFN-inducible genes (Harada *et. al.*, 1989).

Comparison of the recognition sequence (chapter 4) and biochemical properties (chapter 5) of G/CBF with the previously characterised transcription activators NF-Y, CBF, CP-1 and α-CP1 strongly suggested that G/CBF belongs to

a family of heteromeric CCAAT box binding factors. Competition experiments with oligonucleotides containing consensus sequences for these factors and assays using specific antibodies, such as anti-CBF-A (Maity and de Crombrugghe, 1992), are in progress and will elucidate this relationship. If G/CBF is identical to these previously characterised factors, then this factor is probably a transcriptional activator of the human α2(I) procollagen gene. It isn't clear from the data if the SVWI-38 specific complex II proteins is a homo-multimer, a hetero-multimer or a monomer. Previous studies have suggested that it is a hetero-multimeric factor (Smith, 1989). Although not conclusive, the data strongly implied that complex III proteins consists of a monomer. Since the CME binding proteins are not to our knowledge related to any previously characterised DNA-binding factor, they are probably novel factors involved in α2(I) procollagen gene regulation.

CHAPTER 6

CONCLUSION

Previous studies have suggested that a *trans*-acting factor(s) which binds to the proximal promoter of the human α2(I) procollagen gene maybe responsible for the inactivation of this gene in a simian virus 40 transformed embryonic lung fibroblast cell line (SVWI-38) (Parker *et. al.*, 1989, 1992). These fibroblasts secrete a homotrimer of type I collagen consisting of three overmodified α1(I) chains. No detectable amounts of α2(I) protein or mRNA are produced by these cells.

In this study, the SVWI-38 proximal promoter, which has previously been shown to have decreased activity in promoter-CAT constructs (Parker *et. al.*, 1992), was sequenced and shown to be completely normal, thereby ruling out any possibility that mutations within this region of the gene was responsible for its inactivation.

DNA-protein complexes between the proximal promoter and nuclear extracts prepared from SVWI-38 fibroblasts and a type I collagen producing human embryonic lung fibroblast cell line transformed by γ-radiation (CT-1) were investigated using the electrophoretic mobility shift assay. Similar DNA-protein complexes were observed in the -351 to -108 bp region in both cell lines. Nuclear extracts from CT-1 fibroblasts or the parental WI-38 cell line formed two complexes designated complexes I and III with the -110 to +58 promoter fragment. Both these complexes also formed with a smaller promoter fragment extending from nucleotides -110 to -46. An additional complex, complex II, was identified when SVWI-38 nuclear extracts were assayed, as well as much lower levels of complex III when compared with CT-1 cells. Only complexes I and II were identified in a non-collagen producing transformed macrophage cell line, $P388D_1$. Since complex II is present in non-collagen producing cells, it is possible that this factor is a transcriptional repressor. Parker

et. al. (1992) have shown that a small fragment of the proximal promoter (-107 to +58) was able to titrate out an inhibitory effect on α2(I) procollagen gene expression in SVWI-38 fibroblasts, suggesting that a transcriptional repressor binds to this region of the promoter. Also, cells that do not synthesise α2(I) procollagen chains appeared to have no detectable amounts (macrophages) or lower levels (SVWI-38) of complex III than collagen producing cells. These observations suggested that the relative ratios of complex II and III, and not merely the absence or presence of complex II, may be important in determining whether the α2(I) procollagen gene is expressed or not. Interestingly, trace amounts of complex II were detected on longer exposures of CT-1 and WI-38 autoradiograms, suggesting that low levels of this complex was present in type I collagen producing cell lines.

Competition, electrophoretic mobility shift assays and methylation interference assays identified two distinct but overlapping DNA-recognition elements within the -110 to -46 region of the α2(I) procollagen promoter. Complex I proteins, called the GGAGG/CCAAT binding factor (G/CBF), binds to a 21 bp element, $^{5'}$GCCCTCCCATTGGTGGAGGCC$^{3'}$, between nucleotides -92 and -72. This element has been called a GGAGG/CCAAT box element (G/CBE) because it contains an inverted CCAAT box surrounded by an upstream inverted GGAGG box and downstream GGAGG box. The second *cis*-element, $^{5'}$GGAGGCCCTTTT$^{3'}$, extends from nucleotides -78 to -67 and contains the G/CBE downstream GGAGG box and 3'-flanking sequences. CT-1 nuclear extracts formed one complex with this element (complex III), whereas SVWI-38 extracts formed two complexes (complexes II and III) involving the same element. Since complex II protein(s), a candidate transcriptional repressor, bound to this novel 12 bp element, the element has been named the collagen modulating element (CME).

Further characterisation of the G/CBE using DNA-binding assays, methylation interference assays and mutation analysis, suggested the CCAAT box and the inverted upstream GGAGG box were very important motifs for G/CBF binding. Deletion of a single cytosine, which separates these two

motifs, abolished or drastically reduced the formation of complex I. These assays also suggested that the downstream GGAGG box is involved in weak non-specific DNA-protein contacts and that it very loosely forms part of the G/CBE. In fact, replacement of this element with random sequences did not abolish complex I formation. The G/CBE is highly homologous with the CBF element in the mouse α2(I) procollagen promoter (Hatamochi *et. al.*, 1986; Maity *et. al.*, 1988) and the NF-Y site within the thrombospondin 1 promoter (Framson and Bornstein, 1993). The thrombospondin 1 NF-Y site contains the downstream GGAGG box. while the mouse α2(I) procollagen and human thrombospondin promoters contain an upstream inverted GGAGG and GGAGG-like boxes respectively. The crucial cytosine residue which separates the upstream GGAGG and CCAAT boxes is conserved in both these promoters. These observations suggested that G/CBF may be related or identical to NF-Y and CBF. Physico-chemical analysis, such as gel filtration, glycerol gradient sedimentation and UV-crosslinking experiments, of G/CBF also suggested that this factor belongs to the family of previously characterised hetero-multimeric CCAAT box binding factors, NF-Y (Hooft van Huijsduijnen *et. al.*, 1987), α-CP1 (Kim and Sheffery, 1990) and others. UV-crosslinking and denaturing gel filtration data suggested that this factor, like the other hetero-multimeric CCAAT box binding factors, is a multicomponent CCAAT box binding protein. In contrast to CBF (Karsenty *et. al.*, 1988) and NF-Y (Dorn *et. al.*, 1987), which bind to their respective *cis*-elements only if they contain an intact CCAAT motif, mutation of the CCAAT box in the G/CBE did not totally abolish G/CBF binding. The physico-chemical and binding data, however, strongly suggested that G/CBF is related to the transcriptional activators NF-Y, α-CP1, CBF and others. Other studies on the effects of heat, ionic strength and various metal ion chelators on the ubiquitous NF-Y (Hooft van Huijsduijnen, 1987) and G/CBF (Smith, 1989) DNA-binding activities supports this hypothesis.

Mutations of the downstream GGAGG box and the cytosine dinucleotides at positions -72 and -73 (motifs within the CME) resulted in a drastic reduction or abolition of complex II and III formation. Methylation interference and mutation analysis data suggested that the upstream GGAGG and CCAAT

boxes were also involved in weak DNA-protein contacts with complex II and III proteins. Since the CME bears some resemblance to the NF-kB consensus sequence (Lenardo and Baltimore, 1989), the CME could be a NF-kB-like recognition element. The CME also overlaps a large T antigen-like recognition sequence present in the human α2(I) procollagen promoter (Ghosh *et. al.*, 1981). Smith (1989) has shown that large T antigen is not responsible for complex formation with the CME. Although not conclusive, the UV-crosslinking and Southwestern blotting data suggested that complex II protein(s) are multimeric. Since the Stokes radius of this factor did not change during fractionation on a denaturing gel filtration column, the protein-protein interaction between the various subunits of this factor are very stable assuming a multimeric factor. Smith (1989) has postulated that complex II proteins consist of at least two subunits.

UV-crosslinking, Southwestern blotting and denaturing gel filtration data suggested that complex III is a monomer. Additional experiments would need to be done to confirm the polypeptide composition of complexes II and III and their relationship to previously identified proteins.

It was not clear from the data how complex II and III proteins are related to one another. It is possible that complex III proteins are a specific degradation product of complex II proteins. Although unlikely, different post-translational modification of the same factor could result in complex II and III formation. Alternatively, complex II and III proteins could contain a common DNA-binding subunit(s) or, more likely, they may be two distinct transcription factors which bind to the same *cis*-element within the collagen promoter.

Methylation interference and binding assays demonstrated that complex I, II and III proteins bound to their respective elements in a mutually exclusive manner. It can be postulated that the candidate transcription activators, G/CBF and complex III proteins, together with the candidate transcriptional repressor, complex II proteins, tightly regulate the expression of the human α2(I) procollagen gene in type I collagen producing (CT-1 and WI-38) and non-

producing (SVWI-38) cells. Numerous examples have been characterised where a transcriptional repressor and activator binds to distinct overlapping (Barberis *et. al.*, 1987) or the same (Harada *et. al.*, 1989) recognition element in a mutually exclusive manner to regulate gene expression. Kim *et. al.* (1990) and Lim *et. al.* (1992) have identified a strong monomeric activator, α-CP2, which is present in limiting amounts and competes with α-CP1, a weaker but more abundant transcriptional activator, for binding to distinct overlapping recognition sites in a mutually exclusive manner to regulate the α-globin gene. Similarly, in the type I collagen producing CT-1 cell line, complex III proteins may strongly activate transcription when it binds to the CME. The abundant G/CBF may compete with complex III proteins for binding to the G/CBE in a mutually exclusive manner to weakly activate transcription of the α2(I) collagen gene. Conversely, complex II proteins may compete with complex I and III proteins in similar manner to inhibit the expression of this gene in SVWI-38 fibroblasts. The results of this study will make it possible to further examine the functions of these transcription factors and their interaction with one another in the regulation of the human α2(I) procollagen gene.

Finally, since the CME is not present in the mouse α2(I) procollagen promoter, this study also suggested that different mechanisms may be operational in different species as far as regulation of the α2(I) procollagen gene is concerned.

CHAPTER 7

MATERIALS AND METHODS

7.1 SUBCLONING AND PLASMID DNA ISOLATION

7.1.1 PREPARATION OF DNA FRAGMENTS

Plasmids were digested with the required restriction enzymes, the fragments were resolved on either non-denaturing polyacrylamide or agarose gels and the band of interest eluted from the gels using one of the following methods.

7.1.1.1 CRUSH SOAK METHOD

The DNA fragments were resolved on polyacrylamide gels and stained with ethidium bromide. The band of interest was excised, the gel and sliced into small pieces and transferred to microfuge tubes, to which 1 volume of DNA elution buffer (section 7.7) was added. The tube was incubated overnight at 37^{O} C with gentle rotation, the buffer removed with a drawn-out pasteur pipette and transferred to a fresh microfuge tube. The gel slices were washed with 0.5 volumes of DNA elution buffer, added to the above and the DNA precipitated overnight at -20^{O} C by the addition of 2.5 volumes 96% ethanol. The DNA was pelleted, washed with 70% ethanol, dried under vacuum and resuspended in a small volume of 1X TE. The recovered DNA fragments were used in ligation reactions (section 7.1.2) or in DNA-protein binding assays (section 7.4).

7.1.1.2 ELECTRO-ELUTION

The DNA fragments were resolved on agarose gels containing ethidium bromide using 1X TBE as the tank buffer. The DNA band was excised from the gel and placed in a dialysis tube containing 0.5 ml 0.2X TBE. The sealed dialysis tube containing the gel slice was placed in 1X TBE and the DNA

fragment was electro-eluted from the gel slice at 25 mA for 2 hrs. The efficiency of elution was checked under UV light. At the end, the polarity was reversed at 100 mA for 1 min to remove the DNA fragments from the walls of the dialysis tube. The buffer in the dialysis tube was removed and the DNA was precipitated overnight at -20^{O} C by the addition of 2.5 volumes 96% ethanol. The DNA pellet was washed with 70% ethanol, dried under vacuum and resuspended in a small volume of 1X TE. The recovered DNA fragments were used in ligations as described in section 7.1.2.

7.1.1.3 ELECTROPHORESIS INTO WHATMAN 3MM PAPER

The DNA fragments were resolved on a 1% agarose gel containing ethidium bromide using 1X TAE as the tank buffer. A slit was made in front of the band to be eluted and a strip of sterile Whatman 3MM paper about the same width as the band was inserted into the slit so that it protruded a few millimeters from the gel (Errington, 1990). The DNA was then electrophoresed into the paper, the paper removed, trimmed and placed into a 500 ul microfuge tube with a small hole in the bottom. The microfuge tube was place inside a 1.5 ml Eppendorf tube, 10 ul 1X TAE was added to the Whatman 3MM paper and the tubes centrifuged for 20 sec to recover the DNA. The recovered DNA fragments were used in ligations as described in section 7.1.2.

7.1.1.4 QIAEX EXTRACTION PROTOCOL

The QIAEX gel extraction kit (Diagen GmbH, Dusseldorf, Germany) was used to extract DNA fragments from low-melting agarose gels. DNA fragments were extracted from the gels as described by the manufactures and used in ligation reactions (section 7.1.2).

7.1.2 BLUNT-END LIGATION

7.1.2.1 FILLING OF RECESSED 3'-ENDS

The Klenow fragment of *E. Coli* DNA polymerase was used to fill recessed 3'-ends of double stranded DNA (Sambrook *et. al.*, 1989). 1 unit of the Klenow fragment was added per

microgram of DNA in 1X Klenow buffer (section 7.7) in a final volume of 20 ul, containing 0.5 mM each of dATP, dTTP, dCTP and dGTP. The reaction mixture was incubated at room temperature for 30 min, phenol/chloroform extracted, and ethanol precipitated to recover the DNA.

7.1.2.2 BLUNT ENDING WITH T4 DNA POLYMERASE

In order to blunt end DNA molecules with 3'-overlaps, the 3'-5' exonuclease activity of T4 DNA polymerase was used (Sambrook *et. al.*, 1989). One unit of T4 DNA polymerase per microgram of DNA was added to 20 ul of 1X T4 DNA polymerase buffer containing 0.5 mM each of dATP, dTTP, dCTP and dGTP. The reaction mixture was incubated at 37^{O} C for 1 hour, phenol/chloroform extracted and ethanol precipitated to recover the DNA.

7.1.2.3 LIGATION

Ligations were performed in a final volume of 10 ul as recommended by the manufactures. A DNA insert to vector ratio of 5:1 was usually used in the ligation reactions.

7.1.3 PREPARATION AND TRANSFORMATION OF COMPETENT *E.Coli* CELLS

Competent DK-1 cells, an E. Coli K12 strain, were prepared as follows: 10 ml of L-Broth was inoculated with 50 ul of DK-1 from a glycerol stock and grown overnight at 37^{O} C with shaking. One-ml of the overnight culture was used to inoculate 300 ml of L-Broth in a 1000 ml flask. The bacteria were grown with vigorous shaking until the OD_{650} was between 0.2-0.4. The cells were pelleted by centrifugation at 5000 rpm (4^{O} C) for 10 min in a Beckman JA 10 rotor, gently resuspended in 40 ml of ice cold 60 mM $CaCl_2$, 10 mM Pipes, pH 7.2, left on ice for 20 min and re-centrifuged at 5000 rpm and 4^{O} C for 10 min. The cells were resuspended in 4 ml of 60 mM $CaCl_2$, 10 mM Pipes, pH 7.2 containing 15% glycerol, frozen in 200 ul aliquots in a dry ice-ethanol bath and stored at -70^{O} C.

The transformation efficiency was checked using supercoiled pBR322 and was usually between 10^7 and 10^8 colonies per microgram pBR322.

Aliquots of competent cells were thawed on ice. 100 ul of the cells were transferred to 10 ml plastic tubes, 5 ul of ligation mixture was added to the cells, incubated on ice for 30 min and heat shocked at 42^{O} C for 2 min. One-ml of prewarmed (37^{O} C) L-Broth was added to the cells and incubated at 37^{O} C for 1 hour without shaking. 20, 50, 100 or 200 ul aliquots were plated out onto ampicillin plates and incubated overnight at 37^{O} C.

7.1.4 SCREENING OF BACTERIAL COLONIES

The alkaline lysis procedure (Birnboim and Doly, 1979; Birnboim, 1983) was used to prepare plasmid DNA for rapid analysis. Seven-ml of L-Broth containing 0.1 mg/ml ampicillin in a 12 ml sterile tube was inoculated with a single bacterial colony and grown overnight at 37^{O} C with shaking. Cells were harvested at 2000 rpm for 10 min in a Beckman TJ-6 centrifuge at 4^{O} C. The pellet resuspended in 200 ul of "RAPS" solution 1, transferred to a 1.5 ml microfuge tube and incubated for 5 min at room temperature. The bacteria were lysed with 400 ul of "RAPS" solution 2 on ice for 5 min. Three-hundred ul of "RAPS" solution 3 was added to the lysed cells and incubated for 5 to 10 min on ice. The precipitated chromosomal DNA and bacterial proteins were removed by centrifugation in a microfuge for 5 min at 4^{O} C. The supernatants were transferred to a fresh tube, centrifuged again as above and the supernatants transferred to another fresh tube. Six hundred ul (0.8 volume) of isoproponal was added to the supernatants and the nucleic acids allowed to precipitate at -20^{O} C for 30 min. The plasmid DNA and RNA was recovered by centrifugation, washed with 70% ethanol, dried under vacuum and resuspended in 70 ul TE.

7.1.5 LARGE-SCALE PLASMID DNA ISOLATION

Large-scale plasmid preparation was either by CsCl-ethidium bromide centrifugation (Radloff *et. al.*, 1967) or QIAGEN columns.

7.1.5.1 CsCl-ETHIDIUM BROMIDE CENTRIFUGATION

Two hundred ml of overnight cultures were harvested at 5000 rpm for 5 min in a Beckman JA 10 rotor at 4^{O} C. The pelleted cells were resuspended in 2.5 ml of Sucrose/Tris solution (section 7.7) and incubated on ice for 5 min. One ml of Sucrose/Tris solution containing 5 mg/ml lysozyme was added to the resuspended cells and incubated for a further 5 min on ice before the addition of 2 ml of 250 mM EDTA (pH 8.0). After 5 min on ice the cells were lysed by the addition of 4.5 ml of TTE and incubated on ice for 20 min with gentle mixing at 5 min intervals. The lysate was centrifuged at 20 000 rpm for 30 min in a Beckman JA 20 rotor at 4^{O} C to pellet the chromosomal DNA and cellular debris. The supernatant was carefully decanted into a measuring cylinder, CsCl was added to a final concentration of 1 mg/ml and dissolved by gentle shaking before the addition of 80 ul of 10 mg/ml ethidium bromide to every ml of lysate. The lysates were transferred to a 5 ml Beckman Quick-Seal ultracentrifuge tubes, centrifuged at 55 000 rpm for 20 hrs in a Ti70.1 rotor at 20^{O} C.

The plasmid bands were removed from the ultracentrifuge tubes using 20-G needles, transferred to a Corex tubes and extracted with an equal volume of 1-butanol saturated with water to remove the ethidium bromide. The extractions were repeated until all the ethidium bromide was removed from the aqueous phase. The aqueous phase was dialysed against a vast excess 0.1X TE to remove the CsCl, ethanol precipitated, washed twice with 70% ethanol and dried under vacuum. The plasmid DNA was resuspended in about 0.5 ml TE and the purity and concentration of the DNA measured by absorbance at 260 and 280 nm.

Alternatively, 2 volumes of water and 3 volumes (original) of isopropanol were added after butanol extractions and the plasmid DNA was precipitated on ice for 10 min. Chilling to below -20^{O} C was avoided to prevent the CsCl from precipitating out. The plasmid pellet was washed, dried and resuspended in TE.

7.1.5.2 QIAGEN COLUMNS

QIAGEN columns (Diagen GmbH, Dusseldorf, Germany) were used to prepare plasmid DNA from 500 ml overnight cultures as described by the manufactures.

7.1.6 DNA SEQUENCING

DNA sequencing was performed using the Sanger dideoxy chain termination method (Sanger *et. al.*, 1977) as described by the manufactures Sequenace^{TM} (United States Biochemical Corporation, Cleveland, Ohio).

Briefly, the double-stranded plasmids were denatured in 0.2 M NaOH, 0.2 mM EDTA for 5 minutes at room temperature followed by the addition of 0.1 volume of 2 M NH_3Ac (pH 4.6) and 3.5 volumes of ethanol. After leaving at -70^{O} C for 10 minutes, DNA was pelleted, washed with 70% ethanol and dried under vacuum. The DNA was redissolved in 7 ul of distilled water to which 1 ul of primer and 2 ul of Sequenase reaction buffer was added. The reaction mixture was heated at 65^{O} C for 10 minutes, 37^{O} C for 30 minutes and allowed to cool to room temperature for at least 2 hours.

The DNA sequencing reaction was carried out essentially as described by the suppliers and the samples were analysed on 6% denaturing-polyacrylamide sequencing gels.

7.2 CELL CULTURE

Human fibroblasts were grown and maintained in Eagle's Basal Medium (BME), supplemented with 10% heat-inactivated foetal calf serum, 100 ug/ml penicillin and 100 ug/ml streptomycin at 37^{O} C. At confluence, the cells were washed with PBS and harvested with 0.05% trypsin in PBS containing 10 mM EDTA at

37^{O} C for 5 min. The cells were washed in BME, resuspended in BME and split at a ratio of either 1:2 (WI-38) or 1:3 (SVWI-38 and CT-1).

Mouse macrophage cells were grown in suspension in RPMI-1640 medium, supplemented with 10% heat-inactivated foetal calf serum, 100 IU/ml penicillin, 100 ug/ml streptomycin and 0.25 ug/ml fungizone (Flow Laboratories, Irvine, Ayrshire, Scotland).

The cell lines used were:
1) WI-38: a human embryonic lung fibroblast line, obtained from the American Type Culture Collection (ATCC CCL-75).
2) CT-1: WI-38 fibroblasts transformed by gamma-radiation, a gift from Dr M. Namba, University of Tokyo (Namba *et. al.*, 1980).
3) SVWI-38: WI-38 fibroblasts transformed with simian virus-40 (de Haan *et. al.*, 1986).
4) P388D$_1$, a mouse tumour macrophage cell line, which were a gift from Dr L. Thilo, University of Cape Town (Koren *et. al.*, 1975).

7.3 PREPARATION AND FRACTIONATION OF CRUDE NUCLEAR EXTRACTS

The method of Dignam *et. al.* (1983) was used for the preparation of crude nuclear extracts from cultured cells. For small-scale preparations of nuclear extracts, the method of Lee and Green (1990) was used.

7.3.1 PREPARATION OF DIGNAM NUCLEAR EXTRACTS

The monolayers were harvested with a rubber policeman in ice cold PBS, pelleted at 2000 rpm and 4^{O} C for 3 min in a Beckman Model TJ-6 centrifuge, washed in 5 packed cell volumes (PCV) of ice cold PBS and resuspended in 5 PCV off buffer A. For suspension cultures, the medium was centrifuged for 5 min to pellet the cells, washed and resuspended in buffer A as above. The cells were allowed to swell for 10 min on ice, pelleted and resuspended in 2 PCV (original PCV) of buffer A. The resuspended cells were lysed with 10-15 strokes of a Dounce Homogeniser and checked microscopically for cell lysis. The nuclei were pelleted,

the supernatant discarded and centrifuged again at 13 000 rpm and 4^{O} C for further 20 min in a Beckman JA 20 rotor to remove all trace amounts of cytoplasmic material. The nuclei were resuspended in 3 ml of buffer C per 10^{9} cells and homogenised with 10-15 strokes of a Dounce Homogeniser. The homogenate was transferred to a 50 ml tube, stirred gently for 30 min on ice to extract the 0.42 M salt soluble proteins and pelleted at 13 000 rpm and 4^{O} C in a Beckman JA 20 rotor for 30 min to remove the nuclear debri. The supernatant was dialysed against 50 volumes of buffer D for 5 hrs at 4^{O} C, the dialysate centrifuged at 13 000 rpm and 4^{O} C for 20 min in a JA 20 rotor and the supernatant stored in 20 to 100 ul aliquots at -70^{O} C.

7.3.2 SMALL-SCALE NUCLEAR EXTRACT PREPARATION

The monolayer and macrophage cultures were harvested as described above, washed in 30 PCV of ice cold PBS, resuspended in 1 PCV of buffer A and allowed to swell on ice for 5 min. The cell suspension was lysed by 5 cycles of slow aspiration and rapid ejection using a 1 ml syringe with a 26-gauge needle. The syringe was rinsed with buffer A prior to use to prevent excess air accumulation in the cell suspension. Cell lysis was checked microscopically and the homogenate was centrifuged for 20 sec at 4^{O} C in a microfuge. The nuclei were resuspended in two-thirds PCV of buffer C, stirred on ice for 30 min and centrifuged for 5 min at full spead in a microfuge to pellet the nuclear debris. The supernatant was dialysed for 2 hours against buffer D at 4^{O} C. The dialysate was divided into small aliquots, frozen in liquid nitrogen and stored at -70^{O} C.

7.3.3 DIFFERENTIAL SALT EXTRACTION

Nuclei were prepared as described above (section 7.3.1), resuspended in 3 ml buffer C containing 0.1 M NaCl, homogenised with 15 strokes of a dounce homogeniser and stirred gently on ice for 20 min. The nuclei were pelleted at 13 000 rpm and in a Beckman JA 20 rotor at 4^{O} C and the supernatant transferred to a fresh tube. The nuclei were re-extracted stepwise with buffer C containing 0.2 M, 0.3 M or 0.4 M NaCl as described above. The 0.1 M, 0.2 M, 0.3 M

and 0.4 M NaCl supernatants were dialysed against buffer D for 5 hours and the dialysates centrifuged at 13 000 rpm for 20 min in a JA 20 rotor to remove any debris. The supernatants were divided into small aliquots, quick frozen in liquid nitrogen and stored at -70^{o} C.

7.3.4 HEPARIN-AGAROSE CHROMATOGRAPHY

One ml of de-aerated heparin-agarose (Sigma) was packed into a siliconised pasteur pipette plugged with glasswool. The column was equilibrated with 5 to 10 ml of buffer CB (section 7.7) containing 0.1 M KCl and a flow rate of 6 ml/hour. All manipulations were performed at 4^{o} C. Columns were stored in buffer CB containing 0.1 M KCl and 0.02% sodium azide and washed with 5 to 10 ml of 0.1 M CB to remove the sodium azide prior to use.

Six to 12 mg of crude nuclear extract was applied to the column and allowed to flow under gravity. The column was washed with about 7 ml of buffer CB containing 0.1 M KCl and eluted stepwise with 5 ml buffer CB containing 0.2 M or 0.4 M KCl at a flow rate of 6 ml/hour. The columns were regenerated with buffer CB containing 1.0 M KCl to remove any trace amounts of bound protein and re-equilibrated with buffer containing 0.1 M KCl.

The elution of protein from the column was monitored by absorbance at 280 nm (LKB 2158 Uvicord). The fractions were concentrated using Centricon 10 microconcentrators (Amicon corporation, Massachusetts, USA) and the salt concentration adjusted to 0.1 M KCl. The protein concentration of each fraction was determined by the Bradford (1976) method. The electrophoretic mobility shift assay was used to test for DNA-binding activity in each fraction.

7.4 DNA-PROTEIN INTERACTIONS

7.4.1 END-LABELLING OF DNA

DNA fragments were end-labelled using either the Klenow end-labelling method or by kinasing 5'-ends. The Klenow fragment of *E. Coli* DNA polymerse was used to end-label 3'-recessed

ends of double-stranded DNA as previously described (Sambrook *et. al.*, 1989). One-unit of the Klenow DNA polymerase was added to 0.5 ug of DNA in klenow buffer in a final volume of 20 ul. The reaction mixture was incubated at room temperature for 1 to 2 min before 3 ul of dNTP mix containing 3.3 mM each of dATP, dTTP and dGTP and 2 ul (20 uCi) of [α-^{32}P]-dCTP were added. The reactions were incubated for a further 15 min.

Double-stranded oligonucleotides or dephosphorylated linearised DNA fragments were end-labelled using T4 polynucleotide kinase (Ausubel *et. al.*, 1987). One unit of T4 polynucleotide kinase was added per 0.5 ug of dephosphorylated DNA in polynucleotide kinase buffer containing [gamma-^{32}P]-ATP with a final volume of 25 ul and incubated at 37^{O} C for 60 min.

The labelling reactions were stopped by extraction with an equal volume of 24:1 (v/v) chloroform:isoamylalcohol. The unincorporated nucleotides in the aqueous phase was separated from the labeled DNA by chromatography on Sephadex G-50.

When necessary, 1 unit of calf intestine alkalinal phosphatase (CIAP) (Boehringer) was added to DNA and incubated for 30 min at 37^{O} C. At the conclusion of the incubation 1/10 volume of 500 mM EGTA was added and incubated for 45 min at 65^{O} C to inactivate the CIAP. The samples were extracted twice with an equal volume of TE-saturated phenol and once with 1 volume 24:1 (v/v) chloroform isoamylalcohol. The DNA was ethanol precipitated from the aqueous phase, washed, dried under vacuum and resuspended in TE to give a final concentration of 1 ug/ul.

7.4.2 ELECTROPHORETIC MOBILITY SHIFT ASSAYS (EMSA)

Three to 4 ug of crude nuclear extract or partially purified protein fractions were incubated in incubation buffer (IB) containing 4 ug poly dIdC.poly dIdC in a final volume of 20 ul for 10 min at room temperature. One-ng (10^4 cpm) ^{32}P-labelled probe was added and the incubation continued for another 30 minutes at 4^{O} C (Fried and Crothers, 1981). After adding 2 ul of 0.25% bromophenol blue in 35% glycerol, the

DNA-protein complexes were immediately resolved on 5% non-denaturing polyacrylamide gels at 140 V for about 2 hrs at 4^{O} C using 0.5X TBE. After drying, the gels were exposed to X-ray film for 8 to 20 hours.

In competition experiments (Garthew *et. al.*, 1985; Singh *et. al.*, 1986) a 25- to 200-fold molar excess of the appropriate double stranded oligonucleotide was incubated with nuclear extract for 10 min prior to the addition of probe.

Autoradiograms were scanned using a CliniScan densitometer or a Beckman DU 650 spectrophotometer with gel scanning accessory to determine the relative intensities of the retarded bands. In these instances, the gels were exposed to X-ray film so that the image was within the linear range of the film.

7.4.3 MUTUALLY EXCLUSIVE BINDING ASSAYS

Five, 10, 15 and/or 20 ug of partially purified protein fractions were incubated in IB containing an equivalent amount of poly dIdC.poly dIdC (i.e. 5, 10, 15 or 20 ug) in a final volume of 37 ul for 10 min at room temperature. One-ng (10^4 cpm) ^{32}P-labeled DNA was added and the incubation continued for another 20 minutes at 4^{O} C so that DNA-protein complexes could form. Three-ul of partially purified fractions containing different DNA-binding proteins was added and incubated for a further 20 minutes at 4^{O} C. After adding 2 ul 0.25% bromophenol blue in 35% glycerol, the DNA-protein complexes were immediately resolved on 5% non-denaturing polyacrylamide gels at 4^{O} C using 0.5X TBE, the gels dried and exposed to X-ray film for 8 to 20 hours

7.4.4 METHYLATION INTERFERENCE ASSAY

The method of Ausubel *et. al.* (1987) was modified as described below. About 5 ug of plasmid pNJ-59 was linearised with either Hind III or Eco R1, dephosphorylated and end-labelled as described in section 7.4.1. The phosphorylated plasmid was digested with either Eco R1 or Hind III to produce a probe labelled at one end. The digests were

resolved on 7% polyacrylamide gels and the appropriate band was eluted as described in section 7.1.1.1. To methylate the DNA probe, 1 ul of DMS was added to 10^6 cpm kinased probe in a final volume of 30 ul IB and incubated at room temperature for 10 min. Forty-ul of DMS stop buffer (section 7.7), 1 ul of 10 ug/ul tRNA, 130 ul 0.3 M sodium acetate and 1mM EDTA and 600 ul absolute ethanol was added and the DNA precipitated in a dry ice/ethanol bath for 15 min. The DNA pellet was resuspended in 220 ul 0.3 M sodium acetate, 1 mM EDTA and re-precipitated by the addition of 750 ul absolute ethanol as above. The precipitation was repeated for a third time, washed with 70% ethanol, dried under vacuum and resuspended in TE to give a final concentration of 2 X 10^4 cpm/ul.

About 80 ug of crude nuclear extract and 20 ug poly dIdC.poly dIdC was incubated at room temperatue in a final volume of 80 ul IB. After 10 min, 2-5 X 10^5 cpm of methylated probe was added and the incubation continued for another 30 min at 4^O C. After the addition of 4 ul of 0.25% bromophenol blue in 35% glycerol, the DNA-protein complexes were immediately resolved on 5% non-denaturing polyacrylamide gels at 4^O C using 0.5X TBE. The resolved complexes and free probe were electrophoretically transferred from the gel onto DEAE membranes at 500 mA for 1 hour at room temperature in 1X TBE. The DEAE membranes were sealed in a plastic bag and the complexes visualised by autoradiography (1 hr at 4^O C). The complexes and free probe bands were excised, placed in Eppendorf tubes followed by the addition of 200 ul DEAE elution buffer, 2 ul of 10% SDS, and incubated for 60 min at 65^O C. The supernatants were transferred to a separate tubes, the membranes washed with an equal volume of TE and the supernatants pooled. The residual debris were removed by centrifugation and the supernatants transferred to a fresh tube. After adding 1 ul of 10 ug/ul tRNA, the eluted DNA was precipitated by the addition of 3 volumes of absolute ethanol, left at -70^O C for 30 min, washed with 70% ethanol and dried under vacuum.

The dried pellets were resuspended in 1 M piperidine and heated for 30 min at 90^O C to cleave at the modified sites. The cleaved DNA was transferred to a fresh tube, 1.2 ml

1-butanol added and the samples vortexed vigorously until a single phase was obtained. After centrifugation for 5 min, the DNA pellet was resuspended in 150 ul 1% SDS, extracted with 1.2 ml 1-butanol as before, washed with 80% ethanol, dried under vacuum and resuspended in 0.5X formamide loading buffer. Equal counts of the samples were resolved on 12% polyacrylamide/urea sequencing gels, dried and exposed to X-ray film without intensifying screens for at least 16 hrs.

7.4.5 OLIGONUCLEOTIDES SYNTHESIS AND ANNEALING

A1	5'GGGCCCCCAGCCCCAGCCCTCCCATTGGTGGAGGCCC3'
A2	5'GGGCCTCCACCAATGGGAGGGCTGGGGCTGGGGGCCC3'
B1	5'GGGCCCCCAGCCCCAGCCCTCCCATTGG3'
B2	5'CCAATGGGAGGGCTGGGGCTGGGGGCCC3'
C1	5'GGGCCCCCAGCCCCAGCCCTCCC3'
C2	5'GGGAGGGCTGGGGCTGGGGGCCC3'
D1	5'GGAGGCCCTTTTGGAGGCACCCTAGGGCCAGG3'
D2	5'CCTGGCCCTAGGGTGCCTCCAAAAGGGCCTCC3'
E1	5'CAGCCCTCCCATTGGTGGAGG3'
E2	5'CCTCCACCAATGGGAGGGCTG3'
F2	5'GCCTCCAAAAGGGCCTCCACCAATGGGAGGGCTGGGGCTGGGGGCCC3'

MUT-CCAAT	5'GCCTCCAAAAGGGCCTCCAGCAAGGGGAGGGCTGGGGCTGGGGGCCC3'
MUT-CME	5'GCCTCCAAAAGGGAATAAACCAATGGGAGGGCTGGGGCTGGGGGCCC3'
MUT-CME2	5'GCCTCCAAAAGTTCCTCCACCAATGGGAGGGCTGGGGCTGGGGGCCC3'
MUT-US-1	5'GCCTCCAAAAGGGCCTCCACCAATGTTATTGCTGGGGCTGGGGGCCC3'
MUT-US-2	5'GGGCCCCCAGCCCCAGCAATAACATTGGTGGAGGCCCTTTTGGAGGC3'
MUT-dUS-1	5'CAGCCCTCCATTGGTGGAGG3'
MUT-dUS-2	5'CCTCCACCAATGGAGGGCTG3'

Sense and complementary antisense oligonucleotides corresponding to regions of the 59 bp Sma 1 - Bst N1 fragment of the human α_2(I) procollagen promoter were synthesised on a Beckman system 200 A DNA synthesizer. To produce double-stranded oligonucleotides A, B, C, D, MUT-US and MUT-dUS (figures 4.7, 4.18 and 4.20) the complimentary single stranded oligonucleotides A1 and A2, B1 and B2, C1 and C2, D1 and D2, E1 and E2, MUT-US-1 and MUT-US-2 or MUT-dUS-1 and MUT-dUS-2 were denatured at 90^{O} C for 5 min and allowed to anneal at 37^{O} C for 60 min prior to cooling to room temperature.

The 47-mer double-stranded oligonucleotides F, MUT-CCAAT, MUT-CME and MUT-CME2 (figures 4.7 and 4.19) were produced by annealing single-stranded oligonucleotides A1 to F2, MUT-CCAAT, MUT-CME and MUT-CME2 as described above. The Klenow fragment of *E. Coli* DNA polymerase was used to produce double-stranded oligonucleotides as described previously in section 7.1.2.1.

7.5 SDS-POLYACRYLAMIDE GEL ELECTROPHORESIS

Protein extracts were analysed on 10 cm long 5 to 12% SDS-polyacrylamide gels (Laemmli, 1970).

Separating gel:

5 to 7.5 ml	40% Acrylamide Stock Solution (38:2)
6.25 ml	1.5 M Tris-Cl (pH 8.8)
1.25 ml	100 mM EDTA
0.25 ml	10% SDS
to 25 ml	H_2O
250 ul	10% Ammonium Persulphate
25 ul	TEMED

The separating gel was poured, a layer of water was pipetted over the gel surface and left to polymerise for at least 1 hr.

The stacking gel (Smith, I *et. al.*, 1988):

1 ml	40% Acrylamide Stock Solution (38:2)
2.5 ml	0.5 M Tris-Cl (pH 6.8)
0.5 ml	100 mM EDTA
0.1 ml	10% SDS
0.1 ml	0.4% Phenol Red
5.7 ml	H_2O
100 ul	10% Ammonium Persulphate
10 ul	TEMED

An equal volume of 2X treatment buffer and 2 ul of DNA stop buffer was added to between 20 and 50 ug of protein. The samples were heated at 90^{O} C for 3 min, loaded onto the gel and electrophoresed at 20 mA until the samples had run into the separating gel and thereafter at 25 mA in 1X SDS-PAGE

tank buffer. The gels were stained for 4 hrs in Coomassie blue stain and destained for 1 hour in solution 1 and at least 6 hrs with 2 changes of solution 2.

When necessary, dilute protein samples were concentrated acetone precipitated according to Hager and Burgess, 1980. Briefly, 5 volumes of cold acetone (-20^{O} C) was added to each sample and incubated on ice for 30 min. After centrifugation the pellet was washed with absolute ethanol, dried under vacuum and resuspended in 20 ul treatment buffer.

7.6 PHYSICO-CHEMICAL ANALYSIS OF DNA-BINDING PROTEINS

Analytical gel filtration chromatography and density gradient centrifugation (Siegel and Monty, 1966), UV crosslinking (Chodosh *et. al.*, 1986) and south-western blotting (Bowen *et. al.*, 1980) were employed to determine the molecular weights of the *trans*-acting factors in crude or partially purified nuclear extracts.

7.6.1 ANALYTICAL GEL FILTRATION CHROMATOGRAPHY

A de-aerated 80% slurry of Sephacryl S-300 (Pharmacia) in buffer CB containing 0.1 M KCl was packed into a Pharmacia column (1.6 cm X 100 cm) at a flow rate of 1 ml/min at 4^{O} C. The column was equilibrated with CB containing 0.1 M KCl at a flow rate of 6 ml/h (3 $ml/cm^2/h$).

Sephadex G-200 columns were prepared as a 80% slurry (Fine grade, Pharmacia) in CB containing 0.8 M KCl as recommended by the manufactures. The slurry was de-aerated and packed into a 1 cm X 60 cm column at a flow rate of 1 ml/min at 4^{O} C. The column was equilibrated with CB containing 0.8 M KCl and 0.02% sodium azide at a flow rate of 1.8 ml/h (2.3 $ml/cm^2/h$).

To calibrate the Sephacryl S-300 column, protein standards were prepared as follows: 2 mg thyroglobin, 2 mg ferritin, 4 mg catalase, 2 mg aldolase (Pharmacia), 4 mg BSA (Bayer-Miles) and/or 7 mg ovalbumin (Sigma) were dissolved in 0.5 to 1 ml of CB containing 0.1 M KCl and 30% glycerol, layered onto the column and eluted at a flow rate of 6 ml/h. Two mg

each of catalase, aldolase or BSA or 4 mg each of ovalbumin or carbonic anhydrase dissolved in 0.5 ml CB containing 0.8 M KCl were used to calibrate the Sephadex G-200 column at a flow rate of 1.8 ml/h.

The elution volumes (Ve) of the protein standards were monitored by spectrophoretic absorbance at 280 nm (LKB 2158 Uvicord SD UV-monitor). The void (Vo) and the total (Vt) volumes of the columns were determined using 0.2 mg blue dextran 2000 (Pharmacia) and 2 mg thymidine (Sigma) respectively (Kim and Sheffery, 1990). A calibration curve was prepared by plotting the log of the Stokes radius (thyroglobin, 8.50 nm; ferritin, 6.10 nm; catalase, 5.22 nm; aldolase, 4.81 nm; BSA, 3.55 nm and ovalbumin, 3.05 nm) versus k_{av}, where $k_{av}=(Ve-Vo)/(Vt-Vo)$ (Siegel and Monty, 1966).

Between 5 and 18 mg of crude nuclear extract in a final volume of approximately 2 ml of CB containing 0.1 M KCl, 30% glycerol and 0.1% Nonidet P-40 were applied to the Sephacryl S-300 column and eluted with CB containing 0.1 M KCl (without NP-40) at a flow rate of 6 ml/h. The samples were eluted, 1 ml fractions collected, frozen in liquid nitrogen and stored at -70^{o} C until further use. Between 5 and 14 ul of each fraction was used in electrophoretic mobility shift assays containing 2 ug poly dIdC.poly dIdC per reaction, as previously described (section 7.4.2).

The elution volumes of DNA-protein complexes were analysed by binding about 40 ug of crude nuclear extract with 20 ng (2×10^5 cpm) of ^{32}P-labeled oligonucleotide F in a final volume of 80 ul, as described in section 7.4.2. Half-a-ml of CB containing 0.1 M KCl, 30% glycerol and 0.1% NP-40 was added to the binding reaction, layered onto the column and eluted as described above. One-ml fractions were collected from the column and the amount of radioactivety in each fraction was determined.

One to 2 mg of crude extract in a final volume of 400 ul containing 0.8 M KCl and 0.25% NP-40 was layered on top of the Sephadex G-200 column. Once the sample had run into the column, buffer CB containing 0.8 M KCl and 0.02% sodium azide

was gently layered onto the column and the samples eluted with this buffer. The elution of protein from the column was monitored and 200 ul fractions were collected as described above. The fractions were frozen in liquid nitrogen and stored at -70^{O} C.

7.6.2 GLYCEROL GRADIENT SEDIMENTATION

Linear 15 to 40% glycerol gradients in SB (section 7.7) were prepared at 4^{O} C as described by Martin and Ames (1961). In order to calibrate the gradients, 0.5 to 0.75 mg each of catalase, BSA, ovalbumin, carbonic anhydrase and/or lysozyme in a final volume of 100 ul was layered onto the glycerol gradients. The samples were centrifuged for 16 hrs at 53 000 rpm and 4^{O} C in a Beckman SW65Ti rotor with slow deceleration. A hole was punched at the bottom of the tubes and about 36 fractions of 3 drops each were collected. The protein concentration in each fraction was determined by the Bradford (1976) method. Fifteen-ul of each fraction was also resolved on 6 or 10% SDS-polyacrylamide gels as described in section 7.5. A calibration curve was prepared by plotting % glycerol in the fractions were the standard proteins sedimented versus sedimentation coefficient (catalase, 11.3; BSA, 4.4; ovalbumin, 3.7; carbonic anhydrase, 3.2 and lysozyme, 1.9). The percent glycerol in each fractions was calculated as follows:

$$\% \text{ glycerol} = 40\% - (FN \times 25\% / TF)$$

where FN was the fraction number and TF was the total number of fractions collected.

Approximately 0.5 mg of crude nuclear extract and 0.5 mg ovalbumin in a final volume of 100 ul was layered onto the gradients and centrifuged as described above. Three drop fractions were collected from the bottom of each centrifuge tube, frozen in liquid N_2 and stored at -70^{O} C. The protein concentration and the SDS-PAGE analysis of each fraction was determined as decribed above. The DNA-binding activities in 10 ul of each fractions was quantitated by the electrophoretic mobility shift assay as described in section 7.4.2.

7.6.3 UV-CROSSLINKING

The method of Chodosh *et. al.* (1986) was used to UV crosslink DNA-binding proteins to radioactively labelled DNA probes.

7.6.3.1 PREPARATION OF Br-dU SUBSTITUTED PROBES

Complementary oligonucleotides C1 and A2 or C1 and F2 were annealed as previously described (section 7.4.5). Bromodeoxyuridine (Br-dU) and ^{32}P-deoxycytosine (^{32}P-dC) were incorporated into 0.5 ug of the annealed oligonucleotides in a final reaction volume of 20 ul containing 10 units Klenow fragment of *E. Coli* DNA polymerase, ^{32}P-dCTP and 100 uM each of dATP, dGTP and Br-dUTP. The reaction mixtures were incubated at room temperature for 15 min, after which 1 ul of 10 mM dCTP was added and the incubation continued for a further 30 minutes to produce double-stranded Br-dU and ^{32}P-dC substituted double-stranded oligonucleotides. After organic extractions, the free nucleotides were removed by Sephadex G50 gel filtration.

7.6.3.2 *IN-SITU* DNA-PROTEIN CROSSLINKING

Forty ug of crude nuclear extract was pre-incubated with 10 ug Poly dIdC.Poly dIdC in a final volume of 40 ul IB for 10 minutes at room temperature. Ten-ng (10^5 cpm) of ^{32}P-dC and Br-dU substituted probe was added and incubation continued for another 30 minutes on ice (Fried and Crothers, 1981). After adding 2 ul of 0.25% bromophenol blue in 35% glycerol, the DNA-protein complexes were immediately resolved on 5% non-denaturing polyacrylamide gels at 4^{O} C using 0.5% TBE as the tank buffer. The wet gels were sealed in plastic bags and irradiated for 30 min with a UV transilluminator at 305 nm and intensity 7000 uW/cm^2 (Fotodyne) (Wu, C *et. al.*, 1987). The UV crosslinked complexes were visualised by autoradiography of wet gels for 2 to 16 hours.

The complexes were excised from the gels, chopped into slices and eluted with 1 volume of protein elution buffer at room temperature for 3 hours with gentle shaking. The supernatants were transferred to fresh tubes, the gel slices

washed with 100 ul protein elution buffer, the supernatants pooled and centrifuged to remove any debris. Carrier BSA was added to the combined supernatants to a final concentration of 0.1 ug/ul, followed by 5 volumes of cold (-20^{O} C) acetone and the proteins were precipitated on ice for 30 minutes. The protein pellets were washed with 100% EtOH, dried under vacuum for 5 minutes and resuspended in treatment buffer (Hager and Burgess, 1980). The samples were analysed by SDS-PAGE as described in section 7.5.

7.6.3.3 DNA-PROTEIN CROSSLINKING PRIOR TO ELECTROPHORESIS

Ten to 15 ul of partially purified nuclear extracts were incubated with the DNA probes as described in section 7.6.3.2. The binding reaction mixtures were pipetted as drops onto parafilm, covered with inverted Eppendorf tubes, irradiated for 30 min at 305 nm and intensity 7000 uW/cm^2 (Fotodyne transilluminator) and transferred to Eppendorf tubes.

On occasions, 1 ul of 0.4 M $CaCl_2$, 1 unit of micrococcal nuclease and 4 ug DNase 1 were added to each UV irradiated sample and incubated at 37^{O} C for 30 min. For proteinase K treatment, 4 ug of the proteinase K was added to the reaction mixtures and incubated for 60 min at 55^{O} C.

After one volume of 2X treatment buffer was added, the samples were analysed by SDS-PAGE on 8, 10 or 12% gels (section 7.5).

7.6.4 SOUTHWESTERN BLOTTING

Thirty to 50 ug crude nuclear extract were resolved on 6 or 8% SDS-polyacrylamide gels as described in section 7.5, except that the boiling step prior to loading was omitted. The gel was incubated for 20 min in 200 ml of electroblotting buffer with gentle shaking and the proteins electrophoretically transferred from the gel to Amersham Hybond-C nitrocellulose membranes overnight at 10 V followed by 40 V for a further 30 min at 4^{O} C in electroblotting buffer (Towbin *et. al.*, 1979).

The filters were incubated in 50 ml blocking buffer for at least 2 hours at room temperature with gentle shaking. The blocked filters were incubated for a further 2 hours at room temperature in Seal-a-Meal freezer bags containing 2 to 10 ml binding buffer and 5 X 10^5 cpm/ml (50 ng/ml) of klenow labelled DNA probe (Matsuno *et. al.*, 1989). The filters were washed at room temperature with 3 X 50 ml changes of washing buffer for 10 min each with gentle shaking. The washed filters were sealed in plastic bags and exposed to X-ray film.

On occasions filters were stained im 0.3% amido black in 20% ethanol and 7% acetic acid for 15 min with gentle shaking and destained in warm (50^{O} C) 10% acetic acid.

7.7 BUFFERS AND SOLUTIONS

DNA Elution Buffer
0.5 M Ammonium Acetate
10 mM Magesium Acetate
1 mM EDTA

1X TE Buffer
10 mM Tris-Cl (pH 8.0)
1 mM EDTA

1X TBE Electrophoresis Buffer
90 mM Tris-borate (pH 8.0)
90 mM Boric Acid
2.5 mM EDTA

1X TAE Electrophoresis Buffer
40 mM Tris-acetate
5 mM Sodium Acetate
1 mM EDTA

10X Klenow Buffer
0.1 M Tris-Cl (pH 7.5)
0.1 M $MgCl_2$
0.5 M NaCl
10 mM DTE

10X T4 DNA Polymerase Buffer
0.33 M Tris-acetate (pH 7.9)
0.66 M Potassium acetate
0.1 M Magnesium acetate
5 mM DTT
1 mg/ml BSA

L-Broth
10 g/l Tryptone
5 g/l Yeast Extract
5 g/l NaCl

"RAPS" Solution 1

25 mM Tris-Cl (pH 8.0)

10 mM EDTA

50 mM Glucose

"RAPS" Solution 2

200 mM NaOH

1% SDS

"RAPS" Solution 3

3 M Potassium Acetate (pH 4.8)

Sucrose/Tris Solution

50 mM Tris-Cl (pH 8.0)

25% Sucrose

TTE

50 mM Tris-Cl (pH 8.0)

0.5% Triton X-100

6.25 mM EDTA

PBS (pH 7.3)

137 mM NaCl

2.7 mM KCl

4.3 mM $Na_2HPO_4.7H_2O$

1.4 mM KH_2PO_4

Dignam Buffer A

10 mM Hepes (pH 7.9)

1.5 mM $MgCl_2$

10 mM KCl

0.5 mM DTT

Dignam Buffer C

20 mM Hepes (pH 7.9)
25% glycerol
0.42 M NaCl
1.5 mM $MgCl_2$
0.2 mM EDTA
0.5 mM DTT
0.5 mM PMSF
1 ug/ml leupeptin
1 ug/ml pepstatin A

Dignam Buffer D

20 mM Hepes (pH 7.9)
20% glycerol
0.1 M KCl
0.2 mM EDTA
0.5 mM DTT
0.5 mM PMSF
1 ug/ml leupeptin
1 ug/ml pepstatin A

10X Polynucleotide Kinase Buffer

500 mM Tris-Cl (pH 7.5)
100 mM $MgCl_2$
50 mM DTT
0.5 mg/ml BSA

5X IB

100 mM Hepes (pH 7.9)
250 mM KCl
2.5 mM DTT
1.0 mM EDTA
5 mM $MgCl_2$
20% Ficoll 400

DMS Stop Buffer

1.5 M Sodium Acetate (pH 7.0)
1 M ß-mercaptoethanol

DEAE Elution Buffer
10 mM Tris-Cl (pH 8.0)
1 mM EDTA
1 M NaCl

Formamide Loading Buffer
95% formamide
20 mM EDTA
0.05% bromophenol blue
0.05% xylene cyanol

CB Buffer
50 mM Tris-Cl (pH 7.9)
0.1 mM EDTA
0.5 mM DTT
20% Glycerol
0.5 mM PMSF
1 ug/ml leupeptin
1 ug/ml pepstatin A

40% Acrylamide Stock Solution (38:2)
38% Acrylamide
2% BIS
2% Ion exchanger V

The solution was mixed on a roller for 16 hrs at 4^{O} C, filtered and stored at 4^{O} C.

2X Treatment Buffer
0.125 M Tris-Cl (pH 6.8)
4% SDS
20% Glycerol
5 mM EDTA
10% 2-Mercaptoethanol

DNA Stop Buffer
0.05% Bromophenol Blue
0.5% SDS
1 mM EDTA
50% Glycerol

5X SDS-PAGE Tank Buffer

0.25 M Tris-Cl (pH 8.3)
1.92 M Glycine
0.5% SDS

Coomassie Blue Stain

0.125% Coomassie Blue R-250
50% Methanol
10% Acetic Acid

Destain Solution 1

50% Methanol
10% Acetic Acid

Destain Solution 2

7% Methanol
5% Acetic Acid

SB Buffer

20 mM Hepes (pH 7.9)
0.1 M KCl
0.2 mM EDTA
0.5 mM PMSF
0.5 mM DTT
1 ug/ml leupeptin
1 ug/ml pepstatin A

Protein Elution Buffer

50 mM Tris-Cl (pH 8.0)
0.1 mM EDTA
150 mM NaCl
0.1% SDS
5 mM DTT
0.5 mM PMSF
1 ug/ml leupeptin
1 ug/ml pepstatin A

Electroblotting Buffer

0.05 M Tris
0.19 M Glycine
20% Methanol

Southwestern Blocking Buffer

10 mM Hepes (pH 7.9)

0.5% Carnation low fat milk powder

Southwestern Binding Buffer

10 mM Hepes (pH 7.9)

50 mM NaCl

10 mM $MgCl_2$

0.1 mM EDTA

1 mM DTT

0.25% Carnation low fat milk powder

Southwestern Washing Buffer

10 mM Hepes (pH 7.9)

50 mM NaCl

10 mM $MgCl_2$

0.1 mM EDTA

REFERENCES

Adams, S.L. (1989) *American J. Resp. Cell and Mol. Biol.* **1,** 161-168.

Agoff, S.N., Hou, J., Linzer, D.I.H. and Wu, B. *Science* **259,** 84-87.

Andersen, R.D., Taplitz, S.J., Wong, S., Bristol, G., Larkin, B. and Herschman, H.R. (1987) *Mol. Cell. Biol.* **7,** 3574-3581.

Angel, P., Imagawa, M., Chiu, R., Stein, B., Imbra, R.J., Rhamsdorf, H.J., Jonat, C., Herrlich, P. and Karin, M. (1988) *Cell* **49,** 729-739.

Apte, S., Mattei, M-G. and Olsen, B.R. (1991) *FEBS Lett.* **282,** 393-396.

Apte, S., Seldin, M.F., Hayashi, M. and Olsen, B.R. (1992) *Eur. J. Biochem.* **224,** 217-224.

Apte, S.S. and Olsen, B.R. (1993) *Matrix* **13,** 165-179.

Arcot, S.S., Flemington, E.K. and Deininger, P.L. (1989) *J. Biol. Chem.* **264,** 2343-2349.

Arndt, K. and Fink, G.R. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 8516-8520.

Augereau, P. and Chambon, P. (1986) *EMBO J.* **5,** 1791-1797.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. eds (1987) *Current Protocols in Molecular Biology*, Greene Publishing Associates, New York.

Bakker, O. and Parker, M.G. (1991) *Nucleic Acids Res.* **19,** 1213-1217.

Barnhart, K.M., Kim, C.G., Banerji, S.S. and Sheffery, M. (1988) *Mol. Cell. Biol.* **8,** 3215-3226.

Barberis, A., Superti-Furga, G. and Busslinger, M. (1987) *Cell* **50,** 347-359.

Bennett, V.D. and Adams, S.L. (1987) *J. Biol. Chem.* **262,** 14806-14814.

Bennett, V.D., Weiss, I.M. and Adams, S.L. (1989) *J. Biol. Chem.* **264,** 8402-8409.

Bennett, V.D. and Adams, S.L. (1990) *J. Biol. Chem.* **265,** 2223-2230.

Bernard, M., Yoshioka, H., Rodriguez, E., van der Rest, M., Kimura, T., Ninomiya, Y., Olsen, B.R. and Ramirez, F. (1988) *J. Biol. Chem.* **263,** 17159-17166.

Benson-Chanda, V., Su, M-W., Weil, D., Chu, M-L. and Ramirez, F. (1989) *Gene* **78,** 255-265.

Birk, D.E, Fitch, J.M., Babiarz, J.P. and Linsenmayer, T.F. (1988) *J. Biol. Chem.* **106,** 999-1008.

Birkenmeier, E.H., Gwynn, B., Howard, S., Jerry, J., Gordon, J.I., Landschulz, W.H. and Mc Knight, S.L. (1989) *Genes and Dev.* **3,** 1146-1156.

Birnboim, H.C. and Doly, J. (1979) *Nucleic acids Res.* **7,** 1513-1523.

Birnboim, H.C. (1983) *Methods in Enzymology* **100,** 243-255.

Boam, D.S.W., Clark, A.R. and Docherty, K. (1990) *J. Biol. Chem.* **265,** 8285-8296.

Boast, S., Su, M-W., Ramirez, F., Sanchez, M. and Avvedimento, E.V. (1990) *J. Biol. Chem.* **265,** 13351-13356.

Bornstein, P. and Sage, H. (1980) *Ann. Rev. Biochem.* **49,** 957-1003.

Bornstein, P., Mc Kay, J., Morishima, J.K., Devarayalu, S. and Gelinas, R.E. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 8869-8873.

Bornstein, P., Mc Kay, J., Liska, D.J., Apone, S. and Devarayalu, S. (1988) *Mol. Cell. Biol.* **8,** 4851-4857.

Bornstein, P. and Mc Kay, J. (1988) *J. Biol. Chem.* **263,** 1603-1606.

Bornstein, P. and Sage, H. (1989) *Prog. Nucleic Acid Res. Mol. Biol.* **37,** 67-106.

Bowen, B., Steinberg, J., Laemmli, U.K. and Weintraub, H. (1980) *Nucleic Acids Res.* **8,** 1-20.

Boyes, J. and Bird, A. (1991) *Cell* **64,** 1123-1134.

Bradford, M.M. (1976) *Anal. Biochem.* **72,** 248-254.

Breathnach, R. and Chambon, P. (1981) *Ann. Rev. Biochem.* **50,** 349-383.

Breitman, M.L., Clapoff, S., Rossant, J., Tsui, L-C., Globe, M., Maxwell, I.H. and Bernstein, A. (1987) *Science* **238,** 1563-1565.

Brenner, D.A., Rippe, R.A. and Veloz, L. (1989) *Nucleic Acids Res.* **17,** 6055-6064.

Brewton, R.G., Ouspenskaia, M.V., van der Rest, M. and Mayne, R. (1992) *Eur. J. Biochem.* **205,** 443-449.

Briggs, M.R., Kadonaga, J.T., Bell, S.P. and Tjiam, R. (1986) *Science* **234,** 47-52.

Bruggeman, L.A., Xie, H-X., Brown, K. and Yamada, Y. (1991) *Teratology* **44,** 203-208.

Bruggeman, L.A; Burbelo, P.D., Yamada, Y. and Klotman, P.E. (1992) *Oncogene* **7,** 1497-1502.

Bucher, P. and Trifonov, E.N. (1986) *Nucleic Acids Res.* **14,** 10009-10026.

Burbelo, P.D., Martin, G.R. and Yamada, Y. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 9679-9682.

Burbelo, P.D., Horikoshi, S. and Yamada, Y. (1990) *J. Biol. Chem.* **265,** 4839-4843.

Burbelo, P.D., Bruggeman, L.A., Gabriel, G.C., Klotman, P.E. and Yamada, Y. (1991) *J. Biol. Chem.* **266,** 22297-22302.

Burgeson, R.E. (1988) *Ann. Rev. Cell Biol.* **4,** 551-577.

Busslinger, M., Portman, R., Irminger, J.C. and Birnstiel, M.L. (1980) *Nucleic Acids Res.* **8,** 957-977.

Butel, J.S. and Jarvis, D.L. (1986) *Biochim. Biophys. Acta.* **6,** 779-791.

Caridi, G., Pezzolo, A., Bertelli, R., Gimelli, G., Donato, A., Candiano, G. and Ghiggeri, G.M. (1992) *Hum. Genet.* **90,** 174-176.

Carthew, R.W., Chodosh, L.A. and Sharp, P.A. (1985) *Cell* **43,** 439-448.

Cedar, H. (1988) *Cell* **53,** 3-4.

Cereghini, S., Blumenfeld, M. and Yaniv, M. (1988) *Genes and Dev.* **2,** 957-974.

Chang, C-J., Chen, T-T., Lei, H-Y., Chen, D-S. and Lee, S-C. (1990) *Mol. Cell. Biol.* **10,** 6642-6653.

Chen, W and Struhl, K (1988) *Proc. Natl. Acad. Sci. USA* **85,** 2991-2695.

Chiu, R., Boyle, W.J., Meek, J., Smeal, T., Hunter, T. and Karin, M. (1988) *Cell* **54,** 541-552.

Chodosh, L.A., Carthew, R.W. and Sharp, P.A. (1986) *Mol. Cell. Biol.* **6,** 4723-4733.

Chodosh, L.A., Baldwin, A.S., Carthew, R.W. and Sharp, P.A. (1988a) *Cell* **53,** 11-24.

Chodosh, L.A., Olesen, J., Hahn, S., Baldwin, A.S., Guarente, L. and Sharp, P.A. (1988b) *Cell* **53,** 25-35.

Choe, I,. Agcock, R.S., Raghow, R., Myers, J.C., Seyer, M. and Kang, A.H. (1987) *J. Biol. Chem.* **262,** 5408-5413.

Chow, K-L. and Schwartz, R.J. (1990) *Mol. Cell. Biol.* **10,** 528-538.

Christy, R.J., Yang, V.W., Ntambi, J.M., Geiman, D.E., Landschulz, W.H., Friedman, A.D., Nakabeppu, Y., Kelly, T.J. and Lane, M.D. (1989) *Genes and Dev.* **3,** 1323-1335.

Chu, M-L., de Wet, W., Bernard, M., Ding J-F., Morabito, M., Myers, J., Williams, C. and Ramirez, F. (1984) *Nature* **310,** 337-340.

Chu, M-L., Weil, D., de Wet, W., Bernard, M., Sippola, M. and Ramirez, F. (1985) *J. Biol. Chem.* **260,** 4357-4363.

Chu, M-L., Conway, D., Pan, T-C., Baldwin, C., Mann, K., Deutzmann, R. and Timpl, R. (1988) *J. Biol. Chem.* **263,** 18601-18606.

Chu, M-L., Pan, T-C., Conway, D., Kuo, H-J., Glanville, R.W., Timpl, R., Mann, K., and Deutzmann, R. (1989) *EMBO J.* **8,** 1939-1946.

Chu, M-L., Zhang, R-Z., Pan, T-C., Stokes, D., Conway, D., Kuo, H-J., Glanville, R., Mayer, U., Mann, K., Deutzmann, R. and Timpl, R. (1990) *EMBO J.* **9,** 385-393.

Clerc, R.G., Corcoran, L.M., LeBowitz, J.H., Baltimore, D. and Sharp, P.A. (1988) *Genes and Dev.* **2,** 1570-1581.

Cleveland, D.W. (1991) *Curr. Opinion in Cell Biol.* **1,** 1148-1153.

Cohen, R.B., Sheffery, M. and Kim, C.C. (1986) *Mol. Cell. Biol.* **6,** 821-832.

Conaway, J.W. and Conaway, R.C. (1991) *J. Biol. Chem.* **266,** 17721-17724.

Crossley, M. and Brownlee, G.G. (1990) *Nature* **345,** 444-446.

Culotta, V.C. and Hamer, D.H. (1989) *Mol. Cell. Biol.* **9,** 1376-1380.

Cutting, G.R., Mc Ginniss, M.J., Kasch, L.M., Tsipouras, P. and Antonarakis, S.E. (1990) *Genomics* **8,** 407-410.

Davison, B.L., Leighton, T. and Rabinowitz, J.C. (1979) *J. Biol. Chem.* **254,** 9220-9226.

de Crombrugghe, B., Karsenty, G., Maity, S., Vuorio, T., Rossi, P., Ruteshouser, E.C., Mc Kinney, S.H. and Lozano, G. (1990) *Ann. N.Y. Acad. Sci.* **580,** 88-96.

de Haan, J.B., Gevers, W. and Parker, M.I. (1986) *Cancer Res.* **46,** 713-716.

Descombes, P., Chojkier, M., Lichtsteiner, S., Falvey, E. and Schibler, U. (1990) *Genes and Dev.* **4,** 1541-1551.

de Wet, W., Bernard, M., Benson-Chanda, V., Chu, M-L., Dickson, L., Weil, D. and Ramirez, F. (1987) *J. Biol. Chem.* **262,** 16032-16036.

Dickson, L.A., de Wet, W., Liberto, M.D., Weil, D. and Ramirez, F. (1985) *Nucleic Acids Res.* **13,** 3427-3438.

Didier, D.K., Schiffenbauer, J., Woulfe, S.L., Zacheis, M. and Schwartz, B.D. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 7322-7326.

Dignam, J.D., Lebovitz, R.M and Roeder, R.G. (1983) *Nucleic Acids Res.* **11,** 1475-1489.

Doerfler, W. (1983) *Ann. Rev. Biochem.* **52,** 93-124.

Dorn, A., Bollekens, J., Staub, A., Benoist, C. and Mathis, D. (1987) *Cell* **50,** 863-872.

Dozin, B., Quarto, R., Rossi, F. and Cancedda, R. (1990) *J. Biol. Chem.* **265,** 7216-7220.

Dublet, B. and van der Rest, M. (1991) *J. Biol. Chem.* **266,** 6853-6858.

Dynan, W.S. and Tjian, R. (1983) *Cell* **32,** 669-680.

Dynan, W.S. (1989) *Cell* **58,** 1-4.

Ehrlich, R., Maguire, J.E. and Singer, D.S. (1988) *Mol. Cell. Biol.* **8,** 695-703.

Emanuel, B.S., Sellinger, B.T., Gudas, L.J. and Myers, J.C. (1986) *Am. J. Hum. Genet.* **38,** 38-44.

Engel, J., Furthmayr, H., Odermatt, E., von der Mark, H., Aumailley, M., Fleischmajer, R. and Timpl, R. (1990) *Ann. N.Y. Acad. Sci.* **580,** 25-37.

Errington, J. (1990) *Nucleic Acids Res.* **18,** 5324.

Eyre, D.R., Apon, S., Wu, J-J., Ericsson, L.H., Walsh, K.A. (1987) *FEBB Lett.* **220,** 337-341.

Faber, M. and Sealy, L. (1990) *J. Biol. Chem.* **265,** 22243-22254.

Fessler, J.H. and Fessler, L.I. (1978) *Ann. Rev. Biochem.* **47,** 129-162.

Fessler, L.I., Shigaki, N. and Fessler, J.H. (1985) *J. Biol. Chem.* **260,** 13286-13293.

Fessler, J.H. and Fessler, L.I. (1987) in Mayne and Burgeson, 81-103.

Finer, M.H., Aho, S., Gerstenfeld, L.C., Boedtker, H. and Doty, P. (1987) *J. Biol. Chem.* **262,** 13323-13332.

Fisher, P.A. (1991) *Curr. Opinion in Cell Biol.* **1,** 447-453.

Flemington, E., Bradshaw, H.D., Traina-Dorge, V., Slagel, V. and Deininger, P.L. (1987) *Gene* **52,** 267-277.

Forsburg, S.L. and Guarente, L. (1989) *Genes and Dev.* **3,** 1166-1178.

Foulkes, N.S. and Sassone-Corsi, P. (1992) *Cell* **68,** 411-414.

Framson, P. and Bornstein, P. (1993) *J. Biol. Chem.* **268,** 4989-4996.

Francomano, C.A., Cutting, G.R., Mc Cormick, M.K., Chu, M-L., Timpl, R., Hong, H. and Antonarakis, S.E. (1991) *Hum. Genet.* **87,** 162-166.

Freifelder, D. (1982) *Physical Biochemistry: Applications to Biochemistry and Molecular Biology.* **2nd edition,** W.H. Freeman and Company, New York.

Fried, M. and Crothers D.M. (1981) *Nucleic Acids Res.* **9,**6505-6525.

Friedman, A.D., Landschulz, W.H. and Mc Knight, S.L. (1989) *Genes and Dev.* **3,** 1314-1322.

Fromm, M. and Berg, P. (1982) *J. Mol. Appl. Genet.* **1,** 457-481.

Furth, J.J., Wroth, T.H. and Ackerman, S. (1991) *Exp. Cell Research* **192,** 118-121.

Gallinari, P., La Bella, F. and Heintz, N. (1989) *Mol. Cell. Biol.* **9,** 1566-1575.

Ghazal, P., Lubon, H., Fleckenstein, B. and Hennighausen, L. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 3658-3662.

Ghosh, D. (1990) *Nucleic Acids Res.* **18,** 1749-1756.

Gidoni, D., Dynan, W.S. and Tjian, R. (1984) *Nature* **312,** 409-413.

Gidoni, D., Kadonaga, J.T., Barrera-Saldana, H., Takahashi, K., Chambon, P. and Tjian, R. (1985) *Science* **230,** 511-517.

Gil, G., Smith, J.R., Goldstein, J.L., Slaughter, C.A., Orth, K., Brown, M.S. and Osborne, T. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 8963-8967.

Goldberg, M.L. (1979) *PhD Thesis*, Stanford University.

Goldberg, H., Helaakoski, T., Garrett, L.A., Karsenty, G., Pellegrino, A., Lozano, G., Maity, S. and de Crombrugghe, B. (1992) *J. Biol. Chem.* **267,** 19622-19630.

Goldberg, Y.P., Leaner, V.D. and Parker, M.I. (1992) *J. Cell. Biochem.* **49,** 74-81.

Gordon, M.K., Gerecke, D.R. and Olsen, B.R. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 6040-6044.

Gordon, M.K., Gerecke, D.R., Dublet, B., van der Rest, M. and Olsen, B.R. (1989) *J. Biol. Chem.* **264,** 19772-19778.

Gounari, F., De Francesco, R., Schmitt, J., van der Vliet, P.C., Cortese, R. and Stunnenberg, H. (1990) *EMBO J.* **9,** 559-566.

Goyal, N., Knox, J. and Gronostajski, R.M. (1990) *Mol. Cell. Biol.* **10,** 1041-1048.

Graves, B.J., Johnson, P.F. and McKnight, S.L. (1986) *Cell* **44,** 565-576.

Greene, J.M., Larin, Z., Taylor, I.C.A., Prentice, H., Gwinn, K.A. and Kingston, R.E. (1987) *Mol. Cell. Biol.* **7,** 3646-3655.

Greenspan, D.S., Cheng, W. and Hoffman, G.G. (1991a) *J. Biol. Chem.* **266,** 24727-24733.

Greenspan, D.S., Lee, S.T., Lee, B.S. and Hoffman, G.G. (1991b) *Gene Expr.* **1,** 29-39.

Greenspan, D.S., Byers, M.G., Eddy, R.L., Hoffman, G.G. and Shows, T.B. (1993) *Cytogenet. Cell Genet.* **62,** 35-36.

Griffin, C.A., Emanuel, B.S., Hansen, J.R., Cavenee, W.K. and Myers, J.C. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 512-516.

Gronostajski, R.M. (1986) *Nucleic Acids Res.* **14,** 9117-9131.

Gross, D.S. and Garrad, W.T. (1987) *Trends Biochem. Sci.* **12,** 293-297.

Gross, D.S. and Garrad, W.T. (1988) *Ann. Rev. Biochem.* **57,** 159-198.

Guenette, D.K., Ritzenthaler, J.D., Foley, J., Jackson, J.D. and Smith, B.D. (1992) *Biochem J.* **283,** 699-703.

Hager, D.A. and Burgess, R.R. (1980) *Anal. Biochem.* **109,** 76-86.

Har-El, R., Sharma, Y.D., Aguilera, A., Ueyama, N., Wu, J-J., Eyre, D.R., Juricic, L., Chandrasekaran, S., Li, M., Nah, H-D., Upholt, W.B. and Tanzer, M.L. (1992) *J. Biol. Chem.* **267,** 10070-10076.

Harada, H., Fujita, T., Miyamoto, M., Kimura, Y., Maruyama, M., Furia, A., Miyata, T. and Taniguchi, T. (1989) *Cell* **58,** 729-739.

Harrison, S.C. (1991) *Nature* **353,** 715-719.

Hartung, S., Jaenisch, R. and Breindi, M. (1986) *Nature* **320,** 365-367.

Harvey, R.P., Robins, A.J. and Wells, J.R.E. (1982) *Nucleic Acids Res.* **10,** 7851-7863.

Hata, R. and Peterkofsky, B. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 2933-2937.

Hatamochi, A., Paterson, B., and de Crombrugghe, B. (1986) *J. Biol. Chem.* **261,** 11310-11314.

Hatamochi, A., Golumbek, P.T., van Schaftingen, E. and de Crombrugghe, B. (1988) *J. Biol. Chem.* **263,** 5940-5947.

Hatamochi, A., Aumailley, M., Mauch, C., Chu, M-L., Timpl, R. and Krieg, T. (1989) *J. Biol. Chem.* **264,** 3494-3499.

Heckmann, M., Aumailley, M., Chu, M-L., Timpl, R. and Krieg, T. (1992) *FEBS Lett.* **310,** 79-82.

Hennighausen, L. and Lubon, H. (1987) *Methods in Enzymology* **152,** 721-735.

Hideki, (1991)

Hooft van Huijsduijnen, R.A.M., Bollekens, J., Dorn, A., Benoist, C. and Mathis, D. (1987) *Nucleic Acids Res.* **15,** 7265-7282.

Hooft van Huijsduijnen, R., Li, X-Y., Black, D., Matthes, H., Benoist, C. and Mathis, D. (1990) *EMBO J.* **9,** 3119-3127.

Hope, I.A. and Struhl, K. (1985) *Cell* **43,** 177-188.

Horton, W., Miyashita, T., Kohno, K., Hassell, J.R. and Yamada, Y. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 8864-8868.

Hostikka, S.L. and Tryggvason, K. (1987) *FEBS Lett.* **224,** 297-305.

Hostikka, S.L. and Tryggvason, K. (1988) *J. Biol. Chem.* **263,** 19488-19493.

Hostikka, S.L., Eddy, R.L., Byers, M.G., Hoyhtya, M., Shows, T.B. and Tryggvason, K. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 1606-1610.

Howard, B.H, Adams, S.L., Sobel, M.E., Pastan, I. and de Crombrugghe, B. (1978) *J. Biol. Chem.* **253,** 5869-5874.

Huebner, K., Cannizzaro, L.A., Jabs, E.W., Kivirikko, S., Manzone, H., Pihlajaniemi, T. and Myers, J.C. (1992) *Genomics* **14,** 220-224.

Huerre, C., Junien, C., Weil, D., Chu, M., Morabito, M., van Cong, N., Myers, J.C., Foubert, C., Gross, M., Prockop, D.J., Boue, A., Kaplan, J., de la Chapelle, A. and Ramirez, F. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 6627-6630.

Hunter, T. and Karin, M. (1992) *Cell* **70,** 375-387.

Ignotz, R.A. and Massague, J. (1986) *J. Biol. Chem.* **261,** 4337-4345.

Imagawa, M., Chiu, R. and Karin, M. (1987) *Cell* **51,** 251-260.

Janson, L., Bark, C. and Pettersson, U. (1987) *Nucleic Acids Res.* **13,** 4997-5016.

Jeang, K-T. and Khoury, G. (1988) *Bioessays* **8,** 104-107.

Johnson, P.F., Landschulz, W.H., Graves, B.J. and Mc Knight, S.L. (1987) *Genes and Dev.* **1,** 133-146.

Johnson, P.F. and Mc Knight, S.L. (1989) *Annu. Rev. Biochem.* **58,** 799-839.

Jones, K.A. and Tjian, R. (1985) *Nature* **317,** 179-182.

Jones, K.A., Yamamoto, K.R. and Tjian, R. (1985) *Cell* **42,** 559-572.

Jones, K.A., Kadonaga, J.T, Rosenfeld, P.J., Kelly, T.J. and Tjian, R. (1987) *Cell* **48,** 79-89.

Jones, N.C., Rigby, P.W.J. and Ziff, E.B. (1988) *Genes and Dev.* **2,** 267-281.

Junien, C., Weil, D., Myers, J.C., van Cong, N., Chu, M., Foubert, C., Gross, M., Prockop, D.J., Kaplan, J., and Ramirez, F. (1982) *Am. J. Hum. Genet.* **34,** 381-387.

Kadonaga, J.T. and Tjian, R. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 5889-5893.

Kadonaga, J.T., Jones, K.A. and Tjian, R. (1986) *Trends Biochem. Sci.* **11,** 20-23.

Kadonaga, J.T., Carner, K.R., Masiarz, F.R. and Tjian, R. (1987) *Cell* **51,** 1079-1090.

Kadonaga, J.T., Courey, A.J., Ladika, J. and Tjian, R. (1988) *Science* **242,** 1566-1570.

Kageyama, R., Merlino, G.T. and Pastan, I. (1989) *J. Biol. Chem.* **264,** 15508-15514.

Kähäri, V-M., Chen, Y.Q., Su, M.W., Ramirez, F. and Uitto, J. (1990) *J. Clin. Invest.* **86,** 1489-1495.

Karsenty, G., Golumbek, P. and de Crombrugghe, B. (1988) *J. Biol. Chem.* **263,** 13909-13915.

Karsenty, G., and de Crombrugghe, B. (1990) *J. Biol. Chem.* **265,** 9934-9942.

Kennedy, G.C., Andersen, B. and Nilson, J.H. (1990) *J. Biol. Chem.* **265,** 6279-6285.

Khillan, (1986)

Khoury, G. and Gruss, P. (1983) *Cell* **33,** 313-314.

Kim, C.H., Heath, C., Bertuch, A. and Hansen, U. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 6025-6029.

Kim, C.G. and Sheffery, M. (1990) *J. Biol. Chem.* **265,** 13362-13369.

Kim, C.G., Swendeman, S.L., Barnhart, K.M. and Sheffery, M. (1990) *Mol. Cell. Biol.* **10,** 5958-5966.

Kim, Y.K. and Lee, A.S. (1992) *J. Biol. Chem.* **267,** 2723-2727.

Kimura, T., Mattei, M-G., Stevens, J.W., Goldring, M.B., Ninomiya, Y. and Olsen, B.R. (1989a) *Eur. J. Biochem.* **179,** 71-78.

Kimura, T., Cheah, K.S.E., Chan, S.D.H., Lui, V.C.H., Mattei, M-G., van der Rest, M., Ono, K., Solomon, E., Ninomiya, Y. and Olsen, B.R. (1989b) *J. Biol. Chem.* **264,** 13910-13916.

Klausner, R.D. and Harford, J.B. (1989) *Science* **246,** 870-872.

Knight, G.B., Gudas, J.M. and Pardee, A.B. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 8350-8354.

Kohno, K., Sullivan, M. and Yamada, Y. (1985) *J. Biol. Chem.* **260,** 4441-4447.

Koller, E., Hayman, A.R. and Trueb, B. (1991) *Nucleic Acids Res.* **19,** 485-491.

Koren, H.S., Handwerger, B.S. and Wunderlich, J.R. (1975) *J. Immun.* **114,** 894-897.

Kuhn, K. (1981) *Immunochemistry of the Extracellular Matrix,* Furthmayr, E., ed, **vol. I,** pp 1-29, CRC Press, Inc., Boca Raton, FL.

Laemmli, U.K. (1970) *Nature* **227,** 680-685.

Lamb, P. and Mc Knight, S.L. (1991) *Trends Biochem. Sci.* **16,** 417-422.

Landschulz, W.H., Johnson, P.F. and Mc Knight, S.L. (1988) *Science* **240,** 1759-1764.

Landschulz, W.H., Johnson, P.F. and Mc Knight, S.L. (1989) *Science* **243,** 1681-1688.

Law, M.L., Tung, L., Morse, H.G., Berger, R., Jones, C., Cheah, K.S.E. and Solomon, E. (1986) *Ann. Hum. Genet.* **50,** 131-137.

Lee, K.A.W. and Green, M.R. (1990) *Methods in Enzymology* **181,**20-30.

Lenardo, M.J. and Balitmore, D. (1989) *Cell* **58,** 227-229.

Levine, M. and Manely, J.L. (1989) *Cell* **59,** 405-408.

Li, X-Y., Hooft van Huijisduijnen, R., Mantovani, R., Benoist, C. and Mathis, D. (1992) *J. Biol. Chem.* **267,** 8984-8990.

Li, K., Tamai, K., Tan, E.M.L. and Uitto, J. (1993) *J. Biol. Chem.* **268,** 8825-8834.

Liau, G., Yamada, Y. and de Crombrugghe, B. (1985) *J. Biol. Chem.* **260,** 531-536.

Liau, G., Szapary, D., Setoyama, C. and de Crombrugghe, B. (1986) *J. Biol. Chem.* **261,** 11362-11368.

Lichtler, A., Stover, M.L., Angilly, J., Kream, B. and Rowe, D.W. (1989) *J. Biol. Chem.* **264,** 3072-3077.

Lieberman, H.B, Lin, P-F., Yeh, D-B. and Ruddle, F.H. (1988) *Mol. Cell. Biol.* **8,** 5280-5291.

Lim, L.C., Swendeman, S.L. and Sheffery, M. (1992) *Mol. Cell. Biol.* **12,** 828-835.

Linsenmayer, T.F. (1981) *Cell Biology of Extracellular matrix*, Hay, E.C. ed, pp 5-37, Plenum Publishing Corp., New York.

Liska, D.J., Slack, J.L. and Bornstein, P. (1990) *Cell Regulation* **1,** 487-498.

Liska, D.J., Robinson, V.R. and Bornstein, P. (1992) *Gene Expression* **2,** 379-389.

Liu, J., Park, E.A., Gurney, A.L., Roesler, W.J. and Hanson, R.W. (1991) *J. Biol. Chem.* **266,** 19095-19102.

Liu, Y., Yang, N. and Teng, C.T. (1993) *Mol. Cell. Biol.* **13,** 1836-1846.

Lipson, K.E., Chen, S-T., Koniecki, J., Ku, D-H. and Baserga, R. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 6848-6852.

Lozano, G., Ninomiya, Y., Thompson, H. and Olsen, B.R. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4050-4054.

Lu Valle, P., Ninomiya, Y., Rosenblum, N.D. and Olsen, B.R. (1988) *J. Biol. Chem.* **263,** 18378-18385.

Lubon, H. and Hennighausen (1987) *Nucleic Acids Res.* **15,** 2103-2121.

Lum, L.S.Y., Sultzman, L.A., Kaufman, R.J., Linzer, D.I.H. and Wu, B.J. (1990) *Mol. Cell. Biol.* **10,** 6709-6717.

Lum, L.S.Y., Hsu, S., Vaewhongs, M. and Wu, B.J. (1992) *Mol. Cell. Biol.* **12,** 2599-2605.

Maatta, A., Bornstein, P. and Penttinen, R.P.K. (1991) *FEBS Lett.* **279,** 9-13.

Mahoney, C.W., Shuman, J., Mc Knight, S.L., Chen, H-C. and Huang, K-P. (1992) *J. Biol. Chem.* **267,** 19396-19403.

Maity, S.N., Golumbek, P.T., Karsenty, G. and de Crombrugghe, B. (1988) *Science* **241,** 582-585.

Maity, S.N. and de Crombrugghe, B. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 5378-5382.

Maity, S.N. and de Crombrugghe, B. (1992) *J. Biol. Chem.* **267,** 8286-8292.

Maity, S.N., Sinha, S., Ruteshouser, E.C. and de Crombrugghe, B. (1992) *J. Biol. Chem.* **267,** 16574-16580.

Majmudar, G., Schalk, E., Bateman, J. and Peterkofsky, B. (1988) *J. Biol. Chem.* **263,** 5555-5559.

Maniatis, T., Goodbourn, S. and Fischer, J.A. (1987) *Science* **236,** 1237-1245.

Mann, K., Jander, R., Korsching, E., Kuhn, K. and Rauterberg, J. (1990) *FEBS Lett.* **273,** 168-172.

Mariyama, M., Zheng, K., Yang-Feng, T.L. and Reeders, S.T. (1992a) *Genomics* **13,** 809-813.

Mariyama, M., Kalluri, R., Hudson, B.G. and Reeders, S.T. (1992b) *J. Biol. Chem.* **267,** 1253-1258.

Markowitz, A. (1972) *Biochim. Biophys. Acta.* **281,** 522-534.

Martin, R.G. and Ames, B.N. (1961) *J. Biol. Chem.* **236,** 1372-1379.

Martin, J.D., King, D.M., Slauch, J.M. and Frisque, R.J. (1985) *J. Virol.* **53,** 306-311.

Matsuno, K., Suzuki, T., Takiya, S. and Suzuki, Y. (1989) *J. Biol. Chem.* **264,** 4599-4604.

Mauro, V.P., Nguyen, T., Katinakis, P. and Verma, D.P.S. (1985) *Nucleic Acids Res.* **13,** 239-249.

Mayne, R. and Burgeson, R. eds (1987) *Structure and fucnction of the collagen types,* Academic Press, New York.

Mc Keon, C., Ohkubo, H., Pastan, I. and de Crombrugghe, B. (1982) *Cell* **29,** 203-210.

Mc Keon, C., Pastan, I. and de Crombrugghe, B. (1984) *Nucleic Acids Res.* **12,** 3481-3502.

Mc Knight, S.L. and Kingsbury, R. (1982) *Science* **217,** 316-324.

Mermond, N., Williams, T.J. and Tjian, R. (1988) *Nature* **332,** 557-561.

Mermond, N., O'Neil, E.A., Kelly, T.J. and Tjian, R. (1989) *Cell* **58,** 741-753.

Milarski, K.L. and Morimoto, R.I. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 9517-9521.

Mitchell, P.J., Wang, C. and Tjian, R. (1987) *Cell* **50,** 847-861.

Mitchell, P.J. and Tjian, R. (1989) *Science* **245,** 371-378.

Morgan, W.D., Williams, G.T., Morimoto, R.I., Greene, J., Kingston, R.E., and Tjian, R. (1987) *Mol. Cell. Biol.* **7,** 1129-1137.

Morris, N.P. and Bachinger, H.P. (1987) *J. Biol. Chem.* **262,** 11345-11350.

Morrison, K.E., Mariyama, M., Yang-Feng, T.L. and Reeders, S.T. (1991) *Am. J. Hum. Genet.* **49,** 545-54.

Mudryj, M. and de Crombrugghe, B. (1988) *Nucleic Acids Res.* **16,** 7513-7527.

Mueller, C.R., Maire, P., Schibler, U. (1990) *Cell* **61,** 279-291.

Müller, H-P., Sogo, J.M. and Schaffner, W. (1989) *Cell* 767-777.

Muragaki, Y., Kimura, T., Ninomiya, Y. and Olsen, B.R. (1990) *Eur. J. Biochem.* **192,** 703-708.

Muragaki, Y., Jacenko, O., Apte, S. Mattei, M-G., Ninomiya, Y. and Olsen, B.R. (1991a) *J. Biol. Chem.* **266,** 7721-7727.

Muragaki, Y., Mattei, M-G., Yamaguchi, N., Olsen, B.R. and Ninomiya, Y. (1991b) *Eur. J. Biochem.* **266,** 7721-7727.

Myers, J.C., Loidl, H.R., Seyer, J.M. and Dion, A.S. (1985) *J. Biol. Chem.* **260,** 11216-11222.

Myers, J.C., Kivirikko, S., Gordon, M.K., Dion, A.S., Pihlajaniemi, T. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 10144-10148.

Nagata, K., Guggenheimer, R.A. and Hurwitz, J. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 6177-6181.

Namba, M., Nishitani, K. and Kimoto, T. (1980) *Gann* **71,** 300-307.

Narayanan, A.S., Page, R.C. and Swanson, J. (1989) *Biochem J.* **260,** 463-469.

Nehls, M.C., Rippe, R.A., Veloz, L. and Brenner, D.A. (1991) *Mol. Cell. Biol.* **11,** 4065-4073.

Nehls, M.C., Grapilon, M.L. and Brenner, D.A. (1992) *DNA Cell Biol.* **11,** 443-452.

Niederreither, K., D'Souza, R.N. and de Crombrugghe B. (1992) *J. Cell Biol.* **119,** 1361-1370.

Ninomiya, Y. Gordon, M., van der Rest, M., Schmid, T., Linsenmayer, T. and Olsen, B.R. (1986) *J. Biol. Chem.* **261,** 5041-5050.

Ninomiya, Y., Castagnola, P., Gerecke, D., Gordon, M.K., Jacenko, O., Lu Valle, P., Mc Carthy, M., Muragaski, Y., Nishimura, I., Oh, S., Rosenblum, N., Sato, N., Sugrue, S., Taylor, R., Vasios, G., Yamaguchi, N. and Olsen, B.R. (1990) *Extracellular Matrix Genes*, Sandell, L.J. and Boyd, C.D., eds, pp 79-114, Academic press, San Diego.

Nishimura, I., Muragaki, Y. and Olsen, B.R. (1989) *J. Biol. Chem.* **264,** 20033-20041.

Noro, A., Kimata, K., Oike, Y., Shinomura, T., Maeda, N., Yano, S., Takahashi, N. and Suzuki, S. (1983) *J. Biol. Chem.* **258,** 9323-9331.

Nunez, A.M., Kohno, K., Martin, G.R. and Yamada, Y. (1986) *Gene* **44,** 11-16.

Ohkuma, Y., Sumimoto, H., Hoffmann, A., Shimasaki, S., Horikoshi, M. and Roeder, R.G. (1991) *Nature* **354,** 398-401.

Oh, S.P., Taylor, R.W., Gerecke, D.R., Rochelle, J.M., Seldin, M.F. and OLsen, B.R. (1992) *Genomics* **14,** 225-231.

Oikarinen, J., Hatamochi, A. and de Crombrugghe, B. (1987) *J. Biol. Chem.* **262,** 11064-11070.

Oono, T., Specks, U., Eckes, B., Majewski, S., Hunzelmann, N., Timpl, R. and Krieg, T. (1993) *J. Invest. Dermatol.* **100,** 329-334.

Ozer, J., Faber, M., Chalkley, R. and Sealy, L. (1990) *J. Biol. Chem.* **265,** 22143-22152.

Palmiter, R.D., Behringer, R.R., Quaife, C.J., Maxwell, F., Maxwell, I.H. and Brinster, R.L. (1987) *Cell* **50,** 435-443.

Pan, T-C., Zhang, R-Z., Mattei, M-G., Timpl, R. and Chu, M-L. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 6565-6569.

Pang, J.H. and Chen, K.Y. (1993) *J. Biol. Chem.* **268,** 2909-2916.

Parente, M.G., Chung, L.C., Ryynanen, J., Woodley, D.T., Wynn, K.C., Bauer, E.A., Mattei, M-G., Chu, M-L. and Uitto, J. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 6931-6935.

Park, E.A., Roesler, W.J., Liu, J., Klemm, D.J., Gurney, A.L., Thatcher, J.D., Shuman, J., Friedman, A. and Hanson, R.W. (1990) *Mol. Cell. Biol.* **10,** 6264-6272.

Parker, M.I., Judge, K. and Gevers, W. (1982) *Nucleic Acids Res.* **10,** 5879-5891.

Parker, C.S. and Topol, J. (1984) *Cell* **37,** 273-283.

Parker, M.I., de Haan, J.B. and Gevers, W. (1986) *J. Biol. Chem.* **261,** 2786-2790.

Parker, M.I., Smith, A.A. and Gevers, W. (1989) *J. Biol. Chem.* **264,** 7147-7152.

Parker, M.I., Smith, A.A., Mundell, K., Collins, M., Boast, S. and Ramirez, F. (1992) *Nucleic Acids Res.* **20,** 5825-5830.

Patient, R.K. and Allan, J. (1991) *Curr. Opinion in Cell Biol.* **1,** 454-459.

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) *J. Cell Biol.* **116,** 227-236.

Perälä, M., Hänninen, M., Hästbacka, J., Elima, K. and Vuorio, E. (1993) *FEBS Lett.* **319,** 177-180.

Pfeifer, K, Arcangioli, B. and Guarente, L. (1987a) *Cell* **49,** 9-18.

Pfeifer, K, Prezant, T. and Guarente, L. (1987b) *Cell* **49,** 19-27.

Pihlajaniemi, T., Pohjolainen, E-R. and Myers, J.C. (1990a) *J. Biol. Chem.* **265,** 13758-13766.

Pihlajaniemi, T. and Tamminen, M. (1990b) *J. Biol. Chem.* **265,** 16922-16928.

Pogulis, R.J. and Freytag, S.O. (1993) *J. Biol. Chem.* **268,** 2493-2499.

Poli, V., Mancini, F.P. and Cortese, R. (1990) *Cell* **63,** 643-653.

Poschl, E., Pollner, R. and Kuhn, K. (1988) *EMBO J.* **7,** 2687-2695.

Prockop, D.J. and Kivirikko, K.I (1979) *New Engl. J. Med.* **301,** 13-23.

Ptashne, M. (1988) *Nature* **335,** 683-689.

Ptashne, M. and Gann, A.A.F. (1990) *Nature* **346,** 329-331.

Radloff, R, Bauer, W. and Vinograd, J. (1967) *Proc. Natl. Acad. Sci. USA* **57,** 1514-1521.

Raghow, R. and Thompson, J.P. (1989) *Mol. Cell. Biochem.* **86,** 5-18.

Raibaud, O., Gutierrez, C. and Schwartz, M. (1985) *J. Bacteriol.* **161,** 1201-1208.

Ramirez, F., Boast, S., D'Alessio, M., Lee, B., Prince, J., Su, M-W., Vissing, H and Yoshioka, H. (1990) *Ann. N.Y. Acad. Sci.* **580,** 74-80.

Ravazzolo, R., Karsenty, G. and de Crombrugghe, B. (1991) *J. Biol. Chem.* **266,** 7382-7387.

Raymondjean, M., Cereghini, S. and Yaniv, M. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 757-761.

Reichenberger, E., Beier, F., Lu Valle, P., Olsen, B.R., von der Mark, K. and Bertling, W.M. (1992) *FEBS Lett.* **311,** 305-310.

Renkawitz, R. (1990) *Trends Genetics* **6,** 192-197.

Resnick, D., Freedman, N.J., Xu, S. and Krieger, M. (1993) *J. Biol. Chem.* **268,** 3538-3545.

Retief, E., Parker, M.I. and Retief, A.E. (1985) *Hum. Genet.* **69,** 304-308.

Rippe, R.A., Lorenzen, S.I., Brenner, D.A. and Breindl, M. (1989) *Mol. Cell. Biol.* **9,** 2224-2227.

Ristiniemi, J. and Oikarinen, J. (1989) *J. Biol. Chem.* **264,** 2164-2174.

Ritzenthaler, J.D., Goldstein, R.H., Fine, A., Lichtler, A., Rowe, D.W. and Smith, B.D. (1991) *Biochem. J.* **280,** 157-162.

Ritzenthaler, J.D., Goldstein, R.H., Fine, A. and Smith, B.D. (1993) *J. Biol. Chem.* **268,** 13625-13631.

Rosenfeld, P.J. and Kelly, T.J. (1986) *J. Biol. Chem.* **261,** 1398-1408.

Rossi, P., and de Crombrugghe, B. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 5590-5594.

Rossi, P., Karsenty, G., Roberts, A.B., Roche, N.S., Sporn, M.B. and de Crombrugghe, B. (1988) *Cell* **52,** 405-414.

Rossouw, C.M., Vergeer, W.P., du Plooy, S.J., Bernard, M.P., Ramirez, F. and de Wet, W.J. (1987) *J. Biol. Chem.* **262,** 15151-15157.

Rupp, R.A.W., Kruse, U., Multhaup, G., Gobel, U., Beyreuther, K. and Sippel, A.E. (1990) *Nucleic Acids Res.* **18,** 2607-2616.

Ruteshouser, E.C. and de Crombrugghe, B. (1989) *J. Biol. Chem.* **264,** 13740-13744.

Ruteshouser, E.C. and de Crombrugghe, B. (1992) *J. Biol. Chem.* **267,** 14398-14404.

Ryynanen, J., Sollberg, S., Olsen, D.R. and Uitto, J. (1991) *Biochem. Biophys. Res Commun.* **180,** 673-680.

Saitta, B., Stokes, D.G., Vissing, H., Timpl, R. and Chu, M-L. (1990) *J. Biol. Chem.* **265,** 6473-6480.

Saitta, B., Wang, Y.M., Renkart, L., Zhang, R.Z., Pan, T.C., Timpl, R. and Chu, M.L. (1991) *Genomics* **11,** 145-153.

Saitta, B., Timpl, R. and Chu, M-L. (1992) *J. Biol. Chem.* **267,** 6188-6196.

Sakura, H., Maekawa, T., Imamoto, F., Yasuda, K. and Ishii, S. (1988) *Gene* **73,** 499-507.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning, A Laboratory Manual*, **2nd edition**, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Sandmeyer, S., Gallis, B. and Bornstein, P. (1981) *J. Biol. Chem.* **256,** 5022-5028.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. SCi. USA* **74,**5463-5467.

Sangiorgi, F.O., Benson-Chanda, V., de Wet, W.J., Sobel, M.E., Tsipouras, P. and Ramirez, F. (1985) *Nucleic Acids Res.* **13,** 2207-2225.

Santoro, C., Mermod, N., Andrews, P.C. and Tjian, R. (1988) *Nature* **334,** 218-224.

Savagner, P., Miyashita, T. and Yamada, Y. (1990) *J. Biol. Chem.* **265,** 6669-6674.

Sawadogo, M. and Roeder, R.G. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4394-4398.

Saxe, S.A., Lukens, L.N. and Pawlowski, P.J. (1985) *J. Biol. Chem.* **260,** 3812-3819.

Schittny, J.C. and Yurchenco, P.D. (1989) *Curr. Opinion in Cell Biol.* **1,** 983-988.

Schmid, T.M. and Linsenmayer, T.F. (1983) *J. Biol. Chem.* **258,** 9504-9509.

Schmidt, A., Setoyama, C. and de Crombrugghe, B. (1985) *Nature* **314,** 286-289.

Schmidt, A., Rossi, P. and de Crombrugghe, B. (1986) *Mol. Cell. Biol.* **6,** 347-354.

Schnieke, A., Harbers, K. and Jaenisch, R. (1983) *Nature* **304,** 315-320.

Schreier, T., Friis, R.R., Winterhalter, K.H. and Trüeb, B. (1988) *Biochem. J.* **253,** 381-386.

Sealey, L. and Chalkley, R. (1987) *Mol. Cell. Biol.* **7,** 787-798.

Serfling, E., Jasin, M. and Schaffner, W. (1985) *Trends Genetics* **1,** 224-230.

Sharma, A., Bos, T.J., Pekkala-Flagan, A., Vogt, P.K. and Lee, A.S. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 491-495.

Shaw, L.M. and Olsen, B.R. (1991) *Trends Biochem. Sci.* **16,** 191-194.

Sherwood, A.L., Bottenus, R.E., Martzen, M.R. and Bornstein, P. (1990) *Gene* **89,**239-244.

Siebenlist, U. and Gilbert, W. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 122-126.

Siegel, L.M. and Monty, K.J. (1966) *Biocheim. Biophys. Acta* **112,** 346-362.

Sierra, F., Stein, G. and Stein, J. (1983) *Nucleic Acids Res.* **11,** 7069-7086.

Simkevich, C.P., Thompson, J.P., Poppleton, H. and Raghow, R. (1992) *Biochem. J.* **286,** 179-185.

Singh, H., Sen, R., Baltimore, D. and Sharp, P.A. (1986) *Nature* **319,** 154-158.

Slack, J.L., Liska, D.J. and Bornstein, P. (1991) *Mol. Cell. Biol.* **11,** 2066-2074.

Slack, J.L., Liska, D.J. and Bornstein, P. (1993) *American J. Med. Genetics* **45,** 140-151.

Smith, B.D. and Marsilo, E. (1988) *Biochem. J.* **253,** 269-273.

Smith, I., Cromie, R. and Stainsby, K. (1988), *Anal. Biochem.* **169,** 370-371.

Smith, A.A. (1989) *Ph.D. Thesis*, University of Cape Town.

Soininen, R., Huotari, M., Hostikka, S.L., Prockop, D.J. and Tryggvason, K. (1988) *J. Biol. Chem.* **264,** 17217-17220.

Soininen, R., Huotari, M., Ganguly, A., Prockop, D.J. and Tryggvason, K. (1989) *J. Biol. Chem.* **264,** 13565-13571.

Stokes, D.G., Saitta, B., Timpl, R. and Chu, M-L. (1991) *J. Biol. Chem.* **266,** 8626-8633.

Strickland, S., Smith, K.K. and Marotti, K.R. (1980) *Cell* **24,** 347-355.

Struhl, K. (1989) *Trends Biochem. Sci.* **14,** 137-140.

Sturn, R.A., Das, G. and Herr, W. (1988) *Genes and Dev.* **2,** 1582-1599.

Sumimoto, H., Ohkuma, Y., Sinn, E., Kato, H., Shimasaki, S., Horikoshi, M. and Roeder, R.G. (1991) *Nature* **354,** 401-404.

Sundarraj, N. and Church, R.L. (1978) *FEBS Lett.* **85,** 47-51.

Tabor, S. and Richardson, C.C. (1989) *Proc. Natl. Acad. Sci. USA* **86,**4076-4080.

Takahara, K., Sato, Y., Okazawa, K., Okamoto, N., Noda, A., Yaoi, Y. and Kato, I. (1991) *J. Biol. Chem.* **266,** 13124-13129.

Taniguchi, T., O'Neill, M. and de Crombrugghe, B. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 5090-5094.

Tasset, D., Tora, L., Fromental, C., Scheer, E. and Chambon, P. (1990) *Cell* **62,** 1177-1187.

Taylor, S.M. and Jones, P.A. (1979) *Cell* **17,** 813.

Thomas, J.T., Cresswell, C.J., Rash, B., Nicolai, H., Jones, T., Solomon, E., Grant, M.E. and Boot-Handford, R.P. (1991) *Biochem. J.* **280,** 617-623.

Thompson, J.P., Simkevich, C.P., Holness, M.A., Kang, A.H. and Raghow, R. (1991) *J. Biol. Chem.* **266,** 2549-2556.

Timpl, R. and Dzidek, M. (1986) *Int. Rev. Exp. Pathol.* **29,** 1-112.

Towbin, H., Staehelin, T. and Gordon, J. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 4350-4354.

Treacy, M.N., He, X. and Rosenfeld, M.G. (1991) *Nature* **350,** 577-584.

Treacy, M.N., Neilson, L.I., Turner, E.E., He, X. and Rosenfeld, M.G. (1992) *Cell* **68,** 491-505.

Truter, S., Di Liberto, M., Inagaki, Y. and Ramirez, F. (1992a) *J. Biol. Chem.* **267,** 25389-25395.

Truter, S., Andrikopoulos, K., Di Liberto, M., Womack, L. and Ramirez, F. (1993) *Connect. Tissue Res.* **29,** 51-59.

Tsai-Morris, C-H., Cao, X. and Sukhatme, V.P. (1988) *Nucleic Acids Res.* **16,** 8835-8846.

Turner, N., Mason, P.J., Brown, R., Fox, M., Povey, S., Rees, A. and Pusey, C.D. (1992) *J. Clin. Invest.* **89,** 592-601.

Umek, R.M., Friedman, A.D. and Mc Knight, S.L. (1991) *Science* **251,** 288-292.

Unemori, E.N., Amento, E.P., Bauer, E.A. and Horuk, R. (1993) *J. Biol. Chem.* **268,** 1338-1342.

Vallejo, M., Ron, D., Miller, C.P. and Habener, J.F. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 4679-4683.

van der Rest, M. and Mayne, R. (1987) *Structure and Function of Collagen Types*, Mayne, R. and Burgeson, R.E. eds, pp 195-221, Academic Press, Orlando.

van der Rest, M. and Mayne, R. (1988) *J. Biol. Chem.* **263,** 1615-1618.

van der Rest, M. and Garrone, R. (1991) *FASEB J.* **5,** 2814-2823.

van Wijnen, A.J., Massung, R.F., Stein, J.L. and Stein, G.S. (1988) *Biochemistry* **27,** 6534-6541.

Vaughan, L., Mendler, M., Huber, S., Bruckner, P., Winterhalter, K.H., Irwin, M.I. and Mayne, R. (1988) *J. Cell Biol.* **106,** 991-997.

von der Mark, K. (1980) *Curr. Top. Dev. Biol.* **14,** 199-225.

Vuorio, E. and de Crombrugghe, B. (1990) *Annu. Rev. Biochem.* **59,** 838-872.

Vuorio, T., Maity, S.N. and de Crombrugghe, B. (1990) *J. Biol. Chem.* **265,** 22480-22486.

Vuust, J., Sobel, M.E. and Martin, G.R. (1985) *Eur. J. Biochem.* **151,** 449-453.

Wang, L., Balakir, R. and Horton, W.E. (1991) *J. Biol. Chem.* **266,** 19878-19881.

Wegner, M., Cao, Z. and Rosenfeld, M.G. (1992) *Science* **256,** 370-373.

Weil, D., Bernard, M., Gargano, S. and Ramirez, F. (1986) *Nucleic Acids Res.* **15,** 181-199.

Weisbrod, S. (1982) *Nature* **297,** 289-295.

Wells, D.E. (1986) *Nucleic Acids Res.* **14 Suppl,** r119-r149.

Westerhausen, A., Constantinou, C.D., Pack, M., Peng, M.Z., Hanning, C., Olsen, A.S. and Prockop, D.J. (1991) *Matrix* **11,** 375-379.

Williams, T. and Tjian, R. (1991) *Genes and Dev.* **5,** 670-682.

Woodbury, D., Beson-Chanda, V. and Ramirez, F. (1989) *J. Biol. Chem.* **264,** 2735-2738.

Wu, B.J., Williams, G.T. and Morimoto, R.I. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 2203-2207.

Wu, C., Wilson, S., Walker, B., Dawid, I., Paisley, T., Zimarino, V. and Ueda, H. (1987) *Science* **238,** 1247-1253.

Wuarin, J., Mueller, C. and Schibler, U. (1990) *J. Mol. Biol.* **214,** 865-874.

Xiao, J.H., Davidson, I., Ferrandon, D., Rosales, R., Vigneron, M., Macchi, M., Ruffenach, F. and Chambon, P. (1987) *EMBO J.* **6,** 3005-3013.

Yamada, Y., Mudryj, M., and de Crombrugghe, B. (1983) *J. Biol. Chem.* **258,** 14914-14919.

Yamada, Y., Miyashita, T., Savagner, P., Horton, W., Brown, K.S., Abramczuk, J., Hou-Xiang, X., Kohno, K., Bolander, M. and Bruggeman, L. (1990) *Ann. N.Y. Acad. Sci.* **580,** 81-87.

Yamaguchi, N., Benya, P.D., van der Rest, M. and Ninomiya, Y. (1989) *J. Biol. Chem.* **264,** 16022-16029.

Yamaguchi, N., Mayne, R. and Ninomiya, Y. (1991) *J. Biol. Chem.* **266,** 4508-4513.

Yoshioka, H., Zhang, H., Ramirez, F., Mattel, M-G., Moradi-Ameli, M., van der Rest, M. and Gordon, M.K. (1992) *Genomics* **13,** 884-886.

Yuh, C-H. and Ting, L-P. (1991) *Mol. Cell. Biol.* **11,** 5044-5052.

Zabel, U. and Baeuerle, P.A. (1990) *Cell* **61,** 255-265.

Zhou, J., Hertz, J.M., Leinonen, A. and Tryggvason, K. (1992) *J. Biol. Chem.* **267,** 12475-12481.

Zorbas, H., Rein, T., Krause, A., Hoffman, K. and Winnacker, E-L. (1992) *J. Biol. Chem.* **267,** 8478.

APPENDIX A

CANDIDATE TRANSCRIPTION FACTORS AND THEIR RECOGNITION SEQUENCES

DNA sequences recognised by sequence-specific transcription factors in the proximal α2(I) procollagen promoter (-351 to -1) were identified by searching a database of transcription factors, TFD SITES (release 4.3), maintained by Dr D. Ghosh (1990) at the National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health. The following table contains a list of recognition sequences (SEQUENCE) found in this region of the promoter, the candidate transcription factors (TRANS FACTOR), the position of the 5' end of the sequence (POS), the name of the sequence (NAME) and the reference. The abbreviations CS and RS mean consensus sequence and reverse (inverted) sequence respectively.

POS	SEQUENCE	NAME	TRANS FACTOR	REFERENCE
-33	TATAAA	TATA BOX.2	TFIID	Bucher and Trifonov (1986)
-33	TATAAAT	TATA BOX (CS)	TFIID	Breathnach and Chabon (1981)
-33	TATAAA	His3-Tr-TATA	TFIID	Chen and Struhl (1988)
-33	TATAAA	Ad2MLP US.3	TFIID	Sawadogo and Roeder (1985)
-33	TATAAATAG	TFIID/TBP (H2B-1 gene)	TFIID/TBF	Barberis et. al. (1987)
-33	TATAAA	TFIID/TBF?-RS	TFIID/TBF	Wu, C. et. al. (1987)
-33	TATAAATA	B-FACTOR-HSP70	B-factor	Parker and Topol (1984)
-49	AGGGAAAC	Nodulin consensus A (RS)	Unknown	Mauro et. al. (1985)
-82	TGGTGG	HC3 (RS)	Unknown	Augereau and Chambon (1986)
-84	ATTGG	CCAAT Box (RS)	CBF	Maity et. al. (1988)
-104	CCCCCAGCCC	AP-2 (CS 4)	AP-2	Imagawa et. al. (1987)
-115	GGCTGCCC	AP-2 (CS)	AP-2	Mitchell et. al. (1987)
-122	CAGCTGT	GT-IIBa-SV (RS)	GT-IIBa	Xiao et. al. (1987)
-122	CAGCTGT	GT-IIBb-SV (RS)	GT-IIBb	Xiao et. al. (1987)
-122	CAGCTGTGG	GT-IIb/Pvu Box	AP-4	Jones et. al. (1988)
-122	CAGCTGTGG	AP-4 (RS)	AP-4	Jones et. al. (1988)
-123	CCAGCTGTGG	AP-4 (CS)	AP-4	Mermod et. al. (1988)
-128	TCCTCC	malT-malPp	Unknown	Raibaud et. al. (1985)
-128	TCCTCC	malT (CS, RS)	malT	Raibaud et. al. (1985)
-129	CTCCTCCC	JCV repeated sequence (RS)	Unknown	Martin et. al. (1985)
-136	CCGCAGGC	AP-2 (CS 3)	AP-2	Mitchell et. al. (1987)
-143	TGCGCCC	MRE (CS 2)	Unknown	Culotta and Hamer (1989)
-160	CCATTC	H2A conserved US	Unknown	Wells (1986)
-187	GCCAGCCC	AP-2 (CS)	AP-2	Mitchell et. al. (1987)
-201	AACCAAT	CAAT.1	CP1	Chodosh et. al. (1988a)
-229	AAAGTGT	CAP/CRP-lac (RS)	CAP/CRP	Taniguchi et. al. (1979)
-257	GAGTCA	GCN4-ILV1.1	GCN4	Arndt and Fink (1986)
-257	GAGTCA	GCRE (RS)	GCN4	Hope and Struhl (1985)
-257	GAGTCA	GCN4-HIS3.2	GCN4	Arndt and Fink (1986)
-257	GAGTCA	GCN4-HIS (RS)	GCN4	Arndt and Fink (1986)
-257	GAGTCA	GCN4-HIS3.5 (RS)	GCN4	Hope and Struhl (1985)
-257	GAGTCA	GCN4-HIS4.1 (RS)	GCN4	Arndt and Fink (1986)
-257	GAGTCA	GCN4-HIS4.2 (RS)	GCN4	Arndt and Fink (1986)
-258	CGAGTCAG	AP1-Histone H3.2 (RS)	AP-1	Sharma et. al. (1986)

Candidate Transcription Factors and their Recognition Sequences (cont)

POS	SEQUENCE	NAME	TRANS FACTOR	REFERENCE
-258	CGAGTCAG	TRE-Egr-1.1 (RS)	AP-1	Tsai-Morris et. al. (1988)
-269	TATGCA	EFII-RSV	EFII	Sealey and Chalkley (1987)
-278	GGGCGG	BGP1 (RS)	BGP1	Cereghini et. al. (1988)
-278	GGGCGG	Sp1 (CS 2)	Sp1	Briggs et. al. (1986)
-278	GGGCGG	Sp1-IE-3.1 (RS)	Sp1	Jones and Tjian (1985)
-278	GGGCGG	Sp1-IE-3.2	Sp1	Jones and Tjian (1985)
-278	GGGCGG	Sp1-IE-3.4 (RS)	Sp1	Jones and Tjian (1985)
-278	GGGCGG	Sp1-IE-3.5 (RS)	Sp1	Jones and Tjian (1985)
-278	GGGCGG	Sp1-IE-4/5.2 (RS)	Sp1	Jones and Tjian (1985)
-278	GGGCGG	LSF-SV40 (RS)	LSF	Kim et. al. (1987)
-278	GGGCGG	Sp1-SV40.4	Sp1	Gidoni et. al. (1984)
-278	GGGCGG	Sp1?-U2snR.1	(Sp1)	Janson et. al. (1987)
-278	GGGCGG	Sp1?-U2snR.3	(Sp1)	Janson et. al. (1987)
-278	GGGCGG	Sp1 (CS 3)	Sp1	Gidoni et. al. (1984)
-278	GGGCGGAG	JCV repeated sequence	Unknown	Martin et. al. (1985)
-278	GGGCGGAG	Sp1-SV40.2	Sp1	Gidoni et. al. (1985)
-278	GGGCGGAG	Sp1-SV40.3	Sp1	Gidoni et. al. (1984)
-278	GGGCGGAG	Early-Seq1 (RS)	Unknown	Fromm and Berg (1982)
-282	CGGAGGGCGG	AP-2 (CS 4, RS)	AP-2	Imagawa et. al. (1987)
-290	GGAGGA	malT (CS)	malT	Raibaud et. al. (1985)
-290	GGAGGA	malT-malPp (RS)	Unknown	Raibaud et. al. (1985)
-294	GGCGGG	Sp1-hsp70	Sp1	Greene et. al. (1987)
-294	GGCGGG	hsp70.2	Unknown	Wu, C. et. al. (1987)
-294	GGCGGG	Sp1-IE-3.3 (RS)	Sp1	Jones and Tjian (1985)
-294	GGCGGG	Sp1-IE-4/5 (RS)	Sp1	Jones and Tjian (1985)
-295	GGGCGG	Sp1 (CS 2)	Sp1	Briggs et. al. (1986)
-295	GGGCGG	BGP1 (RS)	BGP1	Cereghini et. al. (1988)
-295	GGGCGG	Sp1-IE-3.1 (RS)	Sp1	Jones and Tjian (1985)
-295	GGGCGG	Sp1-IE-3.2	Sp1	Jones and Tjian (1985)
-295	GGGCGG	Sp1-IE-3.4 (RS)	Sp1	Jones and Tjian (1985)
-295	GGGCGG	Sp1-IE-3.5 (RS)	Sp1	Jones and Tjian (1985)
-295	GGGCGG	Sp1-IE-4/5.2 (RS)	Sp1	Jones and Tjian (1985)
-295	GGGCGG	LSF-SV40 (RS)	LSF	Kim et. al. (1987)
-295	GGGCGG	Sp1-SV40.4	Sp1	Gidoni et. al. (1984)
-304	GGCGGG	Sp1-IE-4/5 (RS)	Sp1	Jones and Tjian (1985)
-305	GGGCGGGGGA	AP-2 (CS 4, RS)	AP-2	Imagawa et. al. (1987)
-305	GGGCGG	BGP1 (RS)	BGP1	Cereghini et. al. (1988)
-305	GGGCGG	Sp1-IE-3.1 (RS)	Sp1	Jones and Tjian (1985)
-305	GGGCGG	Sp1-IE-3.2	Sp1	Jones and Tjian (1985)
-305	GGGCGG	Sp1-IE-3.4 (RS)	Sp1	Jones and Tjian (1985)
-305	GGGCGG	Sp1-IE-3.5 (RS)	Sp1	Jones and Tjian (1985)
-305	GGGCGG	Sp1-IE-4/5.2 (RS)	Sp1	Jones and Tjian (1985)
-305	GGGCGG	Sp1 (CS 2)	Sp1	Briggs et. al. (1986)
-305	GGGCGG	LSF-SV40 (RS)	LSF	Kim et. al. (1987)
-305	GGGCGG	Sp1-SV40.4	Sp1	Gidoni et. al. (1984)
-305	GGGCGG	Sp1?-USsnR.1	(Sp1)	Janson et. al. (1987)
-305	GGGCGG	Sp1?-USsnR.2	(Sp1)	Jansom et. al. (1987)
-295	GGGCGG	Sp1?-U2snR.1	(Sp1)	Janson et. al. (1987)
-295	GGGCGG	Sp1?-U2snR.3	(Sp1)	Janson et. al. (1987)
-295	GGGCGGGA	Early-Seq1 (RS)	Unknown	Fromm and Berg (1982)
-295	GGGCGG	Sp1 (CS 3)	Sp1	Gidoni et. al. (1984).
-295	GGGCGGGA	JCV repeated sequence	Unknown	Martin et. al. (1985)
-304	GGCGGG	Sp1-hsp70	Sp1	Greene et. al. (1987)
-304	GGCGGG	hsp70.2	Unknown	Wu, C. et. al. (1987)
-304	GGCGGG	Sp1-IE-3.3 (RS)	Sp1	Jones and Tjian (1985)
-305	GGGCGGGG	Early-Seq1 (RS)	Unknown	Fromm and Berg (1982)
-305	GGGCGG	Sp1 (CS 3)	Sp1	Gidoni et. al. (1984)
-305	GGGCGGGG	(Sp1)-TK.1 (RS)	(Sp1)	Mc Knight and Kingsbury (1982)
-305	GGGCGGGG	JCV repeated sequence	Unknown	Martin et. al. (1985)
-310	GGAAAG	IE1.2 (RS)	Unknown	Ghazal et. al. (1984)
-311	TGGAAAG	SV40.6	Unknown	Pfeifer et. al. (1987a)
-311	TGGAAAG	SV40.16	Unknown	Pfeifer et. al. (1987a)
-311	TGGAAAG	SV40.13	Unknown	Pfeifer et. al. (1987a)
-314	ACTTGG	WAP US6	Unknown	Lubon and Henninghausen (1987)
-344	GCACTC	MT-I.1	Unknown	Andersen et. al. (1987)