MOLECULAR CLONING OF cDNA FOR A NUCLEAR MATRIX PROTEIN

KHWAJA SHAHID HAMEED

ISBN 0-612-01860-1

MOLECULAR CLONING OF cDNA FOR A NUCLEAR MATRIX PROTEIN

By

©Khwaja Shahid Hameed, B.Sc.(Honours), M.Sc.

A thesis submitted to the School of Graduate Studies

in partial fulfillment of the requirements

of the degree of

Doctor of Philosophy

Faculty of Medicine

Memorial University of Newfoundland

October, 1993

St.John's Newfoundland

In memory of my mother

and

sister

ABSTRACT

The eukaryotic nucleus is organized around a proteinaceous structural framework termed the nuclear matrix. The precise way in which nuclear proteins are ordered to form a nuclear scaffolding remains to be established. In order to study structural and functional aspects of nuclear matrix proteins, it is necessary to clone and characterize their genes. As a first step towards achieving this goal, an immunological screening of a λgt11 cDNA library with antiserum raised against a nuclear matrix preparation was performed. Screening of the λgt11 library resulted in the isolation of several immunoreactive positive clones. One of these cDNA clones, λ7E, was further characterized. Immunoblot analysis, using antisera prepared against 7E β gal-fusion protein, resulted in the detection of an immunoreactive nuclear protein with an apparent molecular weight of 52 kDa. Furthermore, cell fractionation combined with immunoblotting showed that, the 7Eb protein was a component of the HeLa cell nuclear matrix cofractionated with nuclei. Subsequent to screening of another cDNA library (λZap II) with a λ7E DNA probe, two overlapping cDNA clones λ8.1 and λ17.1 were isolated. The composite nucleotide sequence deduced from λ8.1 and λ17.1 cDNA inserts revealed that, the cDNA is capable of encoding a polypeptide of 422 amino acid residues with a molecular size of 49.8 kDa which is very close to the molecular weight of the protein identified by western blot analysis.

Based upon the nucleotide sequence data the 7Eb protein was found to be rich in basic amino acids (22%). A computer search for similarity to the predicted amino acid sequence of the 7Eb protein did not reveal any resemblance to previously cloned nuclear matrix proteins including the nuclear lamins.

The amino-terminal half of the 7Eb protein contains a stretch of about 80 amino acids which are very similar to the DNA binding HMG box domain of a number of proteins. The predicted amino acid of the 7Eb sequence also revealed a significant similarity to many of the cytoskeletal proteins. Although the function of the protein identified in this study is not yet known, the presence of a HMG box domain and sequence similarity to cytoskeletal proteins suggest that this protein is a novel DNA binding nuclear matrix protein.

Key words: Nuclear matrix; cDNA; HMG box.

ACKNOWLEDGMENTS

I am deeply grateful to Dr. H.Banfield Younghusband, the chairman of my supervisory committee for his supervision, help and guidance throughout this program. I am also very thankful to Dr. William Davidson for reading the drafts of this thesis and for his helpful suggestions during this research project. I wish to thank Dr. H.C.Smith for providing me with antibody. I wish to thank Dr. Peter Howard for helping me with the database search for my sequences. I am also very thankful to Miss Mary O'Driscoll for invaluable assistance in putting photographs together in the thesis. I am grateful to the School of Graduate Studies and the Faculty of Medicine for financial support.

I am deeply grateful to my brothers, sisters, and especially to my mother without their love, support, and understanding, this work would not have been possible.

Finally, I thank my wife, Shagufta, my children, Hanaa and Sulaiman for their love and patience during this work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

AMV	avain myeloblastosis virus
BSA	bovine serum albumin
bp	base pair
cDNA	complementary DNA
CPM	counts per min
DNase I	deoxyribonuclease I
DTT	dithiothreitol
DEPC	diethyl pyrocarbonate
EDTA	ethylenediaminetetra-acetic acid
EM	electron microscope
gal	galactosidase
HMG	high mobility group
hnRNA	heterogenous RNA
hnRNP	heterogenous ribonucleoprotein
HRP	horseradish peroxidase
HSP70	70K heat shock protein
IAA	iodo acetamide
IPTG	isopropyl-b-D-thiogalactopyranoside
kp	kilobase
kbp	kilobase pairs

LB	Luria-Bertani
LIS	lithium diiodosalicylate
MAR	matrix associated region
MEM	minimal essential medium
MOPS	morpholinopropane sulfonic acid
M_r	relative molecular mass
NaTT	sodium tetrathionate
NEM	N-ethyl maleimide
NP-40	Nonidet P-40
OD	optical density
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PFU	plaque forming unit
PMSF	phenyl methyl sulphonyl fluoride
Rnase	ribonuclease
RNP	ribonucleoprotein
snRNP	small ribonucleoprotein
SDS	sodium dodecy sulphate
SSC	saline sodium citrate
SSPE	saline sodium phosphate-EDTA

SV40	simian virus 40
T antigen	tumor antigen
TEP	Tris-EDTA PMSF
TBS	Tris-buffered saline
TBST	Tris-buffered saline Tween-20
UV	Ultra-violet

CHAPTER 1

REVIEW OF LITERATURE

1.1 INTRODUCTION

The structural components of the nucleus are known to have a central role in the specific topological organization of DNA. The chromatin fibre appears to be organized into domains or loops, which are constrained by a residual framework both in metaphase chromosomes and interphase nuclei. The residual framework of metaphase chromosomes has been called a "scaffold" since this structure retains some of the morphological features of metaphase chromosomes. The structural framework involved in the organization of interphase chromatin is thought to correspond to the nuclear matrix. The nuclear matrix is operationally defined as the residual structure after treatment of nuclei with detergents to remove the nuclear membranes, nuclease to digest nucleic acid, and buffer of high ionic strength to remove chromatin fragments and loosely bound proteins. The resulting residual structure, the nuclear matrix, resembles the nucleus in size and shape and consists of: (1) the residual elements of the nuclear envelope, also known as the pore-complex lamina; (2) the residual nucleoli; (3) a granular and fibrous internal matrix structure that extends

throughout the interior of the nucleus. Functionally, the nuclear matrix not only determines the size, shape and major morphological features of the nucleus, but also provides a solid support upon which the major functions of the nucleus take place. This review will focus on the various structural and functional aspects of the nuclear matrix with an emphasis on the proteins which are the major component of the internal matrix structure.

1.2 THE NUCLEAR MATRIX

The term nuclear matrix was first introduced by Berezney and Coffey (1974) to define a residual structural framework obtained from rat liver nuclei by sequential salt extractions, detergent and nuclease treatments. The resultant isolated three-dimensional structure consisted almost entirely of protein. Subsequent studies showed RNA as the second most abundant component (see section 1.3). The term nuclear matrix is often substituted in the literature with other terms such as nuclear scaffold, nucleoskeleton, nuclear ghost, and nuclear cage, although the term nuclear matrix is the most widely used. The nuclear matrix structure appears to be a universal feature of eukaryotic nuclei. It has been isolated from a wide variety of eukaryotic cell types from protozoan to human (Berezney 1984).

Most procedures for nuclear matrix isolation are based on the original protocols reported by Berezney and Coffey (1974, 1977). Basically, morphologically intact

nuclei are isolated and subjected to a series of treatments involving nuclease digestion, salt (2 M NaCl or 0.4 M $(NH_4)_2SO_4$) and detergent extraction. A key point is that a morphologically recognizable nuclear structure is maintained throughout the extraction protocol despite the removal of most of the chromatin and proteins and disruption of the nuclear membrane with detergent. In general, nuclear matrix preparations from different cells and tissues possess the same common morphological entities: (1) the residual elements of the nuclear envelope, also known as the pore-complex lamina; (2) the residual nucleoli; (3) a granular and fibrous internal matrix structure that extends throughout the interior of the nucleus.

1.2.1 The nuclear matrix controversy

Although there is consensus regarding the overall structure of the nuclear lamina, opinions regarding the internal network are still in conflict. This appears to be due to differences in the isolation conditions used, which can greatly affect the nature of the residual nuclear structures obtained.

The use of relatively harsh conditions to extract chromatin (2 M NaCl) raises the possibility that the residual insoluble structures obtained may be artifacts due to high salt-induced aggregation or precipitation of nuclear components that are soluble *in vivo*. With regard to high salt aggregation, it should be noted that high salt solutions are often used to prevent aggregation of protein in solution rather than to induce it.

Several workers have argued for participation of RNA in the structural integrity of the nuclear matrix. If the ribonuclease (RNase) digestion step is omitted from the isolation procedure, the resulting nuclear structures retain 60-80% of the heterogeneous nuclear RNA (hnRNA) and a dense internal network of ribonucleoprotein (RNP) fibres (Herman *et al.*, 1978; Miller *et al.*, 1978; Ciejek *et al.*, 1982). Subsequent RNase digestion has no significant effect on the morphology or protein composition of these structures, suggesting that RNA is not a structural component of the matrix. However, several groups have found that if RNase digestion precedes the high salt extraction, only the peripheral lamina is isolated (Herman *et al.*, 1978; Kaufmann *et al.*, 1981; Bouvier *et al.*, 1982). This has led to the suggestion that the internal matrix is an artifact formed by aggregation of the RNPs in high ionic strength buffers. Once formed, the protein-protein interactions between the RNP particles are resistant to subsequent RNase digestion, whereas RNase digestion before high salt extraction disrupts the integrity of the RNP particles, rendering their proteins soluble in high ionic strength buffers. However, some workers have isolated all three components of the nuclear matrix, even when the RNase digestion precedes high-salt extraction (Fisher *et al.*, 1982; Van Eekelen *et al.*, 1982). On the other hand, others argue that the RNP network exists *in situ*, and that RNase digestion artificially renders it soluble in high ionic strength buffers. The morphology and distribution of these domains is unaltered in nuclear matrices extracted with high salt in the absence of RNase digestion, suggesting that an RNP

network is present both before and after high salt extraction. RNase digestion of nuclei *in situ*, however, induces alterations in nuclear morphology and distribution of the RNP domains, suggesting that degradation of RNA *in situ* artificially disrupts the RNP network (Smith *et al.*, 1986).

Another factor that can affect the morphology and composition of the nuclear matrix is the extent of disulphide bond formation between nuclear proteins during the isolation procedure. Nuclei are exposed to oxidizing conditions during the various *in vitro* manipulations involved in isolating the nuclear matrix, through exposure to atmospheric oxygen and heavy metals that may contaminate buffers and glassware. Since these conditions vary among laboratories, the different types of residual structures isolated by different investigators may reflect the extent of oxidative cross-linking that has occurred between protein sulphydryl groups during the isolation procedure.

When nuclear matrices are isolated in the presence of .sulphydryl-blocking agents such as iodoacetamide (IAA) or N-ethylmaleimide (NEM), digested with (deoxyribonuclease) DNase and RNase, and extracted with 1.6 M NaCl, the resulting residual structures contain the peripheral lamina but lack the internal matrix components (Kaufmann *et al.* 1981; Kaufmann & Shaper, 1984). These agents irreversibly block free sulphydryl groups, but do not reduce existing disulphide bonds, suggesting that the internal matrix network observed in their absence depends on disulphide cross-links formed *in vitro*, rather than cross-links present *in vivo*. In

the presence of sulphydryl cross-linking agents such as sodium tetrathionate (NaTT), the same isolation procedure yields residual structures containing the peripheral lamina, plus an extensive network of internal fibres and prominent residual nucleoli (Kaufmann *et al.*, 1981; Kaufmann & Shaper, 1984). Subsequent treatment of these cross-linked structures with reducing agents such as dithiothreitol (DTT) or β-mercaptoethanol renders the internal matrix components soluble in 1.6 M NaCl, whereas the lamina remains insoluble (Kaufmann & Shaper, 1984). Similar observations have been made in 3T3 and HeLa cells (Staufenbiel & Deppert, 1984). One interpretation of these results is that the internal matrix observed in the absence of sulphydryl-blocking agents is an artifact due to disulphide cross-linking *in vitro* between proteins that do not interact in the intact nucleus, but do so during matrix isolation as a result of rearrangements induced by 1.6 M NaCl. Alternatively, the internal matrix may exist *in vivo*, but the protein-protein interactions maintaining its integrity may simply be disrupted by high ionic strength buffers unless they are stabilized by disulphide cross-linking, either fortuitously or deliberately. In this case, the disulphide bonds merely stabilize protein-protein interactions which exist in the intact nucleus. Consistent with this interpretation is the observation that NaTT does not result in indiscriminate cross-linking of nuclear proteins, but only a specific subset of non-chromatin, non-lamina proteins are stabilized (Kaufmann & Shaper, 1984; Kaufmann *et al.*, 1986). In contrast to these results, however, others have found that the internal matrix components are isolated regardless of whether IAA or

NEM is present during isolation of residual nuclear structures from HeLa, 3T3 (Capco *et al.*, 1982; Van Eekelen *et al.*, 1982) and *Drosophila* embryo cells (Fisher *et al.*, 1982).

Clearly, resolution of these contradictory results and interpretations requires the demonstration of an internal nuclear matrix *in situ*. Determination of the precise architecture of the internal matrix *in situ* has proven difficult, since this structure is obscured by dispersed chromatin and the mass of other nucleoplasmic constituents. Probably the best visualization *in situ* of internal matrix elements is obtained in cells exposed to the transcription inhibitor α-amantin (Brasch 1982). In the presence of this inhibitor chromatin becomes condensed and RNP aggregation occurs which "clears" the diffuse nucleoplasmic background. This reveals an *in situ* nuclear matrix, visible in the electron micrograph as an interchromatin network of fibres, 10-15 nm in diameter, contiguous with the chromatin and the nuclear envelope. These fibres are also visible when poliovirus-infected human fibroblasts are extracted with non-ionic detergent at physiological ionic strength (Capco *et al.*, 1982). In this case, the virus induces chromatin clumping, revealing the interchromatin fibre network. The same extraction procedure also reveals internal matrix fibres in cells entering prophase, as chromatin starts to condense for mitosis (Capco & Penman, 1983).

Penman and coworkers have used RNase-free conditions and ammonium sulphate (0.25 M) instead of NaCl for extraction (Fey *et al.*, 1986; Nickerson *et al.*, 1989).

In their method, nuclei are fractionated *in situ*, without the prior isolation of nuclei. First cells are extracted with non-ionic detergents, leaving the nucleus still associated with the cytoskeleton. After digestion with RNase-free DNase I, chromatin is extracted with 0.25 M $(NH_4)_2SO_4$, which removes more than 95% of the chromatin while leaving the nuclear matrix intact (Fey *et al.*, 1986). The *in situ* fractionation procedure uncovers a nuclear matrix consisting of two parts. At the periphery of the nucleus is the nuclear lamina, to which is connected the intermediate filaments of the cytoskeleton. The nuclear matrix and the intermediate filaments of the cytoskeleton form a single cell-wide structure (nuclear matrix-intermediate filaments complex) integrated at the nuclear lamina. Inside the lamina are thick, polymorphic fibres connecting the lamina to masses in the nuclear interior. This complex interior matrix structure can be further fractionated by extraction with 2 M NaCl to uncover a highly branched network of 9 nm and 13 nm core filaments with some granular bodies enmeshed in the network (He *et al.*, 1990). This network is completely disrupted if treated with RNase suggesting a role for nuclear RNA in the integrity of the nuclear matrix.

Jackson & Cook (1985a) developed a procedure that allows one to maintain near-physiological ionic conditions throughout the matrix isolation procedure, making use of cells en-capsulated in agarose beads. Agarose embedded cells are permeablized with non-ionic detergents, DNA is digested with restriction endonucleases and subsequently removed by electrophoresis. During the procedure a constant near-

physiological ionic condition is maintained. Using this procedure, the nucleoskeleton in resinless, thick sections were visualized by electron microscopy. In HeLa cells, this procedure has uncovered a "nucleoskeleton", a three-dimensional network of nuclear filaments extending from the lamina throughout the nuclear interior (Jackson & Cook, 1988). The dimensions of these filaments are similar to those of the cytoplasmic intermediate filaments (IFs), i.e. ~10 nm diameter with a 23 nm axial repeat, suggesting that IFs may form the skeleton of both the cytoplasm and the nucleus. The relationship of these intranuclear filament-like structures to the core filaments (He *et al.*, 1990) described above is not clear. Such questions can only be resolved when the composition of the filamentous system is known.

1.2.2 Molecular composition

The precise molecular composition of the nuclear matrix remains to be clarified. Much progress has been made in terms of peripherally located components. A limited number of polypeptides account for the bulk of the residual pore-complex-lamina in most types of nuclei.

In preparing nuclear matrices, the nuclear envelope is exposed to buffers containing non-ionic detergents, nucleases and high-molarity salt buffers. Morphologically, only the pore complex and the lamina are resistant to such treatments. This residual framework, pore-complex-lamina is considered to be a part

of the nuclear matrix.

The nuclear lamina is a fibrillar meshwork of proteinaceous material, which is intercalated between the chromatin and the inner membrane of the nuclear envelope. In mammals, three polypeptides with apparent molecular weights of 70, 67, and 60 kDa predominate. Immunofluorescence and differential solublization studies show that these three proteins are confined to the lamina. They have been designated "lamins" A, B, and C in order of decreasing molecular weight.

The three lamins visualised by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis of rat liver pore-complex-lamina are typical of mammalian and avian somatic cells, but are not universal in nature. Amphibian somatic cells have up to four lamins, many invertebrates have two, and gametes such as amphibian oocytes have one (reviewed by Krohne & Benavente, 1986).

cDNAs for lamins have been isolated and characterized. McKeon *et al.* (1986) and Fisher *et al.* (1986) have characterized the cDNA clones for human lamins A and C. From the protein sequences deduced from these cDNAs it became apparent that the two proteins are identical, except for an additional 9 kDa region at the carboxy-terminus of lamin A and a six amino acid extension at the carboxy-terminus of lamin C. Both are synthesized by different mRNAs (Laliberte *et al.*, 1984), but they may come from a single gene upon differential splicing (McKeon *et al.*, 1986; Fisher *et al.*, 1986). More importantly, these studies have shown that both lamin A and C share striking structural homologies with the entire family of cytoskeletal intermediate

filament proteins (McKeon *et al.*, 1986; Fisher *et al.*, 1986). The homologies concern repeated heptads of amino acids contained within large internal α-helical domains that are believed to be important for the formation of two-stranded parallel coiled-coil structures (Franke, 1987). Moreover, purified lamins A and C were shown to have the potential for assuming all the major structures characteristic of intermediate filament proteins, at the level of assembled filaments (Aebi *et al.*, 1986). Thus based on both sequence analyses and morphological criteria, lamins A and C can be considered as nuclear intermediate filament proteins. This conclusion is supported also by physicochemical, immunological, and morphological evidence indicating a resemblance between nuclear lamins and keratin-type intermediate filaments (Goldman *et al.*, 1986).

Using a mouse monoclonal antibody (IFA) raised against a common domain of all intermediate filament proteins, Lebel and Raymond (1987) as well as Osborn and Weber (1987) have shown that lamin B also shares some sequence homology with the intermediate filament proteins, and recent cDNA analyses demonstrate that B type lamins also belong to the intermediate filament protein family (Krohne *et al.*, 1987).

There are two very striking biological differences between the lamins and the intermediate filament proteins: (1) the former are confined to the nucleus in interphase, the latter probably to the cytoplasm; (2) the former are ubiquitous in eukaryotes, and the latter are not.

How are the lamins linked together in the lamina? Presumably lamin A/C

oligomers are attached to the monomeric inner nuclear membrane associated lamin B. The A/C lamins can polymerize under different ionic conditions, just as cytokeratins do (Aebi *et al.*, 1986). They seem to form a square meshwork pattern *in vivo* with a 10 nm fibre width suggesting end-to-end linking of the molecules. If the B molecules join to A/C at the cross-over points of the mesh, the roughly 50 nm mesh spacing suggests that lamin B is linked to A/C dimers *in situ*. A 32-residue N-terminal sequence of lamin A, synthesized from a cDNA contains the binding site for lamin B. The binding of B to A and C seems to be cooperative, according to binding studies with column-immobilized lamins (Georgatos *et al.*, 1988). Presumably this cooperativity partly accounts for the stability of the interphase lamina. The interphase lamina is a relatively insoluble structure. However, during mitosis, when the nuclear envelope breaks down, the lamina is rapidly disassembled. Lamins A and C become dispersed as monomers throughout the mitotic cytoplasm, while lamin B remains associated with vesicles, presumably derived from the interphase nuclear membrane (Burke & Gerace, 1986).

In vivo studies have shown that the disassembly of the lamina during prophase coincides with hyperphosphorylation of each of the lamins (Gerace & Blobel, 1980; Burke & Gerace, 1986), while during telophase, when nuclear envelopes reassemble about the daughter nuclei, the lamins become dephosphorylated. These observations have led to the view that lamina dynamics are regulated by phosphorylation. *In vitro* studies strongly support this notion (Newport & Spann, 1987).

The pore complexes are large organelles that form channels for nucleocytoplasmic transport through the nuclear envelope (reviewed by Newport & Forbes, 1987). The nuclear pore complex has an estimated molecular weight of 124 million daltons and is mainly composed of protein (Reichelt *et al.*, 1990). Nuclear pore proteins were first identified using the combined approach of raising antibodies against rat liver nuclear fractions and electron microscopic immunolocalization. One of these, gp210, is an integral membrane protein located exclusively at the pores (Gerace *et al.*, 1982). This protein is a glycoprotein rich in asparagine-linked mannose residues. Because of its transmembranous orientation, gp210 is thought to play a role in anchoring the pore complex to the nuclear membrane. Additionally, nuclear pores have been found to contain a group of at least eight novel glycoproteins. These polypeptides with apparent molecular weights of 270, 210, 180, 145, 100, 62, 58, 54 and 45 kDa each contains a single residue of N-acetyl glucosamine O-linked to serine and threonine residues (Snow *et al.*, 1987; Hanover *et al.*, 1987). These proteins were subsequently named nucleophorins. However, each of these proteins is extracted by high ionic strength buffers, and so does not appear to be a component of the residual nuclear matrix.

The internal components of the nuclear matrix have proven much more difficult to characterize. This is due to the molecular complexity and the apparent labile nature of the nuclear matrix structure. It is clear from the above discussion that the morphological complexity of isolated matrices can be significantly affected by such

factors as the sequence of extraction steps, the presence or absence of RNase in the nuclease digestion step, and the extent of disulphide bond formation between nuclear proteins during the isolation procedure. Despite all these variations in matrix isolation protocols, it is remarkable that the resultant matrices often exhibit an internal structure directly comparable to that of the interchromatin region *in situ*. What then is the molecular nature of the internal matrix? The question remains unresolved. However, over the last few years a number of nuclear proteins have been identified as forming part of the nuclear matrix.

Analysis of nuclear matrix proteins from mammalian cells on one dimensional SDS polyacrylamide electrophoresis gels shows the existence of numerous polypeptides ranging from M_r 50 to M_r 200 kDa but most prominent are proteins between M_r 60 and M_r 75 kDa (Berezney, 1984). In addition to peripherally located polypeptides (lamin A, B, and C) and nucleolar proteins, nuclear matrix isolates from most sources also contain significant amounts of RNP-derived elements, including both proteins and RNA.

Two dimensional analyses of nuclear matrix proteins performed by several different groups all stress the high degree of complexity of these polypeptide profiles. Using ^{35}S-methionine labelling for detection, Fey and Penman (1988) have detected over two hundred proteins in the nuclear matrix. Struurman *et al.* (1990) have also found enormous complexity in the two-dimensional profiles with the sensitive silver staining procedure. Despite this complexity, these studies provide valuable

information. For example, the total nuclear matrix proteins can be separated into two major groups. Those which are found in a variety of cell lines (common matrix proteins) and those which are both cell type and differentiation state dependent (Fey & Penman, 1988, Struurman *et al.*, 1990; Diworetzky *et al.*, 1990).

In contrast to our rather extensive knowledge of lamins, the proteins of the nuclear matrix within the lamina have not been well characterized. Only a few of these proteins have been purified, and biochemical and functionally characterized. Both biochemical fractionation procedures and immunological methods have been used for this purpose.

Using a preparative two-dimensional SDS polyacrylamide gel electrophoresis gel system Nakayasu and Berezney (1991) identified at least 13 major proteins from the rat liver nuclear matrix. Of these, 5 proteins identified as nuclear lamin A, B, and C, the nucleolar protein B-23, and a protein from heterogeneous nuclear ribonucleoprotein (hnRNP) particles. The remaining eight major proteins, which they named the nuclear matrins consisted of matrin 3 (M_r 125 kDa, slightly acidic) matrin 4 (M_r 105 kDa, basic), matrin D, E, F and G (M_r between 60-75 kDa, basic) and matrins 12 and 13 (M_r 42-48 kDa, acidic). Using a two-dimensional south-western blotting technique Hakes and Berezney (1991) identified seven major nuclear matrix proteins as DNA binding proteins including lamins A and C but not B, and the internal nuclear matrix proteins, matrin C, D, F, G and 4.

The DNA binding proteins matrins F and G and matrin 3, a nuclear matrix

protein with unknown function and properties, have been cloned recently (Belgrader *et al.*, 1991). Consistent with their DNA binding properties, the predicted amino acid sequence from the coding regions of the matrin F and G cDNA showed that these proteins contain two overlapping putative zinc finger domains which have similarity to the cysteine-cysteine type zinc finger motif, but no known DNA binding motif was identified in matrin 3.

1.2.3 DNA organization

There are at least four levels of DNA organization in the nucleus. At the lowest level the linear, double-stranded DNA molecule combines with histones, thus forming a string of nucleosomes called the 10 nm filament. Nucleosomes are disc-like particles consisting of a core of eight histones surrounded, in a wrap-like manner, by about 140 bp of DNA. The next higher level of organization is the folding of the 10 nm filaments into a fibre made up of helically packed nucleosomes, which has a diameter of 30 nm. Most of the chromatin in metaphase chromosomes and in the interphase nucleus occurs in this folded state. At an even higher level of organization the 30 nm fibre is folded in loops. These loops seem to constitute discrete functional and structural chromatin units.

Evidence for the organization of DNA into loop domains has been obtained by a variety of methods. Cook and Brazil (1975, 1976) noted that the sedimentation rate

of histone-depleted interphase nuclei varied biphasically as a function of the ethidium-bromide concentration in the gradient. Based on experiments with supercoiled plasmid DNA, it was concluded that this sedimentation pattern was most easily explained by postulating that the DNA in the histone-depleted nuclei was topologically constrained and supercoiled. An estimate of the size of the supercoiled domain was made by measuring the number of nicks required to eliminate the biphasic sedimentation response to ethidium bromide. It was found that on average one nick every 220 kbp of DNA in the histone depleted HeLa cell nuclei would eliminate the ethidium bromide response thus suggesting that the DNA loops of average size 220 kbp. Benyajati and Worcel (1976) performed similar experiments, and came up with an average loop size of 85 kbp. Chromatin loops have been directly visualized by fluorescence microscopy of histone-depleted interphase nuclei of 3T3 cells (Vogelstein *et al.*, 1980). After staining with ethidium bromide, the loops are visible as fluorescent haloes of DNA surrounding a residual nuclear skeleton or matrix. The diameter of the halo is dependent upon the concentration of the ethidium. Therefore, the halo appeared to consist of topologically constrained loops of DNA. From the diameter of the halo the average size of the loops has been estimated as 90 kbp, which was consistent with the size of loops measured by biophysical techniques. More recently, an average loop size of 86 kbp in interphase nuclei from HeLa cells has been estimated (Jackson *et al.*, 1990).

This organization of DNA into loop domains is a structural feature of both

interphase and metaphase chromosomes. Laemmli and coworkers have demonstrated that the organization of chromatin loop domains in interphase nuclei is preserved when the chromatin condenses into mitotic chromosomes (Laemmli *et al.*, 1978). Direct electron microscopic measurements show that these loops are the same size as those found in interphase nuclei.

When histones and some other chromatin-associated proteins are extracted from isolated metaphase chromosomes by treatment with solutions of high ionic strength, the DNA remains anchored to a residual proteinaceous scaffold that retains the basic morphological organization of an unextracted metaphase chromosome. Two major proteins, Sc1 (scaffold protein 1) and Sc2, of about 170 and 135 kDa, respectively, are present in these preparations, in addition to a number of minor proteins (Lewis and Laemmli, 1982). Together the proteins represent 1-2% of the total chromosomal proteins. The most abundant of the two major proteins, Sc1, is topoisomerase II (Earnshaw *et al.*, 1985; Gasser *et al.*, 1986). Toposiomerase has been shown to be present exclusively in the central axes of the two sister chromatids of intact, swollen metaphase chromosomes, as visualized by indirect immunofluorescence microscopy and immunoelectron microscopy (Earnshaw & Heck, 1985). The majority of the total immunologically detectable topoisomerse II pool in metaphase cells can be recovered from the chromosomal fraction (Earnshaw *et al.*, 1985). Approximately three molecules of topoisomerase II are present per 63 kbp chromatin loop (Gasser *et al.*, 1986). It is not clear how topoisomerase II molecules interact with the rest of

the scaffold. Immunofluorescence and immunoelectron microscopic data suggest a clustering of toposiomerase II molecules in "Islets" along the central chromatid axis rather than a continuous framework of topoisomerse II polymers (Earnshaw & Heck, 1985).

An obvious question posed by the studies demonstrating the existence of DNA loops in the nucleus concerns the organization of these loops with respect to the DNA sequences within them. In particular, it is important to determine whether the anchorage points of the loops are organized randomly or specific sequences form the anchorage points of the loops. The studies from various labs summarized below have shown that loops are organized non-randomly.

The strategy employed to determine wether specific sequences anchor loops used nuclei treated with 2 M NaCl to remove histones and most other chromatin associated proteins. This resulted in the appearance of a halo of loops of DNA which were anchored to the residual nuclear matrix at the centre of the halo. The DNA sequences in the halo could be progressively removed by treating these matrix-halo structures with DNase I or restriction endonucleases. By fluorescence microscopy the halo could be seen to gradually decrease in size and intensity until only a small faint halo surrounding the nuclear matrix remained. The DNA remaining attached to the nuclear matrix represented those sequences that occur at the base of the DNA loops, thus anchoring them. This anchored DNA, still attached to the matrix, was separated from nuclease released DNA by low-speed centrifugation. The matrix

DNA was then purified and assayed for its relative content of specific sequences. If the DNA was organized randomly with respect to the loop, the concentration of any sequence should be equal in the matrix DNA and total nuclear DNA. If DNA was not randomly organized, then sequences near the anchoring points of the loops should be greatly enriched in matrix. Using this strategy, ribosomal RNA genes of rat liver (Pardoll & Vogelstein, 1980), the SV40 sequences of several SV40-transformed 3T3 cell lines (Nelkin *et al.*, 1980), and sequences coding for mRNA (Jackson *et al.*, 1981) have been found to be preferentially associated with the nuclear matrix.

In their studies on attachment of DNA to the nuclear matrix, Laemmli and coworkers pioneered the use of the ionic detergent lithium diiodosalicylate (LIS) for isolation of subnuclear structures (Mirkovitch *et al.*, 1984). In their procedure, isolated nuclei are stabilized by incubation at 37° C or 42° C, sometimes combined with incubation in buffer containing millimolar concentrations of Cu++ ions. Subsequently, histone and other nuclear proteins are extracted with low concentration of LIS (10-25 mM) and DNA is digested with restriction enzymes to yield a nuclear matrix preparation (Mirkovitch *et al.*, 1984). This method has facilitated the identification and characterization of DNA sequences that remain bound to the nuclear matrix. Such DNA fragments specifically bind to purified nuclear scaffolds and were originally called 'scaffold-associated regions' (SARs). Since in this review I have chosen the term nuclear matrix for the scaffold structure, the DNA fragments that bind specifically to this structure will be referred to here as 'matrix-associated

regions' (MARs), following the nomenclature of Cockerill and Garrad (1986a). Using the criterion of specific binding to nuclear preparations, MARs have been found in the vicinity of a variety of genes. It is thought that MARs form the basis of chromatin loops, anchoring them to the nuclear matrix. Most MARs found so far are located in non-transcribed regions. However, functionally they confer transcriptional activity to genes in which they are inserted. Classic experiments by Stief and coworkers involved utilizing the MAR sequences of the chicken lysozyme gene and inserting this sequence into a transfectable expression vector carrying a reporter gene (Steif *et al.*, 1989). When this reporter system is flanked by the 5' MAR, its expression is significantly increased and is independent of chromosome position (Stief *et al.*, 1989).

Comparison of MARs that have been identified up to now indicate the following general properties: MARs and nuclear matrix from a variety of distantly related species bind each other (Cockerill & Garrad, 1986b) suggesting a common conserved mechanism of MARs and matrix interaction. Comparison of the sequences of a large number of MARs clearly shows that they are not closely related, they do not cross-hybridize and do not reveal strict consensus sequences. They are however, in general 70% A+T-rich and frequently contain topoisomerase II cleavage sites (Gasser & Laemmli, 1987). This is consistent with the fact that topoisomerase II is tightly bound to the nuclear matrix (Berrios *et al.*, 1985). Potentially, topoisomerase II may control, through its interaction with a MAR, the degree of supercoiling of a

chromatin loop. Moreover, in a number of cases enhancer sequences were found in the vicinity of a MAR (Cockerill & Garrad, 1986a; Gasser & Laemmli, 1986b).

A specialized DNA structure formed by certain A+T-rich sequences may be important for the biological function of MARs. The significance of structural characteristics of MARs, such as DNA bending (Von Kries *et al.*, 1990) and a narrow minor groove attributable to oligo(dA) tracts (Adachi *et al.*, 1989), has been proposed. Moreover, MARs from different species are characterized by their strong potential for extensive un-pairing (or unwinding) when subjected to superhelical strain (Bode *et al.*, 1992). This unwinding property has been shown to be important for binding to the nuclear matrix and for the marked increase in gene expression in stable transformants (Bode *et al.*, 1992). Two MARs flanking the immunoglobulin heavy chain gene enhancer were shown to be continuously unpaired over a distance of more than 200 bp in supercoiled plasmid DNA. A short sequence motif, ATATAT, within the MAR located 3' of the immunoglobulin heavy chain enhancer was delineated to be a nucleation site for unwinding (Cockerill *et al.*, 1987). Point mutations substituting three bases in this sequence completely abolished the unwinding property of the MAR (Kohwi-Shigematsu & Kohwi, 1990). In a similar study, Bode and coworkers showed that a concatemerized, double-stranded 25 bp oligonucleotide containing the core unwinding sequence of the 3' MAR behaved like a typical MAR (Bode *et al.*, 1992). This synthetic MAR is capable of unwinding under superhelical strain, strongly binding to the nuclear matrix, and enhancing transcription of a linked

reporter gene in stable transformants (Bode *et al.*, 1992). However, none of these features is observed with a similarly concatemerized, double-stranded oligonucleotide derived from the mutated core sequence; the unwinding property is lost, the binding affinity to the nuclear matrix is greatly reduced, and no enhancement of transcription is detected (Bode *et al.*, 1992).

1.2.3.1 Proteins that binds MARs

In the following section I will focus on a few recent studies of nuclear matrix proteins that I think may be relevant to the question of chromatin organization.

The observations presented above suggest that some of the nuclear matrix proteins that bind MARs might not be strictly sequence-specific DNA-binding proteins, but rather proteins that recognized characteristic DNA structures. This is supported by the observation that distamycin, a drug which selectively binds the minor groove of A+T-rich DNA, abolished DNA binding of all *Drosophila* MARs tested (Kas *et al.*, 1989). Several abundant nuclear proteins, such as topoisomerase II, histone H1, and HMG proteins which can bind any DNA, appear to bind with higher affinity to regions having a narrow minor groove. Topoisomerase II and histone H1 have been shown to interact preferentially with MAR, in competition studies (Adachi *et al.*, 1989; Kas *et al.*, 1989), and this binding is quite efficiently competed by distamycin. However, it seems very unlikely that the general population of histone H1 mediates matrix attachment, because only 5% of this protein remains

in the nuclear matrix preparation (Kas *et al.*, 1989).

The nuclear matrix associated DNA sequences (MARs) often contain a consensus cleavage site for topoisomerase II, which binds preferentially and cooperatively to these sequences (Adachi *et al.*, 1989), and cleaves them (Sperry *et al.*, 1989) *in vitro*. These observations, together with the fact that topoisomerase II is a major protein component of metaphase chromosomes, led to the proposal that topoisomerase II serves an important structural role by anchoring MARs to the nuclear matrix (Gasser & Laemmli, 1987). However, Sperry and coworkers observed that topoisomerase II does not bind strongly to all MARs *in vitro* and that it binds strongly to some non-MAR sequences (Sperry *et al.*, 1989). The authors also point out that the nuclei of resting cells have very low levels of topoisomerase II, yet their isolated nuclear matrices retain the ability to bind MARs, suggesting that other proteins must participate in loop attachment. While toposomerase II binding sites are generally associated with MARs, they are neither necessary nor sufficient to specify a MAR. Kaufmann and Shaper (1991) have shown that the enrichment of topoisomerase II in nuclear matrix preparations may be due in large part to the formation of insoluble topoisomerase II oligomers linked by intermolecular disulphide bonds during sample preparation. They have been unable to demonstrate the existence of these cross-linked complexes in intact cells.

Hofmann and coworkers have developed a procedure to solubilize the yeast nuclear matrix fraction in urea, and then renature the proteins to form a soluble

extract. This soluble extract has been used to purify the yeast MAR-binding proteins RAP1 (*R*epressor-*A*ctivator *P*rotein 1) and ACBP (*A*RS *C*onsensus-*B*inding *P*rotein; Hofmann *et al.*, 1989; Hofmann and Gasser, 1991). RAP1 recognizes the sequences which occur at numerous sites both in regulatory elements such as the mating-type silencer and upstream activation sequences, and in structural DNA elements such as telomeres. At the silent mating-type locus, RAP1 binds at two sites flanking two transcriptionally inactive genes, and both are matrix attached. In reconstitution experiments, RAP1 can form DNA loops *in vitro* with a linear DNA fragment containing MAR sequence and this has been visualized by electron microscopy (Holfmann *et al.*, 1989).

Recently, a new nuclear matrix protein that binds cooperatively and specifically to MARs has been purified from chicken oviduct cells (Von Kries *et al.*, 1991). This protein (ARBP, *a*ttachment *r*egion *b*inding *p*rotein) has an apparent molecular weight of 95 kDa and is a constituent of the internal nuclear matrix. ARBP binds to a variety of homologous and heterologous MAR elements but no unique sequence element responsible for binding could be identified (Von Kries *et al.*, 1991).

Using a MAR DNA found in intron 13 of the human topoisomerase I gene in a DNA binding assay, Romig and coworkers have purified a protein SAF-A (*s*caffold *a*ssociated *f*actor A) from HeLa cells (Romig *et al.*, 1992). This protein has a molecular weight of 120 kDa; it can bind at multiple sites to the human MAR element; it can form large aggregates, and it also mediates the formation of looped

DNA structures.

Most recently, a human cDNA encoding a DNA-binding protein SATB1 (*s*pecial AT *b*inding protein 1) expressed predominantly in thymus has been isolated (Dickinson *et al.*, 1992). The protein encoded by this cDNA clone has a molecular weight of 85.9 kDa and binds specifically to the MARs from homologous and heterologous MARs elements. Chemical interference assays show that SATB1 binds along the minor groove with very little contact with the bases. Moreover, the protein selects for A+T-rich sequences with a high unwinding potential. (Dickinson *et al.*, 1992).

1.2.3.2 Biological significance of loop domains

The interaction between MARs and nuclear matrix seems to be maintained throughout the cell cycle, since *Drosophila* metaphase chromosome scaffolds bind the same MARs as nuclear matrices from interphase cells (Mirkovitch *et al.*, 1988).

The maintenance of MAR binding sites throughout the cell cycle, their conservation during evolution and the observation that a MAR, which may contain one or more consensus cleavage sequences for topoisomerase II, is frequently found near an enhancer, suggest that MARs have an important function in organizing the genome.

Gene activation, for example during development, can be envisaged as a process regulated at two levels. At the first level, regulatory sequences such as promoters

and upstream regulatory elements are made accessible to the transcription machinery via the transition of a chromatin loop from a condensed state to a less condensed state. Now the gene(s) in the domain is (are) committed for transcription. At the second level cellular factors determine the actual transcription rate of the gene involved. The first level of transition may be responsible for the increased, tissue-specific, sensitivity of active or committed genes towards digestion with DNase I as compared with their inactive counterparts (Weisbrod, 1982). In some cases, it has been shown that a large area of increased sensitivity extends outwards from the actual transcribed region of the genes involved. For instance, the high DNase I-sensitivity of the globin coding region in chicken erythrocytes is flanked by regions of at least 8 kbp distinguished by a level of DNase I-sensitivity intermediate between that of an expressed and an inactive gene (Stalder *et al.*, 1980). Likewise, the increased DNase I-sensitivity of the ovalbumin locus comprises a region of approximately 100 kbp of DNA, flanked by areas of low nuclease sensitivity (Lawson *et al.*, 1982). It is very tempting to speculate that the borders between regions of different nuclease sensitivity are formed by matrix attachment sites, preventing the structural changes that underlie domain activation to propagate into neighbouring areas of the chromatin.

1.2.4 DNA replication

The most studied and best characterized function associated with nuclear matrix

is DNA replication. Berezney and Coffey (1975, 1977) reported that newly synthesized DNA is tightly associated with the nuclear matrix. At the earliest time intervals after injections of ^{3}H-thymidine into the hepatic portal vein of partially hepatectomized rats, nuclear matrix fractions were obtained which contained tightly attached DNA highly enriched in label. At later times after the injection, the specific activity of the matrix-associated DNA decreased, while that of the bulk DNA increased (Berezney and Coffey 1975, 1977). They concluded that the newly synthesized DNA is associated with the nuclear matrix and subsequently moved out of the matrix.

Further insight into nuclear matrix-associated DNA replication has been obtained from studies carried out by Pardoll and coworkers, using 3T3 cells in culture. After pulse labelling of growing cells with radioactive thymidine and isolation of nuclei, DNA was made to loop out by extraction of histones with 2 M NaCl. Autoradiography showed that nascent DNA was exclusively found over the nuclear matrix part of the nucleoids and not in the haloes representing the DNA loops (Pardoll *et al.*, 1980). From a detailed analysis of the kinetics of pulse-chase labelling studies, they proposed a fixed site model of DNA replication in which DNA replication complexes are anchored to the nuclear matrix, and that loops of DNA are reeled through these fixed complexes as they replicated. According to the model multiple loops of DNA are attached at their base to the nuclear matrix. These two sites of attachment for each loop contain a fixed site for DNA replication. During

replication, each DNA loop moves downward, through each of the two fixed replicating complexes. As the parent DNA loop passes through the fixed replication sites, two new loops are formed and the old loop disappears. Each of the new loops contains one single strand of the parent DNA combined with a single strand of newly synthesized DNA (Pardoll *et al.*, 1980).

The replication of DNA in the nucleus occurs in a series of tandemly repeated subunits along the continuous DNA molecules; these subunits are termed, replicons. Adjacent replicons are further organized into clusters that replicate as a unit at particular times in S phase (Lau & Arrighi, 1981).

The organization of DNA into loop domains creates discrete functional and structural units of chromatin. A possible relationship between loops and replication is suggested by the fact that DNA synthesis in eukaryotic cells starts from many origins, the distances between which correspond roughly to the size of a loop (Goldman, 1988). The comparable sizes of replicons could well be explained by assuming that loops function also as units of replication or replicons.

Origins of DNA replication appear to be matrix associated throughout the cell cycle, whereas growing replication forks are attached to the nuclear matrix only transiently during the DNA synthetic (S) phase (Dijkwel *et al.*, 1986).

Circumstantial evidence for permanent attachment of origins came from estimations of the length of DNA loops between attachment sites. From measurements of diameters of DNA haloes obtained by treatment of nuclei with 2 M NaCl, it is

inferred that each loop of DNA corresponds to a replicon.

Strong evidence on the permanent binding of origins to the nuclear matrix has been obtained by analysis of the positioning of pulse labelling studies. These studies showed that DNA segments labelled by a short pulse at the very beginning of the S phase remained associated with nuclear matrix after varying periods of chase. In this way it has been found that origins remain bound to the nuclear matrix during the entire cell cycle, including mitosis (Van der Velden *et al.*, 1984).

Recent data further support the clustering of replicons in the nucleus. When replicating DNA is labelled with bromo-deoxy uridine (BrdU) and then reacted with fluorescent conjugated anti-BrdU antibodies, highly defined foci as opposed to diffuse staining have been observed throughout the nucleus (Nakamura *et al.*, 1986).

Studies in Berezney's laboratory also indicate that there are discrete sites of DNA synthesis in the cell nucleus; DNA synthesis sequestered within these discrete sites, termed replication granules, are distributed throughout the nuclear interior(Nakaysau & Berezney, 1989). These granules have been observed by incorporating biotin-11-dUTP into newly synthesized DNA using permeabilized 3T3 cells. The sites of biotinylated, newly synthesized DNA are then directly visualized by fluorescence microscopy following reaction with Texas-red-streptavadin (Nakayasu & Berezney, 1989). The characteristic size and shape of replication granules observed in permeabilized cells remained the same following *in situ* nuclear matrix preparation. These results demonstrate that components of the replicational machinery maintain

sites on the nuclear matrix which corresponds to the presumed replicon cluster site in intact cells.

The proposed model for DNA replication in association with the nuclear matrix requires that enzymes involved in DNA synthesis also be matrix associated. Fulfilling this requirement, a cell cycle and replicative-dependent association has been shown between DNA polymerase α (the lagging strand polymerase) and the nuclear matrix isolated from rat liver cells (Smith & Berezney, 1982). In addition, matrix-bound DNA polymerase α is also capable of synthesizing DNA on matrix-attached DNA fragments *in vitro* and this *in vitro* DNA synthesis occurs at replicational sites previously initiated *in vivo* (Tubo & Berezney, 1985). Moreover, high resolution electron microscopic autoradiography suggested a similar localization of *in vivo* DNA replicational sites and DNA polymerase α in the nuclear matrix interior (Smith *et al.*, 1984).

The DNA replicating complexes associated with the nuclear matrix have been isolated and termed replisomes. The replisome is a 24-30 nm diameter particle with a molecular weight of approximately 5 million daltons and contains at least eight enzymes which include ribonucleoside diphosphate reductase, thymidylate synthetase, dihydrofolate reductase, DNA methylase, topoisomerase, and DNA polymerase (Tubo & Berezney, 1987a, 1987b).

In summary, most of the nuclear machinery required for replication is localized in discrete domains in the nucleus, rather than being diffusely distributed throughout

the nucleus. Since these domains appear to be associated with the nuclear matrix, it is most likely that this scaffolding structure has a key function in DNA replication.

1.2.5 RNA transcription, processing and transport

A major function of the nucleus is the production of specific mRNAs. Primary transcripts are processed to mature mRNAs and transported to the cytoplasm via the nucleopore complex. The nuclear matrix appears to provide a solid support for all these processes related to transcription. A lot of evidence has been accumulated which suggests that transcription takes place at the nuclear matrix. Pulse labelled, nascent RNA remains associated with the nuclear matrix (Jackson *et al.*, 1981, Ciejek *et al.*, 1982; Jackson *et al.*, 1985b). Labelled RNA is never seen associated with DNA that loops out of the nuclei after removal of histones (Jackson *et al.*, 1981).

Numerous studies from various laboratories have shown that essentially all of newly synthesized hnRNA is retained with the nuclear matrix and, in fact, may be an integral component of it, necessary for maintaining nuclear structure (Fey *et al.*, 1986). The association of hnRNA with the nuclear matrix seems to occur as early as during transcription (Schroder *et al.*, 1987). The binding of mRNA and hnRNA to the nuclear matrix is very stable and resists treatment with high concentration of salt, urea, EDTA, or non-ionic detergents (Schroder *et al.*, 1987). Two polypeptides of 41.5 and 43 kDa identified by RNA-protein cross-linking experiments are probably

involved in the binding of hnRNA to the nuclear matrix structure and appear to be components of isolated RNP particles (Van Eekelen & Van Venrooij, 1981).

Several nuclear proteins involved in regulation of transcription, such as steroid hormone receptor complexes, viral T antigens, and adenovirus E1A protein are also associated with the nuclear matrix (Barrack & Coffey, 1982; Steaufenbiel & Deppert, 1983).

The association of transcriptionally active DNA with the nuclear matrix has been demonstrated by various workers. The DNA fraction remaining attached to the nuclear matrix after extraction of the nuclei with high-ionic strength salt and treatment with DNase I or restriction endonucleases has been shown to be enriched in transcriptionally active gene sequences. Using nuclear matrices prepared by either DNase or restriction endonuclease digestion, it has been observed that the transcriptionally active ovalbumin gene is associated with the nuclear matrix in oviduct cells, but not with matrices from tissues where it is not transcribed (Ciejek *et al.*, 1983). This association of the ovalbumin gene with nuclear matrix includes the entire transcription unit with 2-4 kbp of 5'- and 3'-flanking non-transcribed sequences. In addition, the association of ovalbumin sequence with oviduct nuclear matrices is regulated by hormonal stimulation, since hormone withdrawal abolished ovalbumin gene expression and released its matrix association (Ciejek *et al.*, 1983; Robinson *et al.*, 1983).

One might suggest therefore that the apparent association of active genes with

nuclear matrix results from aggregation of nascent RNA and/or non-histone proteins (either RNA polymerses II or the DNA-binding proteins directly involved in transcription reaction) in the course of high-ionic strength salt extraction. However, a non-transcribed restriction fragment 3.8 kbp down stream from the 3' end of the ovalbumin gene is associated with the nuclear matrix, indicating that attachment does not require the presence of transcription complexes (Robinson *et al.*, 1983). This has been further supported by studies on the chicken vitellogenin II gene. This gene is transcribed in the liver of adult laying hens and estrogen stimulated immature chicks, and is associated with nuclear matrix in these cells, but not in tissues where it is transcriptionally silent (Jost & Seldan, 1984). Following primary stimulation of immature chicks, there is a lag period before vitellogenin mRNA is synthesized. However, association of the vitellogenin II gene with the nuclear matrix precedes the vitellogenin mRNA synthesis, indicating that transcription *per se* is not necessary for maintaining the association of active genes with the nuclear matrix.

In *Drosophila* cells it has been shown that transcriptionally active genes are nuclear matrix associated even when they are actually transcribed in less than 1 in 10 cells, suggesting that the anchorage of these sequences to the nuclear matrix was not a result of transcription (Small & Vogelstein, 1985).

In some cases such as *Drosophila* HSP70, and β-actin genes (Small *et al.*, 1985), and the chicken histone H5 gene (Dalton *et al.*, 1986), the matrix attached regions are restricted to the 5'-flanking sequences and/or the 5' end of the gene. The fact

that transcribed sequences downstream from these matrix sites are not matrix associated provides further support to the argument that the association of transcriptionally active genes with nuclear matrix is not merely due to aggregation or precipitation of transcription complexes in high-ionic strength salt.

Studies on a nucleoskeleton preparations from agarose-encapsulated HeLa cells under physiological ionic strength conditions also support this conclusion. After the bulk of the chromatin is removed by nuclease digestion, followed by electrophoresis, most of the nascent RNA and active RNA polymerase II remains associated with the residual nucleoskeleton (Jackson & Cook, 1985b). The fraction of DNA remaining associated with the nucleoskeleton is enriched in transcribing genes. Therefore, these interactions exist under physiological ionic strength conditions and are not artifacts resulting from high-ionic strength salt.

However, the nature of transcriptionally active DNA interactive with the nuclear matrix is far from settled. Several workers did not observe these interaction at all. The exact reason for this is not clear. However, the conditions used to isolate nuclear matrix in all these studies vary from one laboratory to another.

Laemmli and coworkers developed a low-ionic strength salt extraction procedure (see section 1.2.4) to prepare nuclear scaffold (Mirkovitch *et al.*, 1984). They identified specific sequences (MARs) involved in attachment to the nuclear matrix (scaffold). In *Drosophila* they identified at least 18 MARs near a variety of RNA polymerase II transcribed genes extending over 400 kbp of DNA. One such MAR

is found in the non-transcribed spacer in each repeat of the histone gene cluster, on a 657 bp restriction fragment. Exonuclease III digestion has localized two protein-binding domains on this MAR. Each covers roughly 200 bp and the two domains are separated by a nuclease accessible region of about 100 bp (Gasser & Laemmli, 1986a). The protected sequences presumably represent sites of interaction between the DNA and nuclear matrix proteins.

All MARs identified in the *Drosophila* system occur in non-transcribed regions, but for the mouse kappa light chain gene, a MAR is localized adjacent to the enhancer sequences in a transcribed region (Cockerill & Garrod, 1986b). Intragenic MARs have also been identified in the mouse immunoglobin heavy chain, the hamster dihydrofolate reductase gene (Kas & Chasin, 1987), and the human hypoxanthine-guanine phosphoribosyltransferase gene (Sykes *et al.*, 1988). The relationship of all these MARs identified in various genes to those revealed in high-ionic strength salt extracted nuclei is to be determined. It seems that these interactions may not depend on continuous transcription of neighbouring DNA sequences.

The nuclear matrix also appears to be a site for RNA processing. RNA splicing occurs within large RNP assemblies referred to as spliceosomes (reviewed by Ruby & Abelson, 1991). Small ribonucleoproteins (snRNPs) and hnRNPs are required for assembly of fully functional spliceosomes. When nuclear matrices are prepared without RNase digestion, snRNP hnRNP RNA, and proteins components are associated with them (Maundrell *et al.*, 1981; Van Eekelen *et al.*, 1981; Verheijen

et al., 1986). In support for a role of the nuclear matrix in RNA processing, pre-messenger RNA splicing, splicing intermediates, and splicing products remain associated with the nuclear matrix provided it is isolated under conditions that avoid protease and RNase activation. Pre-messenger RNAs are processed upon addition of *in vitro* splicing extracts without the lag time observed when exogenous deproteinized RNA is used as substrate (Zeitlin *et al.*, 1987, 1989). The absence of a lag period suggests that the isolated nuclear matrix contains precursor RNA already assembled into functional spliceosomes.

In summary, circumstantial evidences suggest that during mRNA production, processing, and transport, the nuclear matrix plays a key role.

1.2.6 Summary and statement of objectives

The nuclear matrix is an operationally defined insoluble structure obtained from isolated nuclei by sequential salt extraction, detergent and nuclease treatments. This structure, which is the same size and shape as the original nucleus, contains the nuclear lamina, remnants of nucleoli and an internal fibrogranular network.

Studies from a number of laboratories have led to the conclusion that the nuclear matrix is involved in chromatin organization, DNA replication, and RNA transcription, processing and transport. While all known nuclear functions are

associated with the nuclear matrix, our knowledge of the nuclear matrix itself is very limited. In contrast to our rather extensive knowledge of the lamins, the portions of the nuclear matrix within the lamina have not been well characterized. To date only a few proteins that associate with the internal matrix structure have been purified biochemically and functionally characterized. Most notable are topoisomerase II, matrin F, G and 3.

It is my thesis that a detailed molecular analysis of the individual nuclear matrix proteins is of prime importance for deciphering the structural organization and molecular properties of the nuclear matrix structure and its associated functions. In keeping with this thesis, the specific objectives which were addressed in this research were:

i) To isolate and characterize cDNA for a nuclear matrix protein from cDNA expression library.

ii) To determine the complete nucleotide sequence for isolated cDNA clones and determine if the predicted protein sequence is related to any other previously known protein sequences.

iii) To determine the genomic organization of the gene coding for the protein identified in this study.

It was believed that information thus generated would be useful in elucidating the role of nuclear matrix protein in nuclear organization and function.

CHAPTER 2

MATERIALS AND METHODS

2.1 MATERIALS

The chemical reagents were purchased from Sigma chemical company (Missouri, USA) unless stated otherwise. The oligo(dT)-primed λgt11 cDNA library constructed from human fibroblast cell poly $(A)^+$ RNA, and cDNA Insert Screening Amplimers (λgt11 Insert Screening Amplimers) were obtained from Clontech Laboratories Inc. (California, USA). Restriction endonucleases were purchased either from Bethesda Research Laboratories (Ontario, Canada) or New England Biolabs Ltd (Ontario, Canada). Bluescript plasmid vectors and the oligo(dT)-primed λZap II cDNA library made from HeLa cell poly(A)+RNA were obtained from Stratagene (California, USA). Sequenase kits were purchased from United States Biochemicals (Ohio, USA).

The tissue culture medium and fetal bovine serum were purchased from Flow Laboratories (Ontario, Canada). Geneclean kit, horseradish peroxidase (HRP) conjugated goat anti-rabbit and horseradish peroxidase conjugated goat anti-chicken antibodies were obtained from BIO/CAN SCIENTIFIC (Ontario, Canada). The M13

universal sequencing primer, the M13 reverse sequencing primer, the SK and KS primers were obtained from the University of Calgary oligonucleotide synthesis facility.

Mung bean nuclease, S1 nuclease, nucleotide triphosphates, and Taq DNA polymerase, RNAsin, Avian myeloblastosis virus (AMV) reverse transcriptase, and T4 polynucleotide kinase were obtained from Promega (Wisconsin, USA). The nylon membranes Hybond-N, nitrocellulose membrane Hybond-C, [α-^{32}-P]dCTP (3,000 Ci/mmol),γ-^{32}-P]dATP (3,000 Ci/mmol), [α-^{35}S]dATP (1350 Ci/mmol), and the Multiprime DNA Labelling System were purchased from Amersham (Ontario, Canada). Nitrocellulose membranes were purchased from Schleicher & Schuell (New Hampshire,USA). Oligo(dT)-cellulose (type 3) was from Collaborative Research Inc. (Massachusettes,USA) and Zetaprobe membrane was purchased from BIO-RAD Laboratories California, USA).

2.2 METHODS

2.2.1 Growth and maintenance of cells

The human cell lines used in this study were breast carcinoma cells MDA MB-330 (Calilleau *et al.*, 1980), colorectal adenocarcinoma SW-480 (Leibovitz *et al.*,

1976), T-cell leukemia Jurkat, colon adenocarcinoma SW-620 (Leibovitz *et al.*, 1976), cervical carcinoma HeLa S3 (Purk & Fisher, 1956), human epidermoid carcinoma A-431 (Giard *et al.*, 1973), ovarian adenocarcinoma HEY (Trent *et al.*, 1985), and lung carcinoma BEN (Martin *et al.*, 1987).

HeLa S3, HEY, BEN, MDA MB-330, and A-431 cells were grown as monolayers in Dulbecco's modified minimal essential medium supplemented with 10% fetal bovine serum, penicillin (50 units/ml), streptomycin (50 μg/ml), and $NaHCO_3$ (0.2%, v/v). SW-480 and SW-620 cells were grown in RPMI-1640 medium supplemented with 10% fetal bovine serum, penicillin (50 units/ml), streptomycin (50 μg/ml), and $NaHCO_3$ (0.2%, v/v). Jurkat cells were grown in suspension culture in RPMI-1620 supplemented with 10% fetal bovine serum, penicillin (50 units/ml), streptomycin (50 μg/ml), $NaHCO_3$ (0.2%, v/v), and 2 mM glutamine. The medium was sterilized by filtration through a sterile 0.22 micron filter (Millistack, Millipore). Addition of supplements was done under sterile conditions. Cells were grown in a humidified incubator at 37°C in an atmosphere of 95% air and 5% CO_2.

The continuity of the cell line was maintained by replating the cells when their confluency reached about 70-80%. At that time, cells growing in a 15-cm diameter tissue culture dish were trypsinized with 5 ml of a sterile solution containing Hank's balanced salts, trypsin (0.25%, w/v), 10 mM EDTA and 10 mM Hepes (pH 7.2) at 37°C for 5 min. The cells were replated onto new tissue culture dishes at a density

of approximately $2X10^5$ cells per ml of fresh medium. The cells were allowed to grow until 70-80% confluency was reached.

Stocks of cells were maintained in growth medium containing dimethyl sulfoxide (10%, v/v) under liquid nitrogen. Freshly trypsinized cells were pelleted (1000 x g for 3 min) and resuspended in growth medium at a concentration of about $5X10^6$ cells per ml. A one ml aliquot of this suspension was placed in a 2-ml cryo-vial and immersed in ice for 30 min. The vials then were stored at -70°C overnight before removing to a liquid nitrogen freezer. To revive a frozen stock, a vial of frozen cells was thawed immediately at 37°C and pelleted (1000 x g for 5 min). The pellet was resuspended in 2 ml of growth medium and plated onto a culture dish at a final concentration of about $2X10^5$ cells per ml of growth medium.

2.2.2 Bacterial cell culture

The experiments involving bacteriological techniques were performed using standard aseptic techniques. The bacterial cells, unless stated otherwise, were grown either in the liquid culture medium, Luria-Bertani (LB) medium (0.05% Bacto-yeast extract, 0.1% Bacto-trypton, 0.1% NaCl) containing 0.2% (w/v) glucose and appropriate antibiotic(s) or the same medium with 1.5% (w/v) Bacto-agar. The culture was always started with an isolated colony that was grown on a culture

medium plate containing appropriate antibiotic(s) to verify the bacterial strain.

2.2.3 Storage of bacteriophage stocks

To prepare stocks of bacteriophage, the plate lysate method (Maniatis et al., 1982) was used. The plate lysate was stored with a few drops of chloroform at 4°C or at -70°C with dimethylsulfoxide at a final concentration of 7% if storage for long periods was required.

2.2.4 SDS polyacrylamide gel electrophoresis

Protein fractions from HeLa cells were dissolved in sample buffer supplemented with DTT to disrupt the disulfide bonds and were boiled for 3 min. Proteins were separated by electrophoresis on 10% or 12% SDS polyacrylamide gel according to the procedure of Laemmli (1970).

To separate β gal-fusion protein from bacterial proteins, a 7.5% SDS polyacrylamide gel was run as above.

2.2.5 Immunoblotting

For immunoblotting, proteins resolved on SDS polyacrylamide gels were electrophoretically transferred to nitrocellulose membranes for 18-24 h at a constant current of 150 mA in transfer buffer (25 mM Tris, 192 mM glycine, 20% methanol, 0.05% SDS, pH 8.3) using a Hoefer Transblot apparatus (Hoefer Scientific Instruments, USA) following the instructions provided by the supplier. In later experiments, SDS polyacrylamide gel separated proteins were electrophoretically transferred to nitrocellulose membranes for 1 h using a Mini-transblot apparatus (BioRad) under the conditions recommended by the supplier. Transferred proteins were visualized by staining the nitrocellulose membranes in 0.2% Ponceau S, 3% trichloroacetic acid followed by destaining in Tris-buffered saline (TBS=50 mM Tris-HCl [pH 8.0], 150 mM NaCl). To block non-specific protein binding sites, blots were incubated in blocking buffer (10% fetal bovine serum [v/v] in Tris-buffered saline Tween-20 {TBST=50 mM Tris-HCl [pH 8.0], 150 mM NaCl, .05% Tween-20}) for 2 h at room temperature or 18 h at 4°C. Filters were transferred to sealable bags containing the primary antibody (diluted in blocking buffer). They were then incubated at 4°C for 16 h while agitating slowly on a shaker. At the end of the incubation period, the filters were washed 3 times for 10 min each with TBST and then 10 min with TBS. Filters were then transferred to blocking solution containing HRP-linked second antibody (1:500 diluted) and were incubated at room temperature

for 1 h. Filters were washed as above and then incubated in peroxidase substrate solution prepared by mixing 2 ml of a solution of 4-chloro-1-naphthol in methanol (3 mg/ml) with 10 ml of TBS containing 0.1M imidazole. Just before adding the filter, 5 μl of 30% hydrogen peroxide was added. After color development (15-30 min at room temperature), filters were washed with two changes of distilled water and allowed to dry between Whatman #1 filter papers. Filters were stored protected from light.

2.2.6 Restriction endonuclease digestion of DNA

Digestion of DNA with various restriction endonucleases was carried out using buffer provided by the supplier.

2.2.7 Agarose gel electrophoresis

Agarose gel electrophoresis was performed according to the instructions provided by Maniatis *et al.*, (1982) using either a submerged gel electrophoresis apparatus, BRL Model H4, 20 x 25-cm gel bed or H3, 11 x 14-cm gel bed or BRL Mini-gel, 7 x 10 cm-gel bed. A 0.8-1.2% agarose gel depending upon the size of DNA

fragment to be analyzed was prepared in Tris-acetate-EDTA buffer (TAE=4mM Tris-acetate, 2 mM disodium EDTA pH 8) containing ethidium bromide (final concentration, 0.5 μg/ml). The gel was also submerged in TAE buffer containing 0.5 μg/ml ethidium bromide. The DNA samples were mixed with gel loading dye (0.25% bromophenol blue and xylene cyanol in 15% Ficoll, type 400), loaded in the well and electrophoresed at a constant voltage of about 5 V/cm. Electrophoresis was continued until the bromophenol dye travelled to about 1 cm from the edge of the gel. The positions of the DNA bands were revealed by exposing the gels to UV light to produce ethidium bromide fluorescence. While still exposed to UV light the gel was photographed using Polaroid film 665 or 667.

2.2.8 Recovery of DNA fragments from agarose gels

Purification of DNA fragments from the agarose gel was performed by using a Geneclean kit (BIO/CAN SCIENTIFIC) according to the instructions provided with the kit.

2.2.9 Labelling of DNA probes

Double-stranded DNA was labelled by a random primer method (Fienberg and

Vogelstein, 1983, 1984) using a multiprime DNA labelling system kit obtained from Amersham. The reaction was performed according to the manufacture's instruction and using the reagents provided in the kit.

2.2.10 Preparation of nuclear matrix

The nuclear matrix was prepared from HeLa cells by one of the following methods.

(i) The following method is a modification of a previously published method (Younghusband and Maundrell, 1982).

Exponentially growing monolayer HeLa cells were washed twice on ice with phosphate buffered saline (PBS=0.14 M NaCl, 8 mM Na_2HPO_4, 1.5 mM KH_2PO_4, 2.7 mM KCl [pH 7.4]), detached with a cell scraper and collected by centrifugation at 800 x g for 5 min at 4° C. Cell pellets were suspended in ice-cold nuclear wash buffer (0.25 M sucrose, 10 mM Tris-Cl [pH 7.4], 10 mM NaCl, 5 mM $MgCl_2$, 1 mM phenylmethylsulfonylfluoride [PMSF] containing 0.5% Nonidet P-40 (v/v) on ice for 15 min. The cells were disrupted in a Dounce homogenizer, and lysis was monitored by phase-contrast microscopy. Nuclei were pelleted by centrifugation at 800 x g for 5 min at 4°C and the supernatant was saved as cytoplasmic fraction. The pelleted nuclei were resuspended in nuclear wash buffer at a density of 10^8 nuclei/ml. The concentration of NaCl was increased to 2 M by the addition of 0.67 volume of

5 M NaCl, and DNase I was added to 50 μg/ml. The mixture was incubated at room temperature for 30 min, EDTA was added to 10 mM (final concentration), and the resulting mixture was centrifuged at 10,000 x g for 15 min. The pellet obtained was termed the nuclear matrix and it was stored at -20° C until needed.

(ii) The following method is a modification of a previously published method (Van Eekelen & Van Vanrooij, 1981).

HeLa cell nuclei prepared as above were resuspended in buffer A (10 mM Tris-Cl [pH 7.4], 110 mM NaCl, 1.5 mM $MgCl_2$, 1 mM PMSF) at a density of 10^8 nuclei/ml, and were treated with DNase I (250 μg/ml final concentration) for 30 min at room temperature. An equal volume of 0.8 M $(NH_4)_2SO_4$ in Tris buffer (10 mM Tris-HCl, [pH 7.4], 1.5 mM $MgCl_2$) was slowly added. The mixture was centrifuged at 10,000 x g for 15 min at 4°C. The nuclear matrix pellet was washed once with buffer A, centrifuged again as above, and resupended in sample buffer for SDS polyacrylamide gel electrophoresis.

2.2.11 Estimation of phage titre

The titre of the phage in the cDNA library or in a clear lysate was estimated by a plaque assay method. Ten fold serial dilutions of the phage suspension were made by diluting in SM medium (100 mM NaCl, 50 mM Tris pH 7.5, 10 mM $MgSO_4$, 0.01% gelatin). An aliquot (100-200 μl) of an overnight culture of bacterial host

strain (*E.coli* strains Y1090 for λgt11 and XL-1 blue for λZap II) was mixed with soft agarose (7.5% agarose in Luria-Bertani [LB] medium) containing 0.2% maltose, poured on top of plates containing LB-agar (0.2% maltose) and allowed to set for a few min. An aliquot (5-10 μl) of each of the phage dilutions was then spotted on the agarose, spread to cover a small area about (1 cm diameter) and allowed to adsorb completely by keeping the culture plate at room temperature (with the top cover slightly open) for about 15 min. The plates were then incubated at 42°C to allow for the phage growth and appearance of the plaques (6-8 h). The number of the plaques was counted and used to calculate the plaque forming units (p.f.u)/ml in the original phage suspension.

2.2.12 Antibody screening of the λgt11 cDNA library

The following protocol for screening a λgt11 cDNA expression library is a modification of a previously described method (Huynh et al., 1985) Approximately $5x10^4$ recombinant phages were plated per 150-mm plate on mid-log Y1090 in soft agar on LB/ampicillin plates (pH 7.5) at 43°C for 3 h (until tiny plaques were just visible). The plates were placed at room temperature and an isopropyl β-D-thiogalctoside (IPTG) soaked and dried nitrocellulose filter circle was placed on each plate (filters were prepared by soaking in 10 mM IPTG for 1 h, blotting dry, and allowing to dry 30-60 min at room temperature). The plates were quickly transferred

to a 37° incubator and were incubated for 2-3 h. After the end of the incubation period, before filters were removed from the plates, their position was marked by making three asymmetrically placed holes through the filter and into the agar with a 21-gauge needle on a syringe filled with India ink. The filters were carefully peeled off and processed.

The filters were processed by rinsing a few times in TBST to removfragments of agarose and cell debris. The filters were stored wet, wrapped in an aluminum foil, at 4°, until a second set of filters treated in the same way was ready for processing. The phage plates were sealed with parafilm stored inverted at 4° until needed. The filters that were stored at 4°, were removed to room temperature and incubated (1 h) in blocking buffer (10% fetal calf serum in TBST) to block free protein binding sites. The filters were then incubated in a petri plate containing primary antibody previously treated to remove anti-*E.coli* antibodies (see next section for details), at a dilution of 1:1000 in TBS containing 10% fetal calf serum. The plates containing filters were incubated at 4° for 16 h while agitated slowly on a shaker. The plates were removed to room temperature and the antibody solution was decanted into a clean tube and stored at 4°C for reuse (for a total of three or four times). The filters were washed 3 times 10 min each with TBST and then 10 min with TBS. The filters were transferred to blocking solution containing a biotinylated second antibody (40 μl/10 ml/filter) and were incubated at room temperature for 1 h with gentle agitation on a shaker. The filters were washed as before and then

transferred to TBST containing avidin-biotin horseradish peroxidase complex (preincubated at room temperature for 30 min) and were incubated at room temperature for 30 min with gentle shaking. The filters were washed in TBS 3 times, 5 min each time and then they were incubated in peroxidase substrate solution, which was prepared by mixing 2 ml of a solution of 4-chloro-1-naphthol in methanol (3 mg/ml) with 10 ml of TBS containing 0.1M imidazole. Just before adding the filter, 5 μl of 30% hydrogen peroxide was added. After color development (15-30 min at room temperature), filters were washed with two changes of distilled water and allowed to dry between Whatman #1 filter papers. The filters were stored protected from light, until they were used to locate the areas on the phage plate that corresponded to a positive signal. A plug of agar was removed from this area of the plate using the large end of a sterile Pasteur pipette. The agar plug was transferred to 1 ml of SM plus two drops of chloroform, and the phage were eluted at 4°C overnight. Phage eluents were titered as described above and rescreened using 10-fold fewer phage per plate during the subsequent round. This was repeated until a homogenous population of recombinant phage was obtained, and this was usually achieved after 2-3 rounds of rescreening.

2.2.13 Adsorption of anti-*E.coli* from antiserum

Removal of anti-*E.coli* was done by applying antiserum to a nitrocellulose filter

blotted on a confluent lawn of wild type λgt11 plaques. About 10^4 phage were plated with 10^8 *E.coli* Y1090 cells on a 90-mm LB agar plate and incubated at 43°C for 12 h. The plate was removed to room temperature and an 85-mm nitrocellulose membrane disc was overlaid and the plate was incubated at 37°C for a further 1 h. The membrane was then flipped over on the same agar plate and incubated for an additional 1 h. At the end of the incubation period, the membrane was peeled off and washed in TBS for 10 min, followed by incubation in primary antibody (rabbit or chicken antibody diluted 1:50 in TBST containing 10% fetal calf serum) at room temperature for 1 h. The filter was discarded and the pre-adsorbed antibody was then diluted suitably to achieve the final concentration used in screening (section 2.2.12).

2.2.14 Screening recombinant plaques by hybridization

All the procedures for plaque hybridization were adopted from Maniatis *et al.* (1982). Approximately 1-5x10^4 recombinant phages were plated per 150 mm plates on mid-log host bacteria in soft agar on LB/ampicillin plates (pH 7.5) at 43° C for 12-18 h (until plaques were about 1-2 mm in diameter). The plates were chilled for 1-2 h at 4°C and replica filters (Hybond-N, nylon membranes that were marked asymmetrically at four positions) were made from the plates. While making the replica filters, the marks from the filter were copied on to the plates. The filters were then dried at room temperature for 1 h and treated sequentially with alkali (0.2

N NaOH and 1.5 M NaCl) for 1 min, Tris (0.5 M Tris-HCl, pH 7.5, 1.5 M NaCl) for 1 min, and 5 x SSC (1 x SSC = 150 mM NaCl, 15 mM sodium citrate) for 2 min. The filters were then dried for about 1 h and the DNA was fixed on the filters by exposing them to UV light for 2 min.

For prehybridization, the filters were incubated in a prehybridization solution containing 5 x SSPE (1 x SSPE= 180 mM NaCl, 20 mM sodium phosphate [pH 7.7], 1 mM EDTA), 1% SDS, 5 x Denhardt's solution (1 x Denhardt's solution = 0.02% bovine serum albumin [BSA], 0.02% ficoll, 0.02% polyvinylprolidone), 50% formamide (deionized), and 100 μg/ml sonicated denatured salmon sperm DNA. The filters were incubated at 44°C for 4-6 h and transferred to hybridization solution which contained pre-hybridization solution plus random primed radioactive-labelled probe. The probe was heat-denatured before adding to the hybridization solution. The hybridization was done for 18 h at 44°C. Filters were washed in 6 x SSC, 0.1% SDS twice for 15 min at room temperature, in 1 x SSC, 0.5% SDS twice for 15 min at 37°C, and in 0.1 x SSC, 1% SDS twice for 30 min at 60°C. Filters were then wrapped in Saran wrap and exposed to X-ray film (Kodak X-OmatRP) at -70°C with two intensifying screens. Plaques showing hybridization signal on duplicate filters were identified on their corresponding master culture plates by superimposing the autoradiograms on the culture plates, making use of the orientation marks on the plates, filters and films. The plaques or area showing positive hybridization signals were removed, titred, and screened until the membranes showed 100% positive

plaques.

2.2.15 Isolation of λgt11 DNA

DNA from λgt11 cDNA clones was isolated by the large-scale method of λ preparation (Maniatis *et al.*, 1982).

2.2.16 Preparation of lysogens

To obtain preparative amounts of the β gal-fusion proteins, it was necessary to express the λgt11 recombinant as a lysogen in *E.coli* Y1089 (Huynh *et al.*, 1985). *E.coli* Y1089 was grown in LB medium containing 0.4% maltose and 50 μg/ml ampicillin overnight at 37°C with shaking. Cells were diluted 1 in 50 and grown to an optical density at 600 nm of 0.5. They were next supplemented with 10 mM $MgCl_2$ and stored on ice. Phage was mixed with cells to infect at multiplicity of about 5 and incubated at 32°C for 20 min. Infected cells were inoculated onto LB plates containing ampicillin (50 μg/ml) and incubated over night at 32°C. Single colonies were plated onto two plates, one was incubated at 32°C and another at 42°C. Colonies growing at 32°C but not at 42°C were selected as lysogens.

2.2.17 Isolation of β gal-fusion protein

The β gal-fusion proteins were obtained by inoculating 100 ml LB medium with a single colony of the lysogen and incubated with shaking at 32°C until the optical density at 600 nm reached 0.5. The β gal-fusion protein was induced at 44°C for 20 min and IPTG was added to 10 mM (final concentration). The culture was incubated at 38°C for another 2 h before the cells were harvested by centrifugation (10,000 x g, at room temperature for 15 min) and were resuspended in 1/10 volume of Tris-EDTA PMSF (TEP = 100 mM Tris-HCl [pH 7.4], 10 mM EDTA, and 1 mM PMSF). Cell disruption was completed by freezing in liquid nitrogen and thawing at room temperature. This was repeated once more and the resulting lysate was centrifuged for 10 min at 10,000 x g at 4°C. The supernatant was stored at -70°C until use.

2.2.18 Purification of 7E β gal-fusion protein

The following method for purification of 7E β gal-fusion protein by affinity chromatography is a modification of a previously described method (Germino *et al.*, 1983).

The supernatant containing 7E β gal-fusion protein (section 2.2.17) was treated with three volumes of saturated ammonium sulphate and centrifuged for 20 min at

10,000 x g at 4°C. The pellet was dissolved in cold Tris buffer (50 mM Tris-Cl [pH 7.5], 100 mM NaCl, 10 mM $MgCl_2$, 1 mM EDTA, 10 mM β-mercaptoethanol) at a concentration of about 20 mg/ml. The sample was then loaded on to a column containing a β gal affinity matrix (p-aminobenzyl-1-thio-D-galactopyranoside-agarose, pre-equilibrated with Tris buffer). The flow through fractions were collected and loaded on the column once again. The column was washed with Tris buffer (about 100 bed volumes) until no more material absorbing at OD_{280} nm appeared in the flow through fractions. The 7E β gal-fusion protein was eluted from the column with 100 mM sodium borate, 10 mM β-mercaptoethanol, pH 10.5. The eluent was collected in tubes containing 300 μl 100 mM Tris-Cl (pH 6.8), and stored at -20°C until use.

2.2.19 Preparation of rabbit antiserum

Preimmune serum from the rabbit was collected, saved, and checked to ensure that it gave a low background as assayed by immunoblot analysis. For the preparation of antiserum, HeLa cell nuclear matrix preparation (about 500 μg of total nuclear matrix proteins) was emulsified with an equal volume of Freund's complete adjuvant. The emulsion (3.0 ml) was then injected at 10 sites (subcutaneously and intramuscularly) on the back of a New Zealand white rabbit. After primary injection, the rabbit was boosted with the same antigen preparation in Freund's incomplete adjuvant, 14 and 28 days after primary injection. Serum was collected on days 24

and 38 and assayed by immunoblot analysis.

For the preparation of antiserum against β gal-fusion protein, the lysate prepared from induced lysogen culture was subjected to a 7.5% SDS polyacrylamide gel electrophoresis. The gel was stained with Commassie blue dye, and the band corresponding to β gal-fusion protein was excised. The excised band containing about 200-250 μg fusion protein was rinsed in deionized water and homogenized with an equal volume of PBS by passing through successively smaller gauge syringe needles. The resulting mixture was emulsified as above and used as an immunogen. The rabbit was boosted twice with a 14 days interval between booster injections, and antiserum collected on day 14 and 38 was tested by immunoblot analysis. The same immunization procedure was followed for the nitrocellulose filter membrane blotted β gal-fusion protein, except the filter strip containing fusion protein was dissolved in a small volume (200-400 μl) of dimethyl sulfoxide, prior to emulsification.

For the preparation of antiserum against affinity purified β gal-fusion protein, the immunization procedure described for nuclear matrix protein was followed.

2.2.20 Affinity purification of antibodies

Antibodies directed against the β gal-fusion proteins were affinity purified by specific adsorption to the corresponding β gal-fusion protein on nitrocellulose membranes following the method of Olmsted (1981). Approximately 200-300 μl of

lysate (section 2.2.17) containing β gal-fusion protein was subjected to 7.5% SDS polyacrylamide gel electrophoresis. After electrophoresis, the proteins were transferred to a sheet of nitrocellulose membrane as described in section 2.2.5. Strips of nitrocellulose membrane were cut from the left and right edges of the sheet and immunoblotted with antibodies against β gal-fusion protein as described in section 2.2.25. Using the position of the β gal-fusion protein on these edge strips as guides, a horizontal strip of nitrocellulose membrane, corresponding to β gal-fusion protein band was excised from the preparative nitrocellulose sheet. The strip was washed twice in TBST, and incubated overnight at 4°C in primary antibody (CRLM or rabbit antiserum) diluted 1:50 in TBST/3% BSA. The strip was then washed five times for 10 min each in TBST, loaded into a 5-ml syringe , and washed again by forcing two 5 ml aliquots of TBST through the filter. 1 ml of 200 mM glycine (pH 2.6)/ 0.05% Tween-20 was drawn up into the syringe through the filter, incubated 5 min at room temperature, expelled through the filter, drawn up again, then expelled into 1 ml of 100 mM Tris base. To this mixture 100 μl of 3M NaCl, 100 μl of 1% Tween 20, 400 μl of new born calf serum and 1.4 ml of TBST were added. This antibody was stored at 4°C and used directly in immunoblotting after diluting 1:4 in blocking buffer.

2.2.21 Subcloning

Purified phage DNA was digested with *Eco*RI to release the cDNA inserts. The *Eco*RI fragments were isolated from agarose gels using Geneclean (Section 2.2.8) and ligated into the *Eco*RI site of the plasmid Bluescript SK+. After transformation of competent XL-1 blue cells, single clones were isolated and grown in liquid culture. DNA was purified by the method described below.

In vivo excision of the inserts from the λZap II vector to a bluescript SK environment was performed according to the manufacturer's instructions (Stratagene).

2.2.22 Plasmid preparation

2.2.22.1 Large scale preparation of plasmid DNA

Large preparations of plasmid DNA was made by centrifugation in CsCl/ethidium bromide gradients after cell lysis (Maniatis *et al.*, 1982).

2.2.22.2 Mini-preparation of plasmid DNA

The following rapid method for isolation of plasmid DNA is a modification of a previously described method (Birnboim, 1983). An isolated colony with the appropriate plasmid was added to 5 ml of LB medium with the appropriate antibiotic

and incubated at 37°C with vigorous shaking for about 18 h. About 1.5 ml of an overnight culture was centrifuged for 30 sec in a microfuge tube and the pellet obtained was resuspended in 0.1 ml solution I (50 mM glucose, 25 mM Tris-Cl [pH 8.0], 10 mM EDTA) by vortexing for a few seconds. After storage on ice for 10 min, 0.2 ml solution II (0.2 M NaOH, 1% SDS) was added and the contents of the tube mixed thoroughly by inverting the tube several times, it was then left on ice for 5 min and 0.15 ml 3 M potassium acetate solution, pH 4.8, was added. The contents of the tube were mixed as before and the tube was incubated on ice for 15 min with occasional shaking to allow for precipitation to occur. The solution was centrifuged in an Eppendorf centrifuge for 10 min at 4°C and the supernatant was collected in a clean microfuge tube. Two volumes of absolute ethanol were added, the contents of the tube mixed well by inverting several times and precipitation of plasmid DNA was allowed to occur by incubation of the tube at -70°C for 10 min. The precipitated plasmid DNA was collected by centrifugation in an Eppendorf centrifuge for 15 min at 4°C. The pellet containing the plasmid DNA was washed twice in 70% ethanol, dried under vacuum, dissolved in 20-40 μl of TE and stored at 4°C.

2.2.23 Nucleotide sequencing

The nucleotide sequence of the cDNAs was determined by Sanger's dideoxy nucleotide sequencing method (Sanger *et al.*, 1977). The *Eco*RI generated fragments

were cloned into the plasmid vector Bluescript-SK and their nucleotide sequences were determined using single- or double-stranded plasmid DNA with or without Exonuclease III deletion. Various sequencing primers including the M13 universal primer, the M13 reverse sequencing primer, and the SK and KS primers were used for sequencing. In some cases nucleotide sequences were determined directly from λgt11 DNA using λgt11 sequencing primers. The sequencing reactions were performed using the instructions provided in the manual for the Sequenase kit.

2.2.23.1 Exonuclease III deletion

Unidirectional sets of deletions were created in DNA inserts by partial digestion with Exonuclease III and S1 nuclease (Henikoff, 1987) using the protocol provided by Promega.

2.2.23.2 Isolation and purification of single-stranded plasmid DNA

Single-stranded plasmid DNA was isolated and purified by the polyethylene glycol method (Dente *et al.*, 1983) described in the *Bluescript Exo/Mung DNA sequencing system* instruction manual provided by Stratagene.

2.2.23.3 Denaturation of the double-stranded plasmid for sequencing

Double-stranded plasmid DNA was subjected to denaturation with NaOH (Hattori and Sakaki, 1986) prior to sequencing. About 5 μg of double-stranded plasmid DNA

(10 μl) was mixed with an equal volume of 0.4 M NaOH containing 2 mM EDTA and the resulting mixture was incubated at room temperature for 5 min. The denatured DNA was precipitated by the addition of 0.8 volume of 5 M ammonium acetate pH 7.5 and 4 volumes pre-chilled absolute ethanol and transferred to a -70°C freezer for 15 min. The precipitate was collected by centrifugation in an Eppendorf centrifuge for 15 min at 4°C, dried under vacuum and dissolved in 7 μl deionized water.

2.2.23.4 Sequencing gel

Sequencing gels 0.35 mm x 380 mm x 500 mm containing 6% polyacrylamide and 7 M urea prepared in Tris-borate-EDTA buffer (1 M Tris and Boric acid, 0.4 M EDTA pH 8.3) were pre-run for about 30 min at a constant power of about 60 Watts in a BioRad sequencing apparatus with Tris-borate-EDTA as the running buffer. Samples for analysis were denatured by heating at 80°C for 3 min and then chilled on ice immediately before transferring about 2 μl of each to wells formed using a shark's tooth comb. Electrophoresis was usually performed until the bromophenol blue dye had reached the bottom of the gel (about 2.5 h). To read more than 150 nucleotides, multiple loadings of the same sequencing reaction were performed. After the electrophoresis was completed, the gel was fixed in 1 L of a fixing solution (5% acetic acid/5% methanol) for about 30 min to remove urea from the gel. The gel was dried at 90°C under vacuum using a gel dryer (BioRad-model 583) for about

2 h and then exposed to Kodak X-Omat AR film for 12-18 h. The nucleotide sequences were read manually and analyzed using a computer assisted PC/GENE program (IntelliGenetics, USA)

2.2.24 Polymerase chain reaction (PCR) amplification of λcDNA insert

For the analysis of cDNA inserts, PCR was used to amplify the cDNA directly from isolated plaques. A single isolated λgt11 or λZap II plaque was picked from the culture plate and resuspended in 250 μl of deionized water. To release the λ DNA from the phage, the resuspended plaque was frozen at -70°C for 15 min and then thawed. 25 μl of this lysate was used for PCR amplification. For λgt11 clones, amplification was performed according to instructions provided by the manufacturer (Clontech), using the cDNA Insert screening kit (λgt11 Insert Screening Amplimers). For amplification of cDNA inserts from λZap II, the M13 forward and reverse primers were used. Thirty to 40 cycles of amplification (30 sec at 94°C, 15 sec at 60°C, and 1 min at 72°C) was carried out in a 50 μl reaction buffer (50 mM KCl, 10 mM Tris-Cl (pH 8.3), 1.5 mM $MgCl_2$, deoxy nucleoside triphosphate 0.2 mM each, 20 pM primers 20 pmol each, 0.01% gelatin, 1.25 Units Taq DNA polymerase) using a Perkin Elmer thermal cycler. The PCR product was analyzed by agarose gel

electrophoresis.

2.2.25 Northern blot analysis

2.2.25.1 RNA isolation

Total RNA from HeLa cell monolayers was isolated by the acid guanidinium thiocyanate-phenol-chloroform extraction method described by Chomezynski and Sacchai (1987), with some modifications. For extraction of total RNA, HeLa cells were grown to 80-90% confluency and the growth medium was removed from the cells, and 2.4 ml of 4 M guanidinium thiocyanate, 25 mM sodium citrate, pH 7.0, 0.5% sarcosyl, and 0.1 M β-mercaptoethanol were added to the cells. The cells were scraped off the plates and extracted sequentially with 0.1 volume of 2 M sodium acetate, pH 4.0, an equal volume of water saturated phenol, and 0.2 volume of chloroform/isoamyl alcohol (48:1, v/v). The final suspension was shaken vigorously for 10 sec and cooled on ice for 15 min. The samples were centrifuged at 10,000 x g for 20 min at 4°C. After centrifugation, the aqueous phase was transferred to a fresh tube and mixed with an equal volume of isopropanol. This was placed at -20°C for 1 h and then centrifuged at 10,000 x g for 30 min at 4°C. The RNA pellet was dissolved in 500 μl of guanidinium thiocyanate solution and transferred to a 1.5-ml Eppendorf tube pretreated with diethyl pyrocarbonate (DEPC). 500 μl of isopropanol was added, and the mixture was placed at -20°C for 1 h. After centrifugation for 10

min at 4°C, the RNA pellet was washed with ice-cold absolute ethanol, vacuum-dried, and dissolved in 300 μl of DEPC-treated deionized water. The RNA was precipitated twice with 0.1 volume of 3M sodium acetate, pH 5.2, and 2.2 volumes of absolute ethanol. The pellet was dissolved in DEPC-treated water. The pellet obtained after the second ethanol precipitation, was resuspended in 70% ethanol, and centrifuged at 10,000 x g for 15 min at 4°C. The washed pellet of RNA was dried under vacuum, dissolved in DEPC-treated water, subjected to spectrophotometric measurement and one more round of ethanol precipitation. The resulting RNA solution was then stored at -70°C until oligo(dT)-cellulose chromatography was performed as described in next section.

2.2.25.2 Oligo(dT)-cellulose chromatography

About 0.5 g oligo(dT)-cellulose (type 3) was resuspended in DEPC-treated water, poured into a small column (10 ml, Econo column, BioRad) and activated by passing about 10 column volumes of 0.1 N NaOH through the column. This was followed by washing the column with 10 or more column volumes of water until the pH of the effluent dropped to neutral. The column was then equilibrated by passing through about 10 column volumes of binding buffer (25 mM Tris-HCl [pH 7.4], 0.5 M NaCl containing 0.5% SDS (w/v). The RNA in binding buffer was heated to 65°C for 5 min and cooled to room temperature. The denatured RNA was then passed through the equilibrated oligo(dT)-cellulose column. Unbound material from the first passage

of RNA through the column was applied once more to achieve maximum binding of the poly(A)$^+$ RNA. The column was then washed with binding buffer until the OD_{260} of the eluent reached basal levels. The RNA was then eluted in 3 ml DEPC-treated water, subjected to ethanol precipitation, and the precipitate was collected by centrifugation. The precipitate was dried under vacuum and dissolved in DEPC-treated water and stored at -70°C until use.

2.2.25.3 Analysis of RNA

Poly(A)$^+$ RNA (~5 μg) was dissolved in 11 μl DEPC-treated water and the following added; 2 μl of 10 x running buffer {1 x running buffer = 40 mM MOPS [3-N-Morpholinopropane-sulphonic acid] (pH 7.0), 10 mM sodium acetate, 1 mM EDTA (pH 8.0)}, 7 μl 37% formaldehyde, and 20 μl formamide (pH 7.0), to give a final volume of 40 μl. After incubation at 60°C for 15 min, 4 μl of sample buffer (50% glycerol, 1 mM EDTA, 0.4% bromophenol blue) was added. RNA was fractionated by electrophoresis through 1% agarose gel containing 1 x running buffer, 2.2 M formaldehyde, and 0.2 μg/ml ethidium bromide. Gels were run at constant voltage (5 V/cm) for about 4 h in 1 x running buffer. Following electrophoresis, the RNA was visualized by UV transillumination and photographed. The gel was rinsed briefly in transfer buffer (20 x SSPE) and blotted onto a nylon membrane (Hybond-N) for about 18 h as described for Southern blotting (section 2.2.26) but without any further treatment of the gel. Transfer of RNA was monitored by restaining the gel

in ethidium bromide followed by visualization under UV light. The membrane with the transferred RNA was air dried and baked at 80°C for 2 h to fix the RNA on membrane. The blot was then stored in a sealed plastic bag until required for hybridization.

Filters were pre-hybridized with 0.25 ml/cm^2 of 50% formamide (pH 7.4), 5 x Denhardt's solution, 5 x SSPE, 0.1% SDS, and sonicated denatured salmon sperm DNA (100 μg/ml) at 44°C for 4 h. Filters were hybridized with 0.1 ml/cm^2 of 50% formamide (pH 7.4), 1 x Denhardt's solution, 5 x SSPE, 0.1% SDS, sonicated denatured salmon sperm DNA (100 μg/ml) plus denatured random primed radioactive-labelled probe DNA (2-4 x 10^6 cpm/ml), at 42°C for 24 h. The filters were washed and exposed to X-ray film as for DNA blots.

2.2.26 Primer extension analysis

Primer p31, GGGCATTTGTGTTGCAGGAG (+6 to -14) was end-labelled with [γ-^{32}P]dATP (3,000Ci/mmol) according to the method described by Sambrook *et al.*, (1989). Unincorporated [γ-^{32}P]dATP was removed following ethanol precipitation of the primer (Sambrook *et al.*, 1989). The primer had a specific activity of 2.2 x 10^7 cpm/10 pmoles. 5 μg poly(A)$^+$ RNA or 50 μg of total RNA from HeLa cells were co-precipitated with 1 x 10^6 cpm of the primer p31. The pellet was resuspended in 15 μl of hybridization buffer (150 mM KCl, 10 mM Tris.Cl [pH

8.3], 1 mM EDTA), denatured at 65° C for 1 h, and slowly cooled to 40°C. To this, 30 μl of reverse transcription mixture (20 mM Tris.Cl [pH 8.3], 10 mM $MgCl_2$, 5 mM DTT, 0.01% actinomycin D, 150 μM each of the 4 dNTPs, 25 units RNasin, 7 units AMV reverse transcriptase) was added. The reaction mixture was incubated at 42°C for 1 hr. The product was treated with RNase A at room temperature and recovered by phenol:chloroform extraction followed by ethanol precipitation. The reaction samples were resuspended in 10 μl of formamide loading buffer (95% formamide, 0.05% [w/v] bromophenol blue, 0.05%[w/v] xylene cyanol, 20 mM EDTA) and denatured at 80°C for 3 min before loading (2 μl/lane). Single stranded M13 DNA, sequenced by the method described above (section 2.2.23), was used as size marker. The extension products and the sequencing reactions were electrophoresed side by side on a 6% denaturing polyacrylamide sequencing gel. The gel was washed for 30 min in gel fixing solution (5% acetic acid, 5% methanol), dried and autoradiographed using a X-ray film (Kodak X-OmatRP).

2.2.27 Genomic Southern blot analysis

The human genomic DNA used in this experiment was a generous gift from Dr. Roger Green. DNA was digested with restriction endonuclease in the digestion buffer recommended by the suppliers. Digestion products were separated by electrophoresis through 0.8% agarose gels in Tris-acetate-EDTA (TAE) buffer [40

mM Tris-HCl (pH 7.9), 5 mM Na_3 acetate, 1 mM EDTA] containing 0.5 μg/ml ethidium bromide. Electrophoresis was for 16 h at 30 V and gels were photographed under short wavelength UV light on a transilluminator.

DNA was transferred to nylon filters (Hybond-N) by the blotting method of Southern (1975). First, the DNA was partially depurinated by soaking gels in 0.25 N HCl for 10-15 min, then denatured by soaking in 0.5 M NaOH, 1.5 M NaCl twice for 15 min each. DNA was blotted onto filters in 10 x SSC (1 x SSC = 0.15 M NaCl, 0.015 M sodium citrate) for 12-16 h, and fixed by baking for 2 h at 80 in a vacuum oven.

Filters were pre-hybridized with 0.25 ml/cm^2 of 6 x SSC, 5 x Denhardt's solution, 1% SDS, 100 μg /ml sonicated denatured salmon sperm DNA and 10 μg/ml poly-A at 44°C for 24 h. Filters were then hybridized with 0.1 ml/cm^2 of 50% formamide (pH 7.4), 6 x SSC, 1% SDS, 100 μg/ml sonicated denatured salmon DNA, and denatured labelled probe DNA (2-4 x 10^6 cpm/ml), at 44°C for 24 h. Filters were washed in 6 x SSC, twice for 15 min at room temperature, in 2 x SSC, 0.1% SDS for 30 min at 65°C, and in 0.2 x SSC, for 10 min at 68°C. Filters were then wrapped in Saran wrap and exposed to a X-ray film (Kodak X-OmatRP).

CHAPTER 3

RESULTS

3.1 ISOLATION AND CHARACTERIZATION OF cDNA USING RABBIT ANTISERUM AGAINST HeLa CELL NUCLEAR MATRIX

The primary goal of this research as described in section 1.2.6 was to isolate a cDNA from an expression library and characterize a clone specific for a nuclear matrix protein. Two types of antibodies were used in the screening of cDNA libraries. A rabbit polyclonal serum raised against HeLa cell total nuclear matrix and the screening of a λgt11 cDNA library constructed from human fibroblast RNA is described in this section. Another antibody, specific to rat liver nuclear matrix proteins, was obtained from Dr. H.C. Smith of Rochester University. This antibody was also used to screen the same cDNA library and this study is described in section 3.2.

3.1.1 Preparation of nuclear matrix

Nuclear matrix was prepared from HeLa cells using the method described in section 2.2.10 (method i). As a first step in the preparation of the nuclear matrix, HeLa cells were Dounce homogenized in buffer containing detergent (NP40) to disrupt the cell membrane and to release nuclei. The lysis of cells was monitored by phase contrast microscopy, and nuclei were obtained essentially free of cytoplasmic contamination. The nuclei were subsequently treated with DNase I to degrade DNA as completely as possible. The digested nuclei were then extracted with 2 M NaCl (final concentration) to extract most of the DNA, histones and other chromosomal proteins. The mixture was centrifuged and the resulting pellet was nuclear matrix.

3.1.2 Generation of polyclonal antiserum

For the production of a high titre antiserum, one rabbit was immunized with whole nuclear matrix preparation. The rabbit was injected on days 1, 15, and 29, with about 500 μg of total nuclear matrix proteins. For the first injection, Freund's complete adjuvant was used; for subsequent injections, Freund's incomplete adjuvant was used. Injections were given at multiple sites: subcutaneous and intramuscular. The rabbit was bled from the ear on days 0, 24, and 38. The specificity of the antiserum obtained at days 24 and 38 was assessed by probing a western blot as

described in section 2.25.

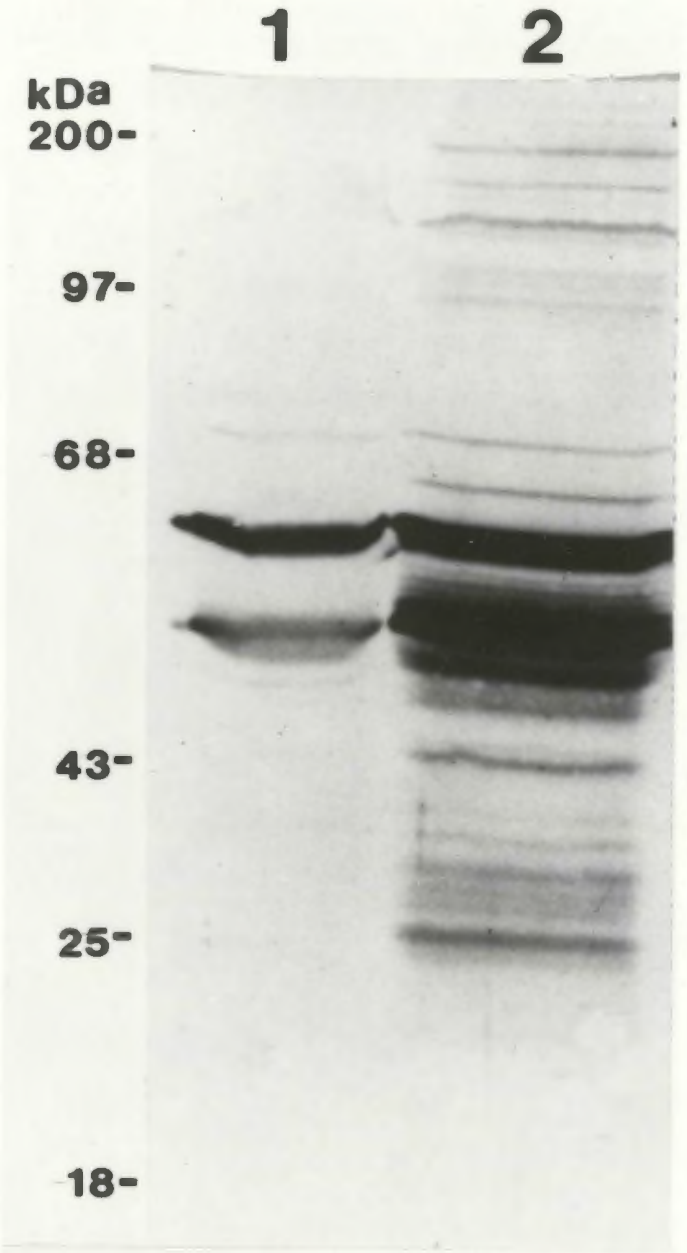
Figure 3.1, shows a western blot analysis of HeLa cytoplasmic and nuclear matrix fractions obtained during nuclear matrix preparation. The blot showed that the rabbit antiserum collected on day 38 recognized several proteins over a wide range of sizes. These proteins were predominantly in the nuclear matrix fraction (lane 2), and were reduced in the cytoplasmic fraction (lane 1). The most prominent bands reactive with antiserum were in the molecular weight range of 40-70 kDa. This western blot analysis indicated that the antiserum generated in the rabbit was enriched for the antibodies reacting with the nuclear matrix proteins.

3.1.3 Screening of the cDNA library

The polyclonal rabbit antiserum was used to screen $2X10^6$ recombinant phage of a λgt11 human fibroblast cDNA library. Lambda gt11 is a 43.7 kbp, linear, double-stranded DNA bacteriophage derived cloning vector into which a cDNA is inserted at a unique *Eco*RI site via linkers ligated to its ends (Huynh *et al.*, 1985). The unique *Eco*RI site in λgt11 is located in the carboxyl-terminus of the β gal gene and the inserted DNA is expressed under the control of the lac promoter. Thus, this library of recombinant phages could be screened with antibodies that recognized the nuclear matrix proteins to identify phage that express DNA segments encoding the antigenic determinants of these proteins. Twenty nitrocellulose membranes

Fig. 3.1 Specificity of rabbit polyclonal antiserum.

Total cytoplasmic (lane 1) and nuclear matrix proteins (lane 2) were prepared from HeLa cells and resolved on a 12% SDS polyacrylamide gel. The separated proteins were electrophoretically transferred to a nitrocellulose membrane and probed with rabbit polyclonal antiserum raised against HeLa nuclear matrix (dilution 1:200).


1
2
kDa
200-
97-
68-
43-
25-
18-

containing a total of $2x10^6$ plaques (about $1x10^5$ plaques/membrane) were lifted in duplicate, and screened with rabbit polyclonal antiserum as described in section 2.2.12. Initial screening of the library yielded 12 positive λgt11 clones, only those plaques that were present on duplicate membranes were considered true positive clones. Each positive clone was then plaque purified by repeated rescreening at successively lower plaque densities. The twelve individual clones denoted by the numbers λ101, λ102, λ103, λ104, λ106, λ107, λ108, λ109, λ110, λ111, λ112, and λ113 were grown up on a large scale. The DNA was extracted from each of the twelve clones, and digested with *Eco*RI to release the cDNA inserts from the cloning site in the vector. The digestion mixtures were resolved on a 1% agarose gel, and DNAs were visualized by ethidium bromide fluorescence. Based upon size of insert, λ103, λ109, λ110, λ111, λ112 were selected for further studies.

3.1.4 Identification of proteins encoded by cDNA clones

To identify the proteins encoded by the cDNA clones, polyclonal antibodies against five individual fusion proteins from λ103, λ109, λ110, λ111, λ112 were generated in rabbits. *E.coli* Y1089 was lysogenized separately with each of the recombinant phage, and induced with IPTG to produce β gal-fusion proteins (see section 2.2.16 and 2.2.17 for details). Liquid lysate made from each of the induced cultures was separated individually on a 7.5% preparative SDS polyacrylamide gel.

About 200-250 μg of β gal-fusion protein was loaded on a single preparative gel. The gel was stained with Coomassie blue dye and a gel slice containing the fusion protein was cut out of the gel, macerated, emulsified in Fruend's complete adjuvant (section 2.2.19), and used for immunization of a rabbit. After booster injections on day 14 and on day 28, antiserum was collected from each rabbit on day 40 and the specificity of the antiserum was tested by a western blot analysis (figure 3.2). The results of this analysis showed that of five rabbits immunized, only three produced antibodies that recognized the fusion protein (lanes 3, 4, and 5), the other two rabbits did not produce antibodies that reacted with the fusion protein (lanes 1 and 2). The rabbit antisera were also tested for immunofluorescent staining of HeLa cells grown on cover slips None of the antisera provided any conclusive results. These clones were not further characterized (see Discussion).

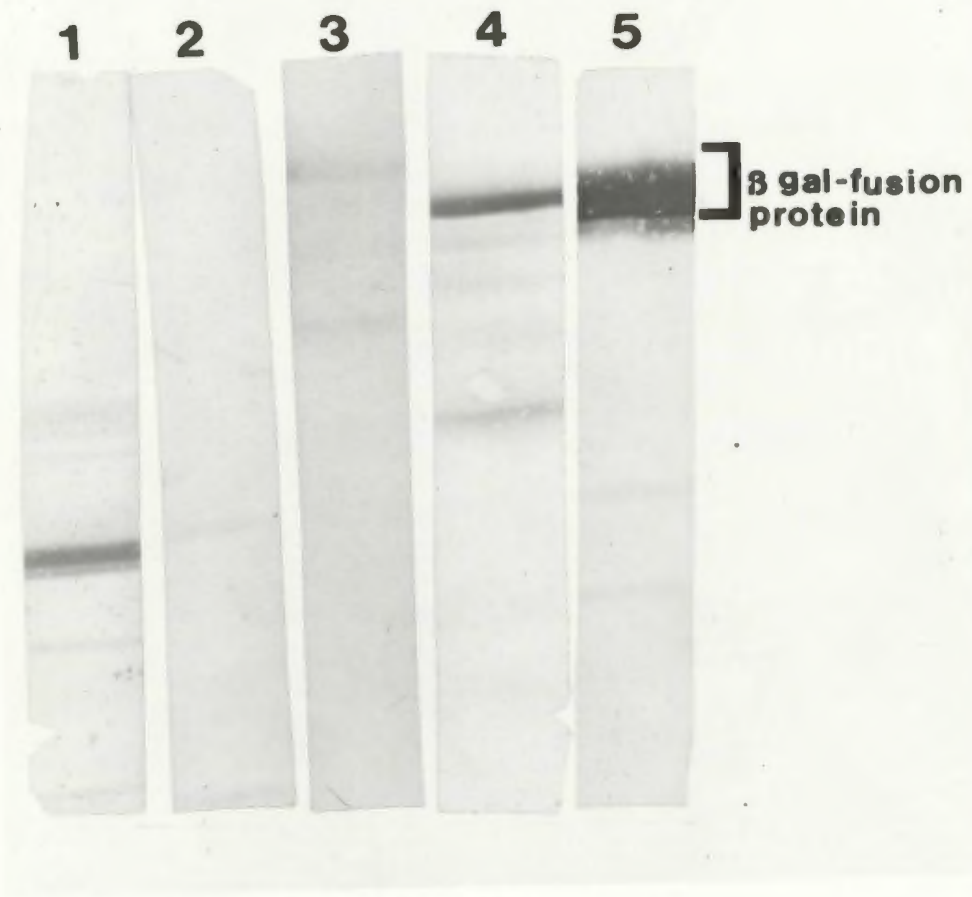
3.2 ISOLATION AND CHARACTERIZATION OF cDNA CLONES USING CHICKEN POLYCLONAL ANTIBODY RAISED AGAINST RAT LIVER NUCLEAR MATRIX

3.2.1 Specificity of chicken polyclonal antibody

As a second attempt to isolate cDNA clones for nuclear matrix proteins, the

Fig. 3.2 Reactivity of rabbit antisera for β gal-fusion protein.

Bacterial cell lysates were prepared from the lysogens of λ103 (lane 1), λ109 (lane 2),λ110 (lane 3), λ111 (lane 4), and λ112 (lane 5). Each sample were resolved on a 7.5% SDS polyacrylamide gel and separated proteins were electrophoretically transferred to a nitrocellulose membrane. Individual sample lanes were cut out and probed with their corresponding β gal-fusion protein antiserum raised in rabbit (dilution 1:100).


1
2
3
4
5
ß gal-fusion protein

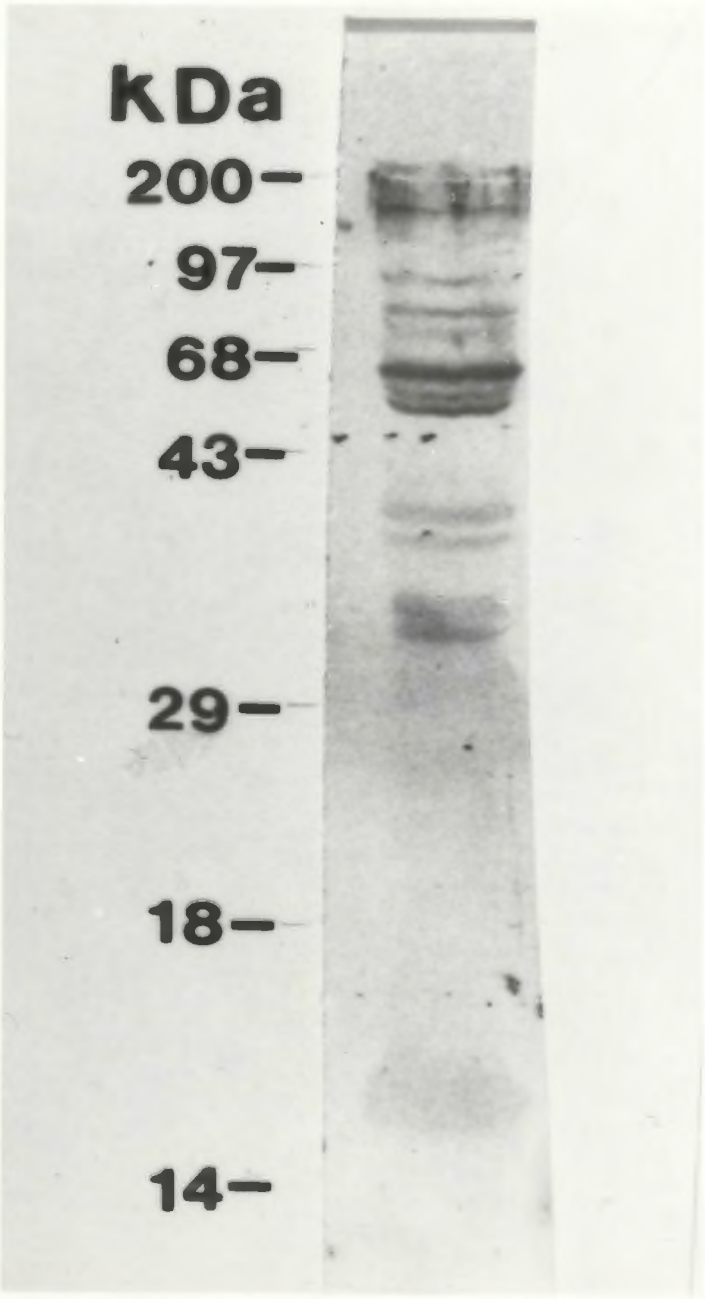
λgt11 library was screened with another polyclonal antibody preparation (CRLM), provided to us by Dr. H.C. Smith of Rochester University. The antiserum had been raised in a chicken against total nuclear matrix proteins prepared from rat liver (Smith *et al.*, 1985). Although CRLM was claimed (Smith *et al.*, 1985) to be reactive with only one major nuclear matrix protein (p107 in rat and p65.5 in human), it showed several immunoreactive bands over a wide size range from HeLa cell total nuclear proteins by immunoblot analysis (figure 3.3).

3.2.2 Screening of a cDNA library

Screening of the library was performed according to the method described in section 2.2.14. Screening of about $2x10^6$ individual recombinant phage (about $2x10^5$ plaques/membrane) resulted in the isolation of 18 positive λgt11 clones reactive with CRLM on duplicate plaque replicas. After three rounds of screening,13 remained positive and were designated as λ5A, λ7A, λ7B, λ7E, λ7F, λ7H, λ8A, λ8B, λ8C, λ8D, λ8F, and λ8H. DNA was purified from all 13 positive λgt11 clones, digested with *Eco*RI, separated on an agarose gel, and the sizes of the inserts determined. The inserts could not be released from five positive clones (λ5A, λ7D, λ7H, λ8A and λ8B). Two clones λ7E and λ7F, both contained two *Eco*RI fragments of about 950 and 350 bp, and another clone λ8F also contained two *Eco*RI fragments of 1100 and 550 bp. Two clones, λ7A and λ7B both contained an insert of 1300 bp. Three

Fig. 3.3 Specificity of chicken polyclonal antibody (CRLM).

Total nuclear proteins were prepared from HeLa cells and resolved on a 12% SDS polyacrylamide gel. The separated proteins were electrophoretically transferred to a nitrocellulose membrane and probed with the CRLM (dilution 1:1000). The numbers on the left of the lane represent approximate positions of molecular weight standards.


kDa
200
97
68
43
29
18
14

clones, λ8C, λ8D and λ8H contained inserts of 450, 550, and 950 bp, respectively.

To determine if there were DNA sequence identity between positive λgt11 clones, a Southern blot was prepared from *Eco*RI digested lambda DNAs. The blot was probed successively with each of the random primed radioactive-labelled cDNA inserts. Based on cross-hybridization experiments, the nine clones were grouped into six non-cross-hybridizing classes. These were: 1) λ7A and λ7B; 2) λ7E and λ7F; 3) λ8C; 4) λ8D; 5) λ8F; and 6) λ8H. Four clones which were strongly immunoreactive with CRLM, λ7A, λ7E, λ8D and λ8H were selected for further study.

3.2.3 Characterization of fusion proteins

3.2.3.1 Lysogen production

To confirm the reactivity of CRLM with antigenic determinants encoded by the cDNA inserts, the positive recombinant phages were used to infect *E.coli* strain Y1089. Lysogens were made from each of the recombinant phages (section 2.2.16); the individual lysogens were induced by IPTG to produce β gal-fusion proteins, and after two h incubation at 38°C, cultures were separated into cell pellet and supernatant fractions (section 2.2.17). The SDS polyacrylamide gel electrophoresis analysis of proteins from the pellets and supernatant indicated that almost all of the β gal-fusion protein remained in the cell pellets, therefore, cell pellets obtained after

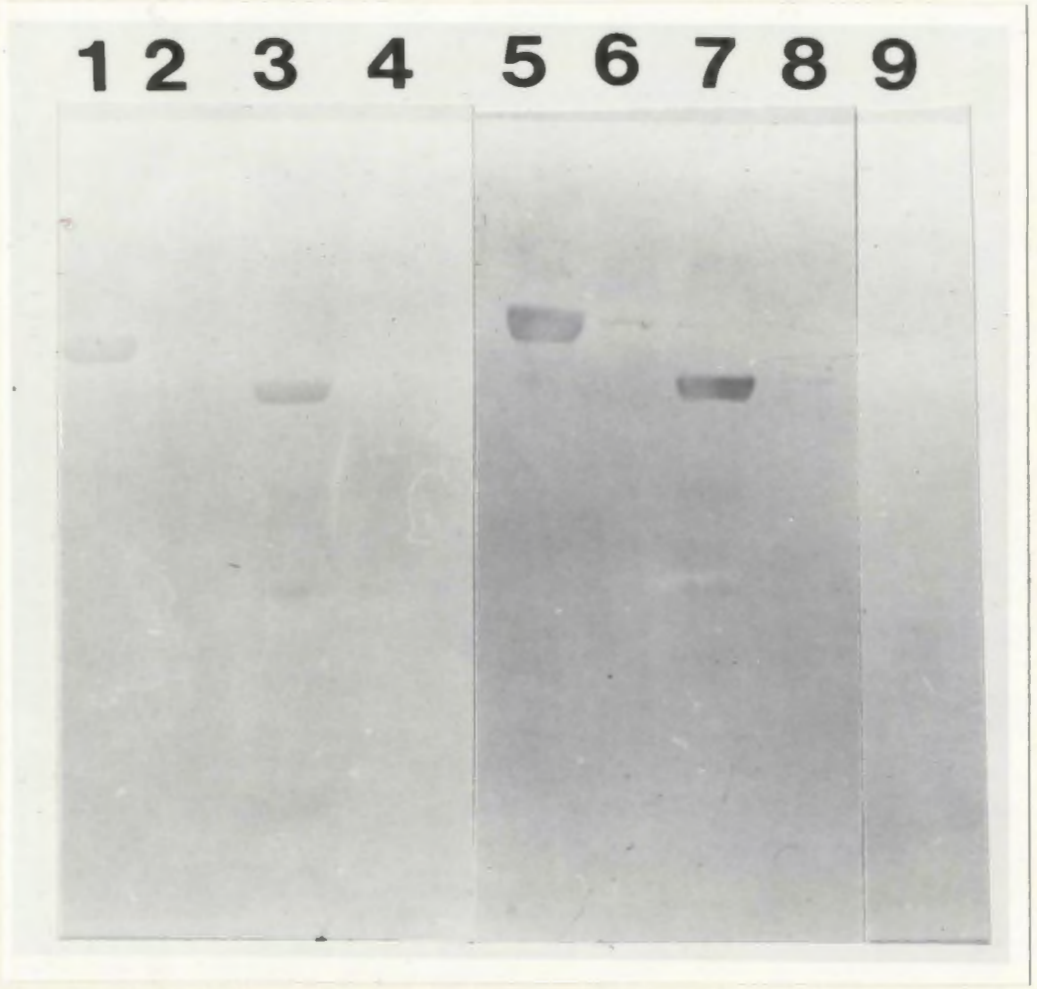
centrifugation were resuspended in TEP buffer (section 2.2.17). To release the β gal-fusion proteins from cells, each resuspended pellet was freeze thawed twice in liquid nitrogen, and the cleared lysate was used as a source of fusion proteins.

3.2.3.2 Reactivity of CRLM for β gal-fusion protein

Bacterial lysate prepared from each induced culture was resolved on a 7.5% SDS polyacrylamide gel (section 2.2.4). The separated proteins were transferred to a nitrocellulose membrane and after blocking non-specific protein binding sites on the membrane, the blot was incubated for 2 h in CRLM diluted 1:400 in blocking buffer. The blot was washed three times (10 min each) with TBST buffer, once (10 min) in TBS buffer, and incubated in goat anti-chicken antibody conjugated to horseradish peroxidase for one h. The membrane was then washed as described above and developed (section 2.2.5). In each of these cultures, IPTG induction resulted in the production of a β gal-fusion protein, which was reactive with CRLM (figure 3.4, lanes 1, 3, 5, and 7), whereas in a non-induced culture the CRLM reactive band was absent (lane 2, 4, 6, and 8). As a negative control, a separate lane containing β gal was also probed with CRLM (lane 9), as expected CRLM did not react with β gal. This clearly showed that the reactivity of CRLM with β gal-fusion proteins expressed by various cDNA clones was due to its immune reaction with the fusion part but not with β gal portion of the fusion proteins.

Fig. 3.4 Reactivity of CRLM for β gal-fusion protein.

Bacterial cell lysate were prepared from the lysogens of λ7E (lanes 1 and 2), λ7A (lanes 3 and 4), λ8H (lanes 5 and 6), and λ8D (lanes 7 and 8). After heat shock at 44°C, the lysogen cultures were either induced with IPTG (lanes 1, 3, 5, and 7) or not induced (lanes 2, 4, 6, and 8). Samples were resolved on a 7.5% SDS polyacrylamide gel, and separated proteins were electrophoretically transferred to a nitrocellulose membrane. The membrane was then probed with the CRLM (dilution 1:400). Lane 9 contained pure β gal and served as negative control.


1 2 3 4 5 6 7 8 9

3.2.4 Identification of cDNA clones

3.2.4.1 Generation of rabbit antiserum against fusion proteins

To determine if the immunoreactive fusion proteins encoded by λ7A, λ7E, λ8D, and λ8H were *bona fide* nuclear proteins, affinity purification of β gal-fusion protein specific antibodies from CRLM was attempted using a nitrocellulose membrane as the antigen carrier (section 2.2.20); this was unsuccessful. Therefore, I decided to raise antibodies against fusion proteins that could be used to identify the proteins encoded by the cDNA clones.

Large scale cultures of lysogens were grown to produce the fusion proteins. The β gal-fusion proteins used in immunization of rabbits were partially purified on a preparative SDS polyacrylamide gel as described in section 3.1.4. The gel slices corresponding to about 200-250 μg fusion protein were cut out of the gel, and used for immunization without any further purification. Four rabbits were immunized with the different β gal-fusion proteins as described in section 3.1.4. After six weeks, antiserum was collected from each immunized rabbit and tested against fusion protein on a western blot. All four rabbits produced antibodies, which were reactive with their corresponding antigen i.e., β gal-fusion protein. The rabbit immunized with 7E β gal-fusion protein gave strongest signal on a western blot, and therefore, it was analysed further.

3.2.4.2 Affinity purification of 7E β gal-fusion protein specific antibody from rabbit antiserum

Affinity purification of 7E β gal-fusion protein specific antibody was performed as described in section 2.2.20. To purify 7E β gal-fusion protein specific antibodies from the rabbit antiserum, the bacterial lysate containing the fusion protein was resolved on a 7.5% preparative SDS polyacrylamide gel. The gel containing separated proteins was electroblotted onto a nitrocellulose membrane. A horizontal strip, the entire width of the filter paper, corresponding to the 7E β gal-fusion protein was cut off and after blocking non-specific protein binding sites, the filter strip was incubated in antiserum containing anti 7E β gal-fusion protein antiserum. After 18 h at 4°C, the strip was washed in buffer as described in section 2.2.20. The anti 7E β gal-fusion protein antibodies were then eluted in a low pH buffer, immediately neutralized and stored at 4°C until used.

3.2.4.3 Identification and subcellular localization of protein encoded by λ7E cDNA

To determine the molecular weight and subcellular location of the protein encoded by λ7E cDNA, immunoblotting experiments were performed. Subcellular fractions (cytoplasmic and nuclear) were prepared from HeLa cells (section 2.2.10), and equal amounts of protein sample from each fraction were separated on a 12% SDS polyacrylamide gel. A western blot was prepared and probed with affinity-purified

anti 7E β gal-fusion protein rabbit antibodies (RA-7E) (section 2.2.25).

Figure 3.5 shows a western blot analysis of HeLa cell proteins with RA-7E. RA-7E recognized a single protein band highly enriched in the nuclear fraction (lane 2), which was undetectable in the cytoplasmic fraction (lane 1). The approximate molecular weight of the immunoreactive protein (referred hereafter as 7Eb protein) was calculated by comparing the mobility of size markers loaded in a parallel lane on the same gel. Thus the western blot analysis strongly suggested that 7E could be considered as a cDNA containing part of the mRNA sequence for a 52 kDa nuclear protein (7Eb protein).

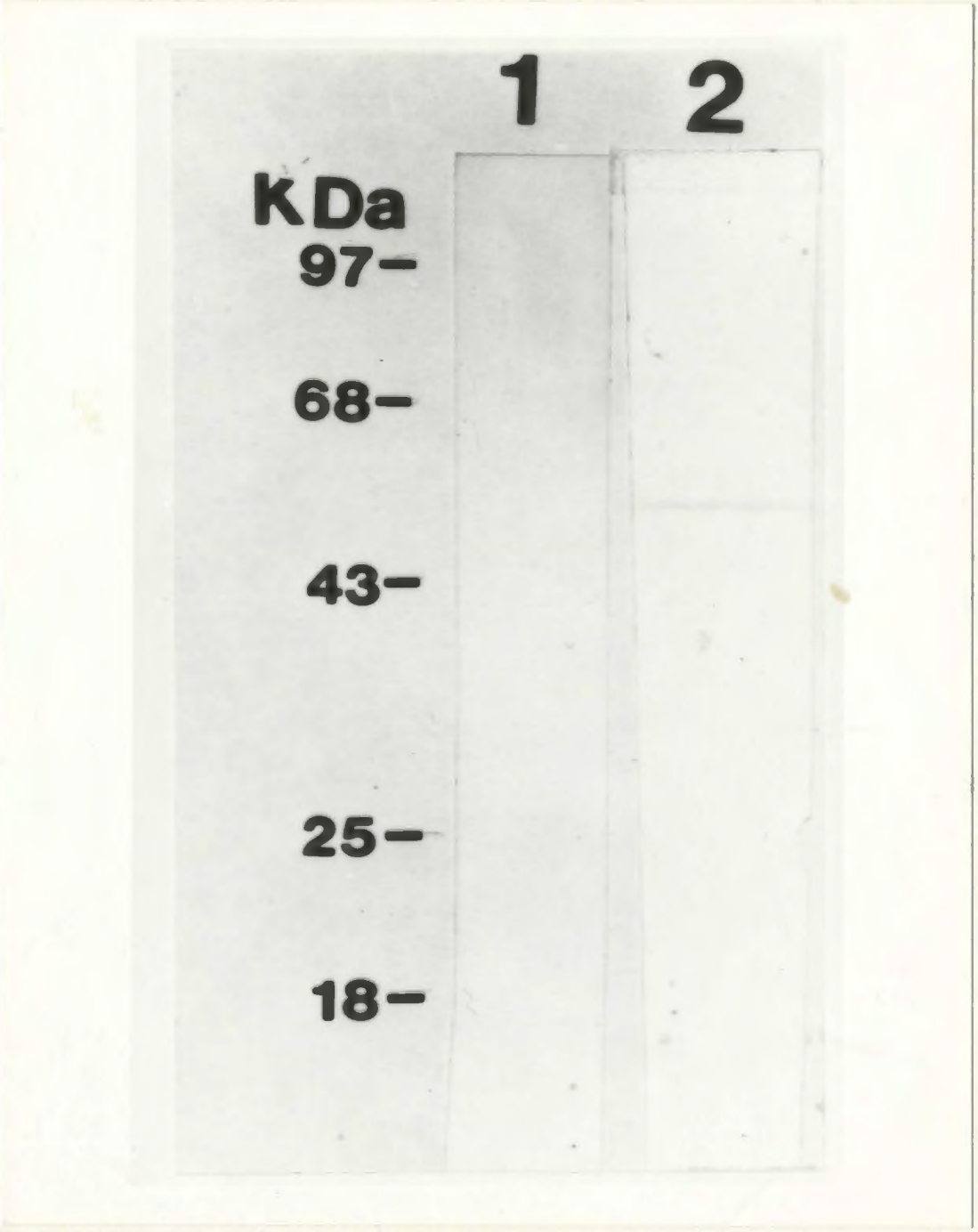
To independently test the nuclear location of the 7Eb protein, indirect immunofluorescent experiments were also carried out with RA-7E. These experiments did not provide any conclusive evidence for the localization of the 7Eb protein probably because the polyclonal antibodies obtained from the rabbit were against a denatured fusion-protein, and antibodies generated against a denatured protein often do not recognize cellular antigens. Therefore, I decided to prepare a batch of polyclonal antibodies to 7E β gal-fusion protein, which was subjected to the least possible denaturating conditions.

3.2.4.4 Affinity purification of 7E β gal-fusion protein

Affinity purification of 7E β gal-fusion protein was carried out as described in section 2.2.18. To purify the 7E β gal-fusion protein from the lysate it was first

Fig. 3.5 Identification of a protein encoded by λ7E cDNA and its subcellular localization.

Total cytoplasmic (lane 1) and nuclear proteins (lane 2) were prepared from HeLa cells and resolved on a 12% SDS polyacrylamide gel. The separated proteins were electrophoretically transferred to a nitrocellulose membrane and probed with affinity purified anti-7E β gal-fusion protein rabbit antibody (RA-7E).


1
2
KDa
97-
68-
43-
25-
18-

concentrated by ammonium sulphate precipitation, and then loaded on to a column containing containing a β gal affinity matrix (p-aminobenzyl-1-thio-D-galctopyranoside-agarose). The flow through fractions were collected and passed through the column once again. After washing the column with several bed volumes of Tris buffer, the fusion protein was eluted by 100 mM sodium borate buffer (pH 10.5), in a relatively pure form. A small aliquot was analyzed by SDS polyacrylamide gel electrophoresis, and as shown in figure 3.6, only the fusion protein band was detected. This indicated the effectiveness of the affinity chromatography procedure in separating the fusion protein from other contaminating bacterial proteins.

3.2.4.5 Generation of rabbit antiserum to 7E β gal-fusion protein

A total of six rabbits were used to generate polyclonal antiserum against 7E β gal-fusion protein. Two rabbits received affinity purified 7E β gal-fusion protein, while two other were immunized with a gel slice containing fusion protein and still another two were injected with nitrocellulose membrane containing fusion protein (see section 2.2.19). After the primary immunization, the rabbits were boosted twice on day 14 and day 28, respectively. Serum was collected from each immunized rabbit on day 40 and tested by western blot analysis of lysate containing 7E β gal-fusion protein (figure 3.7). Five of the rabbits produced antibodies which reacted strongly with 7E β gal-fusion protein in addition to *E coli* proteins (lanes 1, 2, 4, 5 and

Fig. 3.6 SDS polyacrylamide gel electrophoresis analysis of affinity purified 7E β gal-fusion protein.

An aliquot (about 5 μg) of affinity purified 7E β gal-fusion protein was resolved on a 7% SDS polyacrylamide gel. The gel was stained with Coomassie blue.

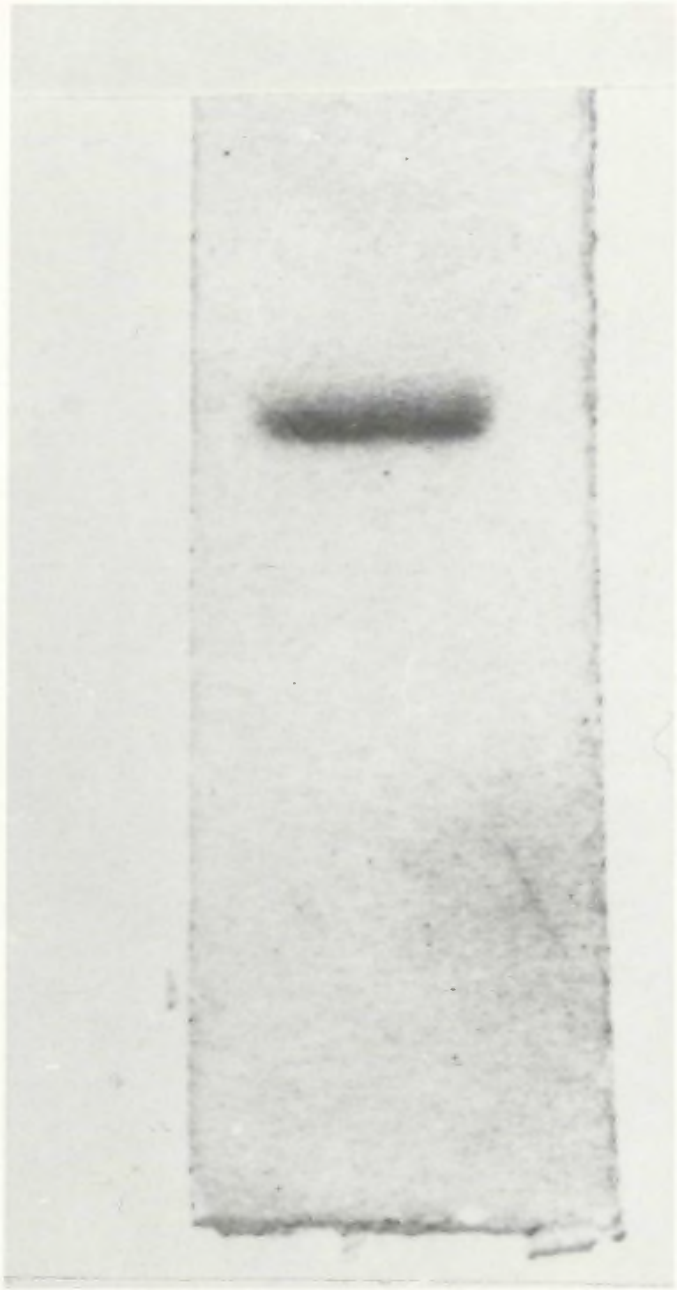
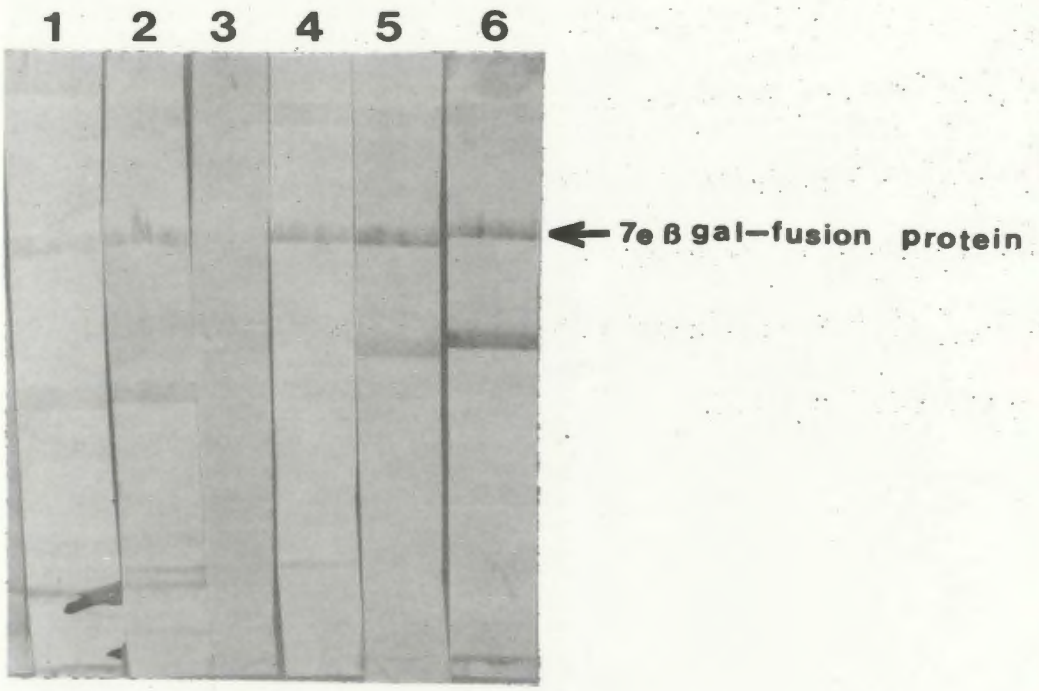


Fig. 3.7 Reactivity of rabbit antisera for 7E β gal-fusion protein.

Bacterial lysate prepared from the lysogen of λ7E and resolved on a 7.5% SDS polyacrylamide gel. The separated proteins were electrophoretically transferred to a nitrocellulose membrane and probed with rabbit antisera raised against different preparations of 7E β gal-fusion protein. Lane: 1 and 2, blot probed with antisera collected from rabbits immunized with affinity purified 7E β gal-fusion protein; 3 and 4, blot probed with antisera collected from rabbits immunized with nitrocellulose membrane containing 7E β gal-fusion protein; 5 and 6, blot probed with antisera collected from rabbits immunized with a gel slice containing fusion protein. Antisera were used at a dilution 1:100.


1 2 3 4 5 6
7e ß gal–fusion protein

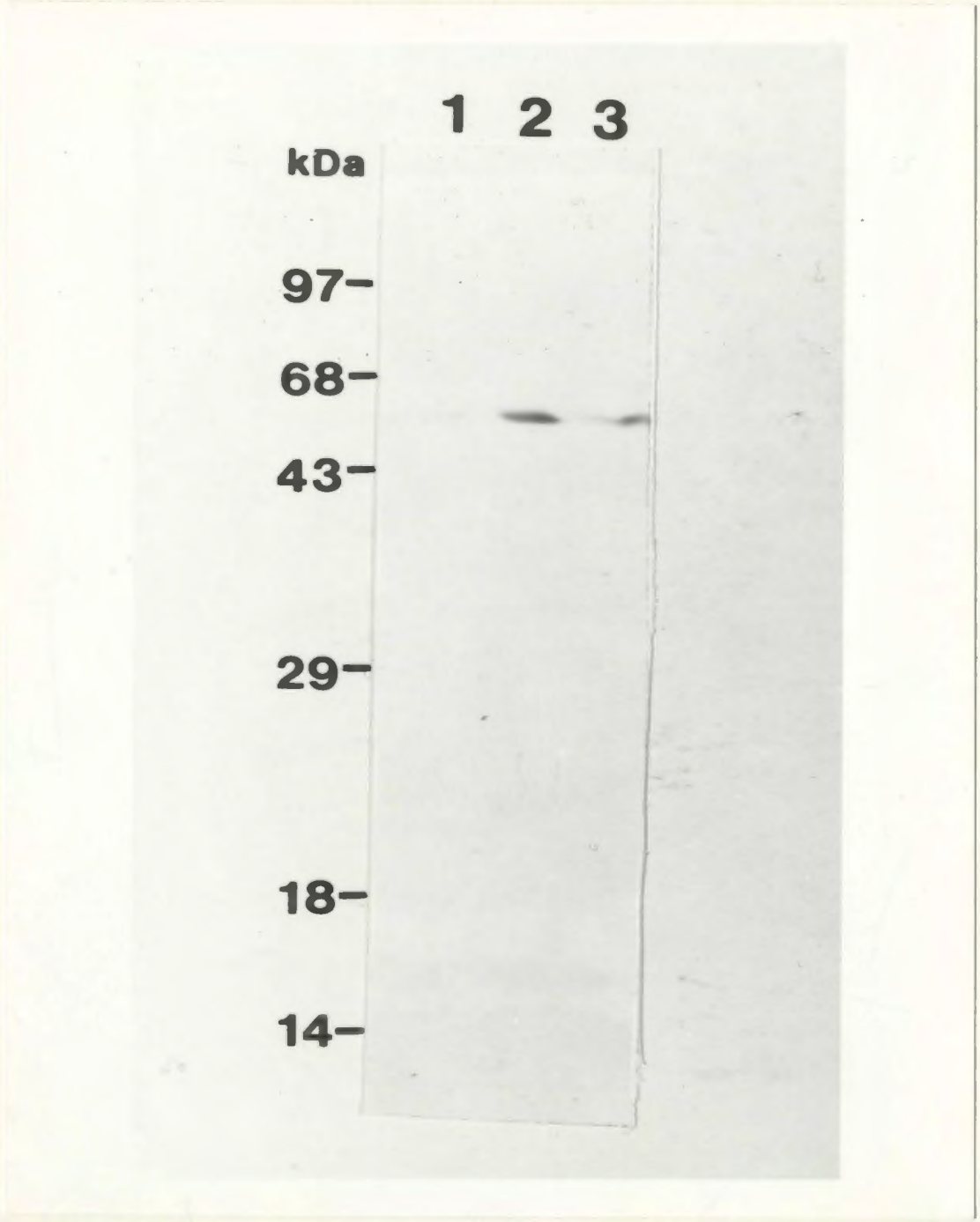
6). The rabbit wich was immunized with nitrocellulose containing fusion protein, produced antibodies which gave a weak signal on western blot (lane 3). Despite the fact that three different types of preparations of fusion protein were used for immunization, no difference in specificity or titre of the antiserum was observed. Specific antibodies to 7E β gal-fusion protein were purified from rabbit antiserum as described in section 3.2.3.2, and tested by immunoblot analysis. These preparations of affinity purified antibodies reacted with a single 52 kDa nuclear protein as detected earlier (section 3.2.4.3). Furthermore, a weak but exclusively nuclear staining was also observed when these antibodies were used in immunofluorescent studies of HeLa cells. However, it was not possible to take photographs.

3.2.4.6 Localization of the 7Eb protein in the nuclear matrix fraction

To further confirm the subcellular location of the protein encoded by λ7E, a western blot was prepared from various subcellular fractions that were obtained during nuclear matrix preparation from HeLa cells (section 2.2.10, method ii). The proteins were resolved on a 12% SDS polyacrylamide gel and electroblotted to a nitrocellulose membrane. The blot was then probed with RA-7E. The antibody specifically reacted with a 52 kDa protein band in the lane containing total nuclear proteins (figure 3.8, lane 2) or nuclear matrix proteins (lane 3). No antigen-antibody reaction was observed with cytoplasmic proteins (lane 1). These results of this western blot analysis indicated that the 7Eb protein is a component of the nuclear

Fig. 3.8 Localization of the protein encoded by λ7E in the nuclear matrix fraction.

Total cytoplasmic (lane 1), nuclear (lane 2), and nuclear matrix proteins (lane 3) were prepared from HeLa cells and resolved on a 12% SDS polyacrylamide gel. The separated proteins were electrophoretically transferred to a nitrocellulose membrane and probed with affinity purified anti-7E β gal-fusion protein rabbit antibody (RA-7E).


1 2 3
kDa
97-
68-
43-
29-
18-
14-

matrix.

3.3 NUCLEOTIDE SEQUENCING OF λ7E CLONE

To determine the nucleotide sequence of the cDNA inserts from clone λ7E completely, it was necessary to subclone the fragments into a plasmid vector. For this purpose λ7E DNA was digested with *Eco*RI and the two *Eco*RI fragments 7Ea and 7Eb of sizes 360 and 948 bp, respectively were purified from the gel by using Gene clean (section 2.2.8), and ligated into a bluescript SK+ plasmid using the unique *Eco*RI cleavage site in the polylinker region. The resulting plasmid subclones designated as p7Ea and p7Eb were used to transform *E.coli* strain XL1-blue cells (section 2.2.21). A large number (800-1000) of transformants were obtained, and 24 of them were picked, grown and subjected to plasmid DNA isolation using the mini-prep method (section 2.2.22). The digestion of the resultant plasmid DNA with *Eco*RI followed by agarose gel analysis revealed that 21 plasmid isolates contained cDNA inserts of the right size. One clone selected from each of the cDNA fragments (7Ea and 7Eb) was used for sequencing. The cDNAs cloned in the bluescript plasmid vector were sequenced by the dideoxy sequencing method (Sanger *et al.*, 1977) using a sequenase kit. Both double-stranded and single-stranded plasmid DNA was sequenced using various sequencing primers including the M13 universal sequencing primer, the M13 reverse sequencing primer and the SK and KS primers

which are specific for the bluescript plasmids. The sequences of about 150 nucleotides from each end of the insert were determined directly from the λ7E DNA using λgt11 sequencing primers. In order to to avoid any ambiguity in the sequence data, 7Eb insert DNA which contained 360 bp was sequenced in both directions, using SK and KS primers. Direct sequencing of the λ7E clone with λgt11 primers showed that the 7Eb cDNA insert sequence contained an open reading frame which was adjacent to, and in register with, the β gal gene of the λgt11 phage. For the determination of the complete sequence for 7Ea DNA, various restriction endonuclease generated fragments were recloned into bluescript plasmids and used for sequencing. The location of the restriction endonuclease sites, the strategy used for sequencing, and the complete nucleotide sequence for λ7E DNA is shown in figure 3.9. A computer search for open reading frames in the λ7E sequence discovered one large open reading frame of 121 amino acid residues in λ7E DNA (figure 3.9). The reading frame, which was in register with the β-gal sequence of λgt11 and therefore, is expressed by λ7E, extends from the first nucleotide to nucleotide 363, where a TGA stop codon was encountered. A molecular weight for the polypeptide predicted from the open reading frame was calculated to be 14.9 kDa. A computer search for open reading frames in 7Ea alone did not show any open reading frame longer than 59 amino acids in all six reading frames. A homology search done on the EMBL DNA sequence data bank, (release 18.0) did not reveal any significant homology between 7Eb and any data bank entry. However, the 7Ea

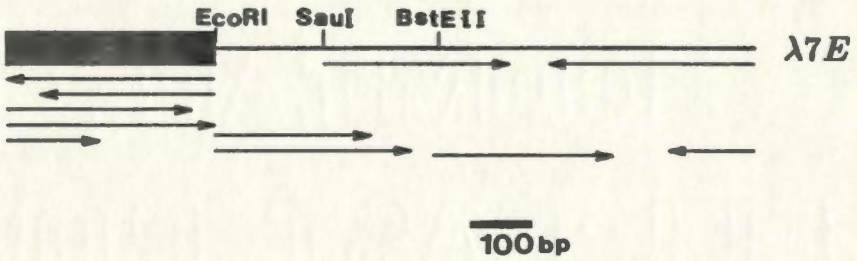
Fig. 3.9 Nucleotide sequence of λ7E DNA and the deduced amino acid sequence of the largest open reading frame.

(A) The nucleotide sequence is written in the 5' to 3' direction. The nucleotide sequence is numbered on the right and the amino acid sequence is numbered at the left. Nucleotides in bold characters represent *Eco*RI sites. An asterisk sign in the protein sequence indicates the end of reading frame. The asterisk denotes a termination codon.

(B) Schematic representation of λ7E cDNA. The filled box represents the coding region and the solid line represents non coding DNA. Horizontal arrows indicate the direction and appropriate positions of the region sequenced.

```
    GAATTCCGCGGGCAGAAAAAGCGCCCGCAGCTCCGGCTGAGCGCAGTCAGAGCAGCATCG    60
  1  E  F  R  G  Q  K  K  R  P  Q  L  R  L  S  A  V  R  A  A  S

    TTCCTGAGGAAGAACAAGCAGCTAACAAAGGCGAGGAGAAGAAAGACGACGAGAACATTC   120
 21  F  L  R  K  N  K  Q  L  T  K  A  R  R  R  K  T  T  R  T  F

    CGATGGAGACAGAGGAGACACACCTTGAAGAAACAACAGAGAGCCAACAGAATGGTGAAG   180
 41  R  W  R  Q  R  R  H  T  L  K  K  Q  Q  R  A  N  R  M  V  K

    AAGGCACGTCTACTCCTGAGGACAAGGAGAGTGGGCAGGAGGGGGTCGACAGTATGGCAG   240
 61  K  A  R  L  L  L  R  T  R  R  V  G  R  R  G  S  T  V  W  Q

     AGGAAGGAACCAGTGATAGTAACACTGGCTCGGAGAGCAACAGTGCAACAGTGGAGGAGC  300
 81   R  K  E  P  V  I  V  T  L  A  R  R  A  T  V  Q  Q  W  R  S

     CACCAACAGATCCCATACCAGAAGATGAGAAAAAAGAATAAGTGTTGCCTTGTGGAATTC  360
101   H  Q  Q  I  P  Y  Q  K  M  R  K  K  N  K  C  C  L  V  E  F

    CGGTGACTCACTCTGTCTGCCCAGGCTGGAGTGCAGTGGCGGATCTAGCTCACTGCAACC   420
121  R  *

    TCCGCCTCCCAGGTTCAAGCGATCCTGCCTCAGCCTCCTGAGTAGCTGGGATTATAGGTG   480
    TGTGTGCCACGTCTCGGCTAGTTTTTTTGTATTTTTAGTAGAGATGGGTGTTTCACCATT   540
    TTGTTCAAGGCTGGTCTTGAACTCCTGACCTCAGGTGATCCGCCCACCTTTGGCCTCCCA   600
    AAGGTGCTGGGATTCAAGGCGTAAGCCGCGCTCCGGCCTGCGGCCTGCAAATGCTATTTC   660
    AAAAGACGTTCTTCATGATGCTGCTGTGTTTCTCAGTGGATGGCTTTGTAGCACTGGTGT   720
    TTCTTGATAATACTTAGGGGGTCACCAGATTTCATATTTTCCGTTTTTAAAAAAATTCAG   780
    AATTTCTGTGAATCTTGCTGGAAACGATTTAGTTTCTCTGGATTATTGGATGCCATTTGG   840
    AAATCACGAATATGTCTGATAGTTTCATCACCTGGCTGTCATAAGTCTCTGATCCGCATC   900
    TCTCATTCATTCTTGCTAAAGCTAACAAGGTGACTATATCAGATTGGCTGCTTCCACAGC   960
    TATGCTCAAAGGGTCTTTCCCTTCTTCATCAGTGGCATGTTGATTGGCACCTCGTTTTAG  1020
    GAATAAACATACCTGCCCTGTGTGCCCTAAGACGGTGGCATGGTGCAATGGTCCCCCGCC  1080
    CTTGGACATCTCTTTGGTTGACATTAGCACCATTCTGTAGGAGGAACTCACACGTCACCA  1140
    AAGAGCCCCCTAATACAGCCTGAATTAAGTGGTGTCGCTTTGTTTCTCTGAATTGGCCAG  1200
    TTCACGTCTGCACCATGAGCCAAAGCCTCAGCCATTTTTAGGAAGTTTTTTTCATATGAC  1260
    GCCCTATAAAGCTAAGTCCTGGAGGAATTC                                 1290
```

(B)



sequence showed stretches of homology with mammalian repeat sequences (Alu sequence). Because the λ7E clone contained two *Eco*RI fragments, 7Ea and 7Eb, it was possible that they were two independent cDNA clones, ligated together in a single λgt11 DNA molecule. Therefore, I concentrated on characterization of 7Eb.

3.4 SCREENING OF THE cDNA LIBRARY WITH 7Eb DNA

The molecular weight estimated for the 7E β gal-fusion protein was about 130 kDa from SDS polyacrylamide gel electrophoresis analysis. This is quite consistent with its size deduced from sequence data, since the protein (M_r14.9 kDa) deduced from the coding sequence was expressed as a fusion with *E.coli* β gal (M_r=116 kDa). However, to code for a 52 kDa protein which the western blot analysis suggested at least 1.5 kbp of DNA is required.

To isolate full-length cDNA, the library was screened again according to the method described in section 2.2.14, using the 7Eb cDNA insert as a probe. Screening of about $3x10^5$ recombinant clones ($5x10^4$ plaques/membrane) resulted in the isolation of five positive λgt11 clones designated as λ1B, λ2B, λ1G, λ3A, λ6A, and λ6D. Each positive clone was plaque purified by two rounds of rescreening at lower densities. The isolated positive plaques were used to prepare liquid lysate, from which DNAs were purified (section 2.2.15). DNAs were then digested with *Eco*RI, separated on an agarose gel, and the size of the inserts were determined.

Among all positive clones, λ1B contained the largest sized cDNA insert of about 950 bp, whereas, λ1G, λ3A,λ6A, and λ6D contained insert of 650, 500, 500, and 650 bp respectively. Digestion of the 2B clone DNA did not release any insert, suggesting that the *Eco*RI site may have been lost during cloning.

3.5 NUCLEOTIDE SEQUENCING OF 7Eb-SELECTED CLONES

Prior to subcloning into a plasmid vector, a sequence of about 100 to 150 nucleotides was determined directly for each positive clone from both ends using λgt11 primers. This nucleotide sequence data showed that λ3A and λ1G had nucleotide sequences similar to the nucleotide sequences of λ6A and λ6D, respectively.

The cDNA inserts from λ1B, λ3A, and λ6D were excised from the vector DNA (λgt11) and purified by Geneclean prior to ligation with the bluescript plasmid (SK+). The ligated DNAs were then used to transform *E.coli* XL1-blue cells (section 2.2.21). Transformants that contained the appropriate plasmids, as determined by restriction endonuclease analysis of mini-preps, were grown in large scale to prepare DNA for sequencing. To determine the complete sequence for λ1B and λ6D, which were 947 and 682 bp long, respectively, various restriction endonuclease sites within the cDNA inserts were used and the restriction fragments were subcloned into the bluescript plasmid. The restriction endonuclease sites within

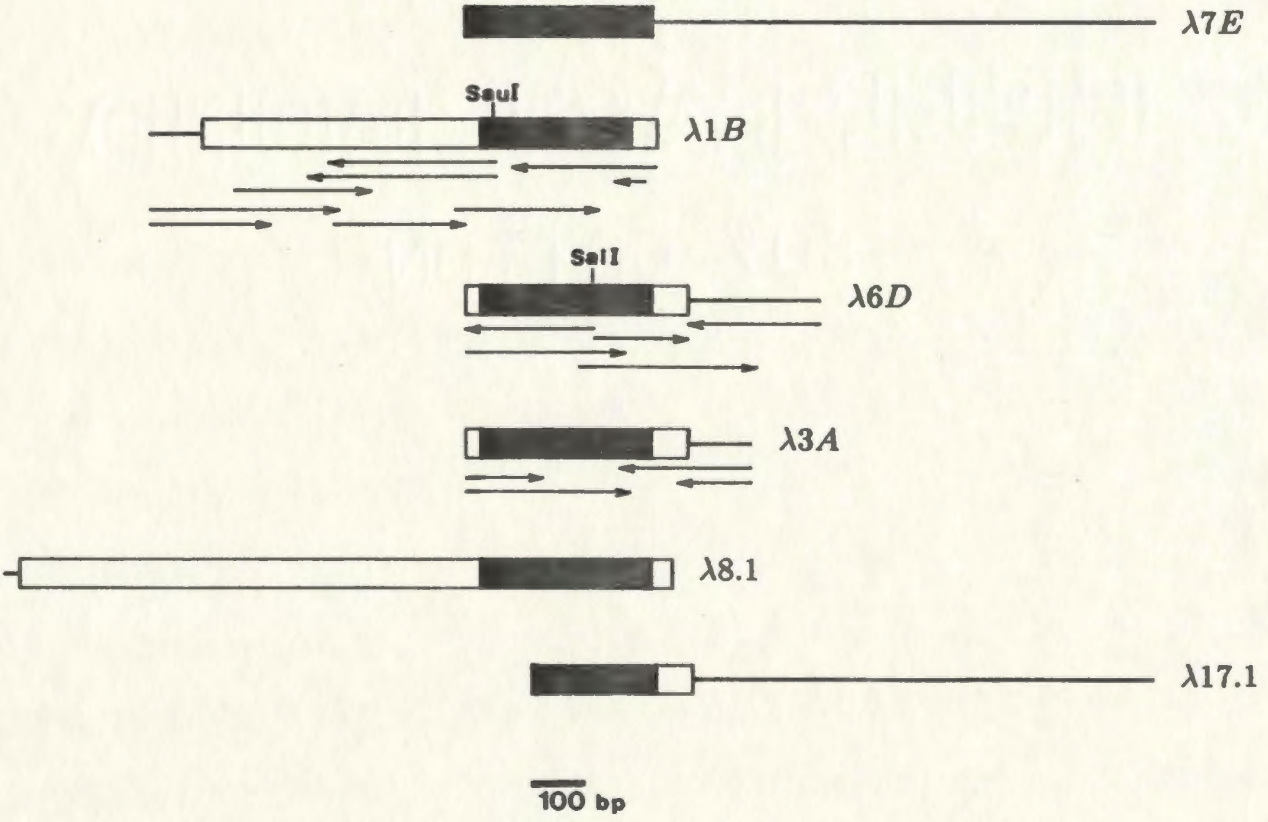
the cDNA inserts of λ1B and λ6D, which were used for subcloning and sequencing are also shown in figure 3.10. The complete nucleotide sequences for λ3A λ6D and λ1B inserts were determined (appendices 1, 2, 3). The sequences were aligned using PC/GENE computer program. The alignment of the λ1B sequence with that of 7Eb showed sequence identity of 276 bp. This is shown schematically in figure 3.10. Similarly, the alignment of the λ6D sequence with that of 7Eb revealed sequence identity of 318 bp between the two (figure 3.10). This comparision analysis also showed that the λ6D nucleotide sequence completely overlapped the λ3A sequence and contained an additional 140 bp (figure 3.10). Moreover, λ1B contained a single open reading frame of 855 bp (from nucleotide 92 to 947), coding for 285 amino acids. λ6D contained an open reading frame of 426 bp (1 to 426), coding for 142 amino acids.

3.6 SCREENING OF THE cDNA LIBRARY WITH λ1B cDNA CLONE

While the sequencing of λ1B DNA was still in progress, the analysis of preliminary sequence data indicated a possible reading frame in λ1B cDNA with several stop codons in the middle. This was later found not to be true when the insert DNA from λ1B was completely sequenced (see above). However, based on this early information about reading frames, I decided to rescreen the same library

Fig. 3.10 Alignment of 7Eb-selected cDNA clones.

The complete nucleotide sequences of λ3A, λ6D, λ1B, λ8.1 and λ17.1 were determined. The sequence data generated were used to align the sequences with the λ7E sequence using the PC/GENE program. The alignment of various clones is shown here schematically. The horizontal bar represents coding regions. The filled area within the horizontal bar indicates sequence identity between the 7Eb sequence and various other cDNAs. The solid line represents untranslated sequences. The sequence strategy is also shown with relevant restriction endonucleases sites. The horizontal arrows indicate the direction and appropriate portion of the regions sequenced.


λ7E
SauI
λ1B
SalI
λ6D
λ3A
λ8.1
λ17.1
100 bp

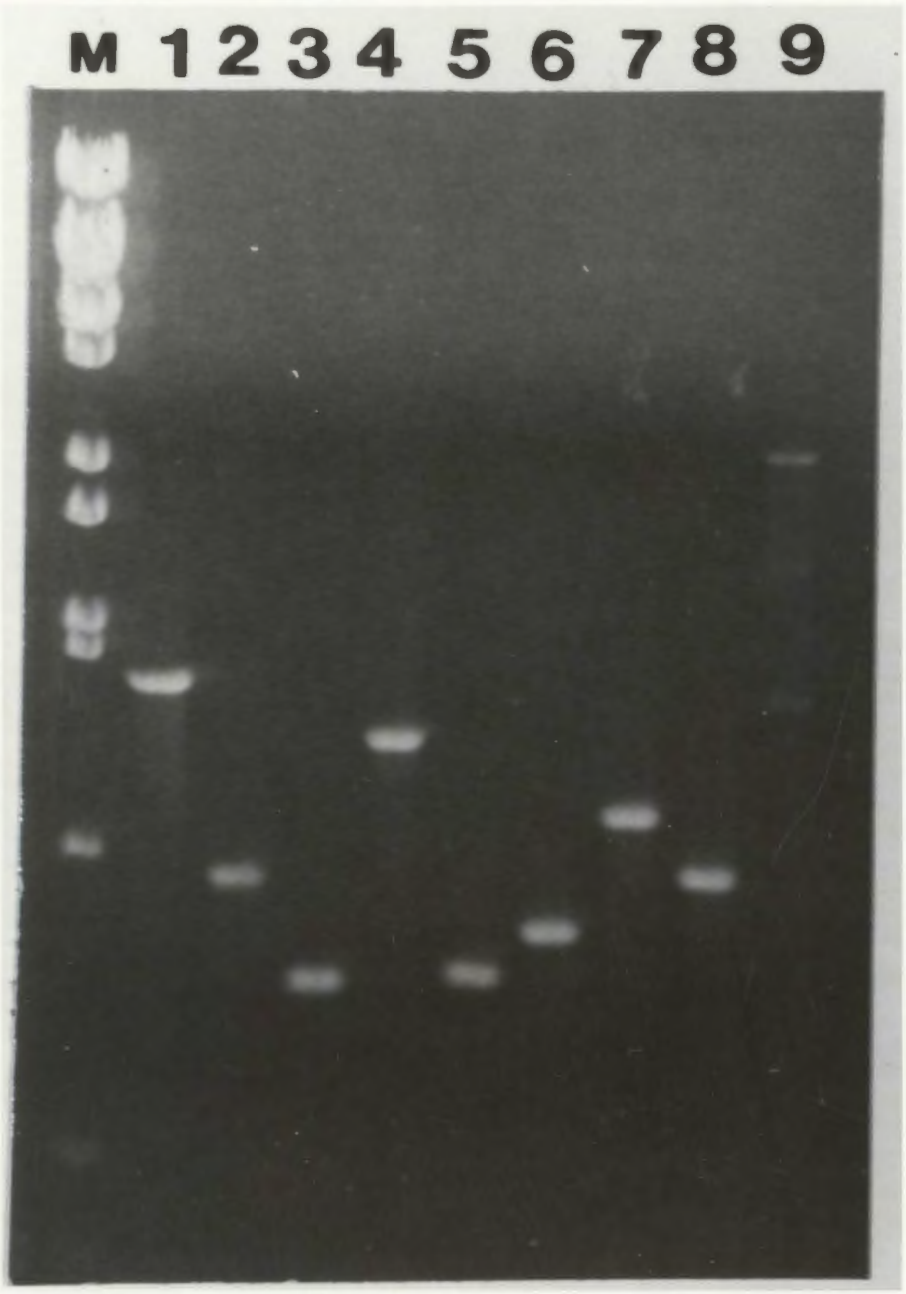
to isolate a longer cDNA, particularly one that contained more 5' sequence.

λ1B DNA was subjected to double digestion with *Eco*RI and *Xho*II to isolate the 242 bp *Eco*RI-*Xho*II fragment from the 5' end of λ1B. The fragment was purified from agarose gel, radioactively-labelled with ^{32}P by random priming, and used as a probe to rescreen the λgt11 library (section 2.2.14).

Screening of about $1x10^5$ recombinant phages, resulted in the isolation of seven positive λgt11 clones designated as λ5a, λ5b, λ5c, λ5d, λ5e, λ5f, λ5g, and λ5h, respectively. Each of the positive clones repeatedly gave a positive signal after secondary and tertiary screening. The cDNA inserts from the positive clones were directly amplified from purified plaques by the polymerase chain reaction (PCR; section 2.2.27). Each positive phage plaque was picked, and resuspended directly into deionized distilled water. To release λ DNA from bacteriophage particles, the resuspended plaque was frozen at -70°C and then thawed at room temperature and subsequently part of freeze thaw extract was used in the amplification reaction. The amplification of the cDNA insert was performed on a temperature cycler in a reaction containing λgt11 PCR primers as described in section 2.2.27. After 30-40 cycles of amplification, aliquots of the PCR reactions were loaded on a 1% agarose gel along with DNA size markers. The amplified insert size was deduced by comparing the mobility of the PCR product with those of the size markers. As shown in figure 3.11, lanes 2-9, λ5a, λ5b, λ5c, λ5d, λ5e, λ5f, λ5g, and λ5h contained cDNA inserts of about 650, 500, 1,100, 500, 550, 800, 600 and 2,400 bp respectively.

Fig. 3.11 Agarose gel electrophoresis of PCR-amplified inserts DNA from λ1B-selected clones.

DNA fragments were detected by staining the gel with 0.5 μg/ml ethidium bromide. Lane: 1, λ1B; 2, λ5a; 3, λ5b; 4, λ5c; 5, λ5d; 6, λ5e; 7, λ5f; 8, λ5g; 9, λ5h. Lane M represents DNA size marker (λ*Bst* E II digest; New England BioLabs) .


M 1 2 3 4 5 6 7 8 9

A good yield of PCR products was obtained from each positive plaque, except in the reaction containing λ5h. Because the extension time of one min was chosen during PCR temperature cycling, it is possible that this was not sufficient for the amplification of longer DNA molecules such as the insert from λ5h which as mentioned above was about 2.4 kbp long.

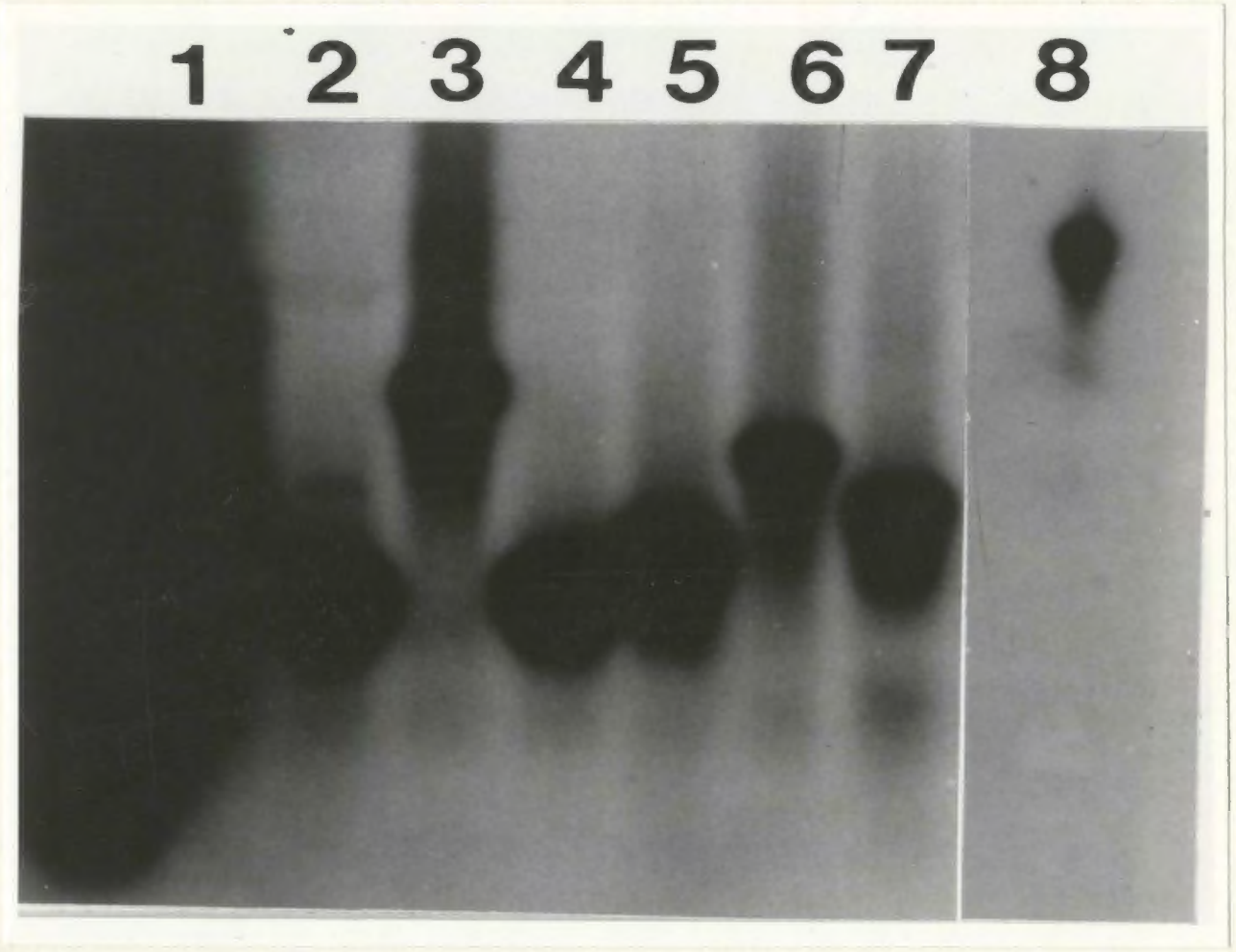
To determine the specificity of PCR generated products, a Southern blot was prepared, and probed with the labelled *Eco*RI-*Xho*II λ1B fragment (figure 3.12). After 6 h hybridization, the blot was washed at high stringency (0.2X SSC at 68°C), and exposed for 4 h. As shown in figure 3.12, hybridization signals were detected in every lane, showing sequence similarity between the λ1B insert DNA and the λ5 series of cDNA clones.

3.7 NUCLEOTIDE SEQUENCING OF λ1B-SELECTED CLONES

Based on the sizes of the cDNA inserts, the clone λ5h was initially selected for nucleotide sequencing and later, partial sequences were determined from the rest of the clones. To determine the complete nucleotide sequence of the cDNA insert of λ5h, which was about 2,400 bp, it was necessary to subclone the fragment into a plasmid vector. The cDNA clone λ5h was grown in liquid culture medium and DNA was isolated and purified by CsCl step gradient centrifugation (section 2.2.15). The purified DNA was digested with *Eco*RI to release the cDNA insert from the vector

Fig. 3.12 Hybridization of ^{32}P-labelled *Eco*RI-*Xho*II λ1B fragment with a Southern blot of cDNA inserts PCR-amplified from λ1B-selected clones.

A Southern blot prepared from the gel shown in figure 3.11, was probed with ^{32}-P labelled *Eco*RI-*Xho*II λ1B fragment. The blot was hybridized at 44°C and washed with a final stringent wash in 0.2 X SSC at 65°C. Lane: 1, λ5a; 2, λ5b; 3, λ5c; 4, λ5d; 5, λ5e; 6, λ5f; 7, λ5g; 8, λ5h.


1 2 3 4 5 6 7 8

DNA and purified on an agarose gel. The gel slice containing the insert DNA fragment was separated from the rest of the gel and the DNA was further purified by the Geneclean method (section 2.2.28). The purified insert DNA was then used for subcloning into a bluescript SK+ plasmid, and the resulting plasmid was designated as p5h.

To determine the nucleotide sequence of p5h insert DNA completely, unidirectional deletions were created in the insert DNA using Exonuclease III (section 2.2.23.1), and the resulting subclones were used for making plasmid preparations for nucleotide sequencing.

Plasmid DNAs were subjected to double-strand DNA sequencing, and a complete nucleotide sequence of 1237 bases was determined (appendix 4). Comparision of the nucleotide sequence of λ1B with that of λ5h revealed a sequence identity of only 64 bp (nucleotide 12 to 75 of λ1B) with λ5h; outside this 64 bp region, the rest of the sequence was unrelated to λ1B. Therefore, λ5h was not characterized any further. The same 64 bp were found in all of the clones (λ5a, λ5b, λ5c, λ5d, λ5e, λ5f, and λ5g; appendicis 5, 6, 7, 8, 9, 10, 11) DNA; these clones were also not characterized further.

3.8 SCREENING a λZAP II cDNA LIBRARY WITH λ6D DNA

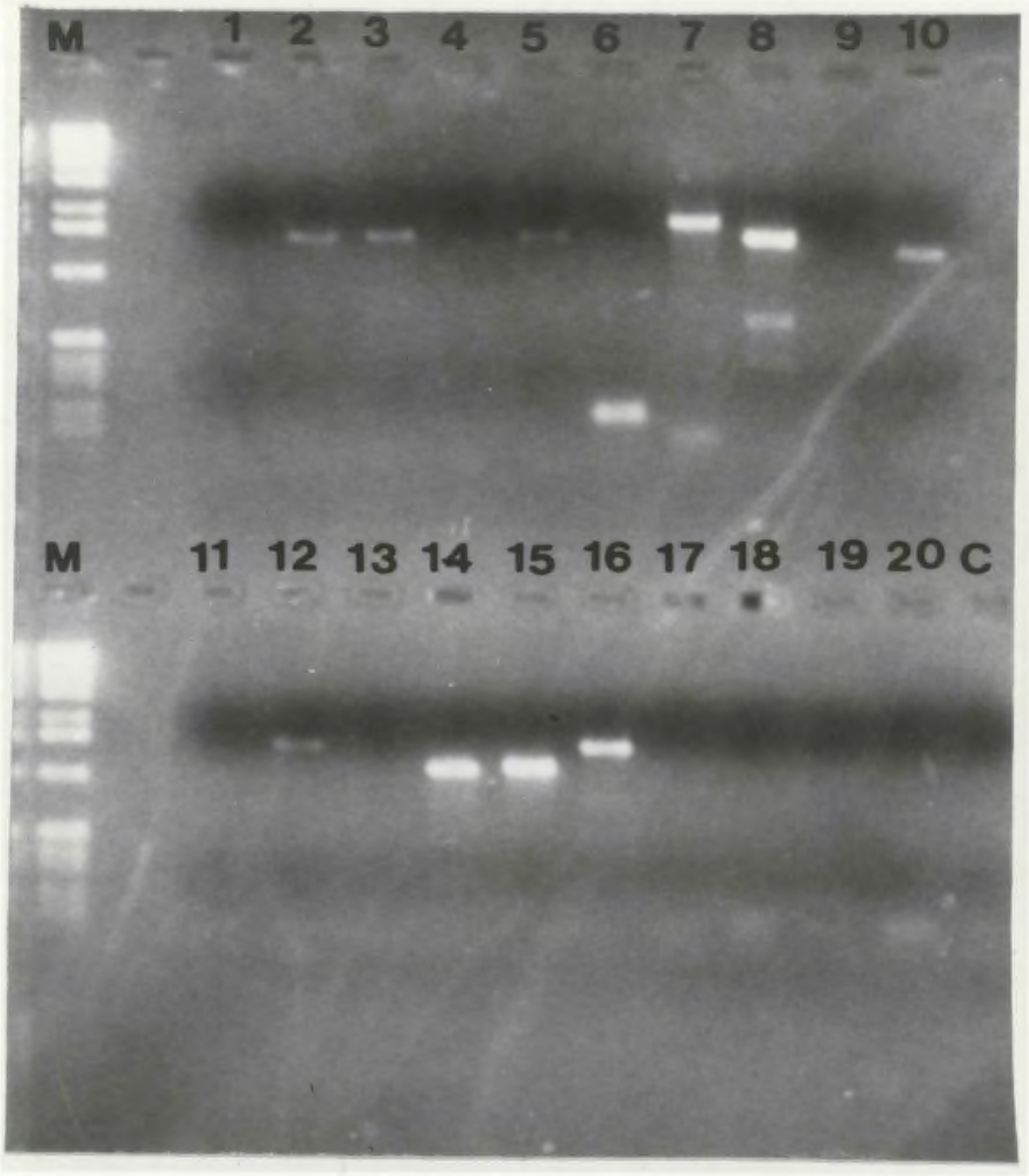
Despite several rounds of screening the λgt11 library, a full-length cDNA for the

nuclear matrix protein identified in the immunological screening was not obtained. Therefore, I decided to screen another cDNA library, which was made from HeLa cell mRNA in a λZap II vector. The λZap II library was screened by a DNA probe (*Eco*RI-*Ava*II fragment) isolated from the λ6D clone DNA (figure 3.10). The library was screened as described in section 2.2.14, and initial screening of about $2X10^5$ phages, resulted in the isolation of 55 positive clones reactive with the probe. After three rounds of rescreening, 20 strongly positive plaques were isolated and characterized further.

To determine the sizes of the inserts, the cDNA inserts from all 20 positive clones were directly amplified from purified plaques by PCR (section 3.4). Each positive phage plaque was used in PCR under the conditions described in section 3.4, except that M13 reverse and forward sequencing primers were substituted for the λgt11 primer. The binding site for the M13 forward primer in the bluescript plasmid is located 125 bp upstream from the unique *Eco*RI cloning site. Similarly, the M13 reverse primer binding site is situated 139 bp downstream from the unique *Eco*RI site. Therefore, the PCR amplified inserts contained an extra length of approximately 250 bp of vector DNA. After 35 cycles of amplification, aliquots of the PCR reactions were analyzed on an agarose gel. The DNA was visualized by staining the gel with ethidium bromide, and the size of each insert was calculated by comparing the mobility of the PCR product with those of the size markers. As shown in figure 3.13, upon ethidium bromide staining of the gel, the PCR amplified inserts were only

Fig. 3.13 Agarose gel electrophoresis of PCR-amplified DNA inserts from λ6D-selected clones.

DNA fragments were detected by staining the gel with 0.5 μg/ml ethidium bromide. Lane: 1, λ1.1 ; 2, λ3.1; 3, λ4.1 ; 4, λ5.1; 5, λ8.1; 6, λ10.1; 7, λ11.1; 8, λ13.1; 9, λ17.1; 10, λ19.1; 11, λ20.1; 12, λ21.1; 13, λ23.1; 14, λ24.1; 15, λ25.1 ; 16, λ29.1; 17, λ33.1; 18, λ35.1; 19, λ43.1 ; 20, λ55.1; c, negative control (without template DNA). Lane M represents DNA size marker (1 Kb ladder; Bethesda Research Laboratories).

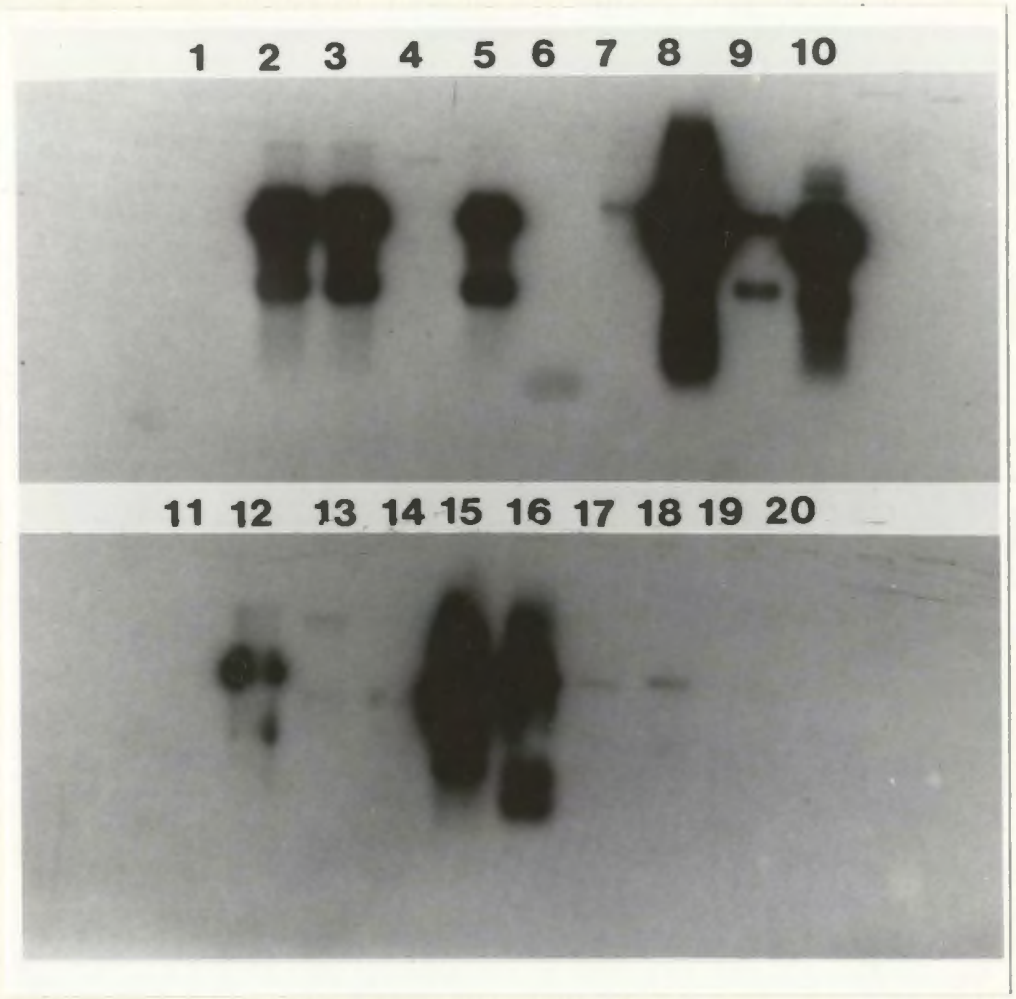

M
1
2
3
4
5
6
7
8
9
10
M
11
12
13
14
15
16
17
18
19
20
C

detected in 10 positive λZap II clones (lane 2, 3, 5, 6, 7, 8, 12, 14, 15, and 16), and in the other 9 positive clones (lane 1, 4, 9, 11, 13, 17, 18, 19, and 20) inserts were not detected, lane c served as a negative control. The sizes of inserts ranged between about 700 to 1,400 bp. In lane 10, a band which migrated at about 300 bp was most probably due to the amplification of a sequence flanking the *Eco*RI site in plasmid DNA (see above).

To determine the specificity of PCR amplified inserts, a Southern blot was prepared from PCR amplified DNA and hybridized with random primed radioactive-labelled *Eco*RI-*Ava*II fragment probe from λ6D. After 8 h, the blot was washed at high stringency (0.2 X SSC at 65°C), and exposed to film for 16 h. As shown in figure 3.14, hybridization signals were detected in lane 2, 3, 5, 8, 10, 12, 15, and 16, showing the identity between PCR amplified cDNA inserts and λ6D. In each lane, besides the major band, a faster migrating band was also detected. This faster migrating band might represent the single-strand DNA molecules generated from the PCR amplification of the insert DNA. Furthermore, in lanes 9, 17, and 18, PCR products were also detected which were not seen by ethidium bromide staining of the gel. Based on the sizes of the cDNA inserts, λ8.1 (1,263 bp) and λ17.1 (1,187 bp) were selected for the determination of nucleotide sequence.

Fig. 3.14 Hybridization of ^{32}P-labelled *Eco*RI-*Ava*II λ6D fragment with a Southern blot of cDNA inserts PCR-amplified from λ6D-selected clones.

A Southern blot prepared from the gel shown in figure 3.13 was probed with labelled *Eco*RI-*Ava*II 6D fragment. The hybridization was performed at 44°C and the blot was washed with final stringency of 0.2 X SSC at 65°C. Lane: 1, λ1.1 ; 2, λ3.1; 3, λ4.1 ; 4, λ5.1; 5, λ8.1; 6, λ10.1; 7, λ11.1; 8, λ13.1; 9, λ17.1; 10, λ19.1; 11, λ20.1; 12, λ21.1; 13, λ23.1; 14, λ24.1; 15, λ25.1 ; 16, λ29.1; 17, λ33.1; 18, λ35.1; 19, λ43.1 ; 20, λ55.1.


1 2 3 4 5 6 7 8 9 10
11 12 13 14 15 16 17 18 19 20

3.9 NUCLEOTIDE SEQUENCING OF λ6D-SELECTED CLONES

Prior to nucleotide sequencing, the bluescript plasmid containing the inserts from λ8.1 and λ17.1 were excised from the λZap II vector DNA *in vivo* with the help of helper phage R 408 as described in section 2.2.21. The bacteria harbouring the *in vivo* excised plasmids were then grown in large scale culture to prepare the excised plasmid DNA for sequencing. To determine the complete nucleotide sequence for λ8.1 and λ17.1 DNA, undirectional deletions were created in the insert DNA using Exonuclease III (section 2.2.23.1). The resulting subclones were then used to make plasmid preparations for sequencing. The nucleotide sequences of the two overlapping clones λ8.1 and λ17.1 was determined using the sequence strategy shown in figure 3.15. The cDNAs, which span a total of 2,171 nucleotides, were sequenced completely. A large open reading frame encoding a polypeptide of 422 amino acids was found, starting with an ATG translation initiation codon at nucleotide 26 and terminating with a stop codon TAA at nucleotide 1291. Since the open reading frame was not preceded by any in-frame stop codon at any position, the position of the initiation codon at nucleotide 26 is tentative. However, the calculated molecular weight 49.8 kDa for the polypeptide deduced from the sequence data is very close to the size of the protein, identified by western blot analysis (see section 3.2.4.3). The 3' end of the cloned cDNA did not contain a polyadenylation consensus sequence (AATAAA), or a poly A tail. Furthermore, the major messenger RNA in

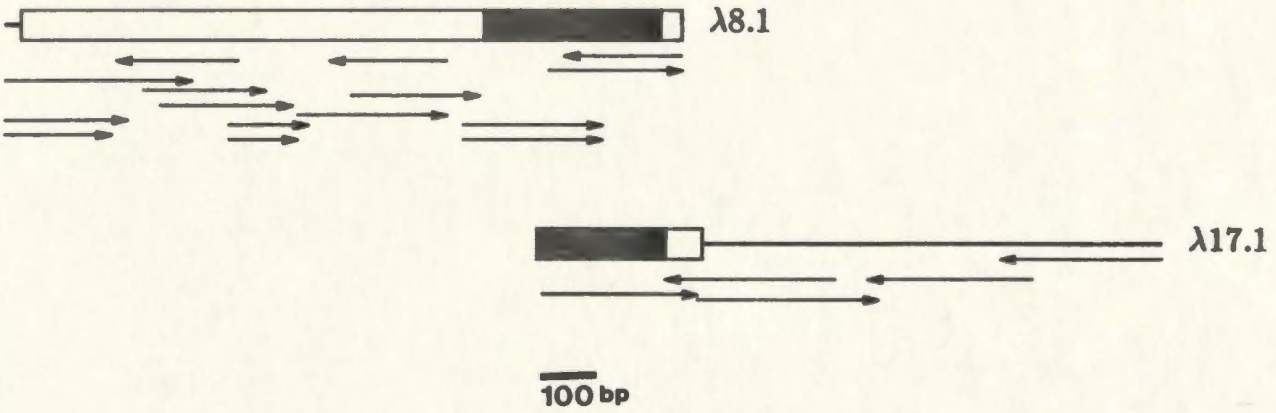
Fig. 3.15 Combined nucleotide sequence of λ8.1 and λ17.1 clones with deduced amino acid sequence of the largest open reading frame.

(A) The nucleotide sequence is presented in the 5' to 3' direction starting with the initiation ATG as number 1. The deduced amino acid residues are numbered beginning with the initiation methionine and an asterisk sign in the protein sequence indicates the end of reading frame. Potential sites for post translational modification, their consensus sequences and their locations are: glycosylation site, ^{109}N (N-{P}-S/T-{P}); protein kinase C phosphorylation sites, ^{59}S, ^{169}S, ^{175}T, ^{203}T, ^{316}T, ^{319}T, ^{328}T, ^{348}T, (S/T-X-X-R/K); cAMP- and cGMP-dependent protein kinase sites ^{316}T, ^{317}T, ^{328}T, ^{356}S, ^{374}T, (R/K-R/K-X-S/T); casein kinase II sites, ^{69}S, ^{90}T, ^{129}S, ^{144}S, (S/T-X-X-D/E); and a nuclear targeting signal ^{390}R^{391}K^{392}K^{393}N^{394}K, (R/K/T/A-K-K-R/Q/N/T/S/G-K). In the above consensus sequence, the symbol 'X' is used for a position where any amino acid is accepted. Any amino acid that is not accepted at a given position is shown within a pair of curly brackets. The region of sequence similarity to human HMG-1 protein is also shown (underlined region, residues 45-166). (B) Schematic representation of the size and location of the two overlapping cDNA clones sequenced. The horizontal bar represents the coding region, the filled area within the horizontal bar indicates the sequence identity to 7Eb DNA. The solid line represents untranslated sequences and arrows indicate the direction and approximate position of the regions sequenced.

```
                                                    GGAATTCCCAGCTCCTGCAACACAA   -1

   ATGCCCAGCACACCAGGGTTTGTGGGATACAATCCATACAGTCATCTCGCCTACAACAAC    60
 1 M  P  S  T  P  G  F  V  G  Y  N  P  Y  S  H  L  A  Y  N  N

   TACAGGCTGGGAGGGAACCCGGGCACCAACAGCCGGGTCACCGGATCCTCTGGTATCACG   120
21 Y  R  L  G  G  N  P  G  T  N  S  R  V  T  G  S  S  G  I  T

   ATTCCAAAACCCCCAAAGCCACCAGATAAGCCGCTGATGCCCTACATGAGGTACAGCAGA   180
41 I  P  K  P  P  K  P  P  D  K  P  L  M  P  Y  M  R  Y  S  R

   AAGGTCTGGGACCAAGTAAAGGCTTCCAACCCTGACCTAAAGTTGTGGGAGATTGGCAAG   240
61 K  V  W  D  Q  V  K  A  S  N  P  D  L  K  L  W  E  I  G  K

   ATTATTGGTGGCATGTGGCGAGATCTCACTGATGAAGAAAAACAAGAATATTTAAACGAA   300
81 I  I  G  G  M  W  R  D  L  T  D  E  E  K  Q  E  Y  L  N  E

   TACGAAGCAGAAAAGATAGAGTACAATGAATCTATGAAGGCCTATCATAATTCCCCCGCG   360
111 Y  E  A  E  K  I  E  Y  N  E  S  M  K  A  Y  H  N  S  P  A

   TACCTTGCTTACATAAATGCAAAAAGTCGTGCAGAAGCTGCTTTAGAGGAAGAAAGTCGA   420
121 Y  L  A  Y  I  N  A  K  S  R  A  E  A  A  L  E  E  E  S  R

   CAGAGACAATCTCGCATGGAGAAAGGAGAACCGTACATGAGCATTCAGCCTGCTGAAGAT   480
141 Q  R  Q  S  R  M  E  K  G  E  P  Y  M  S  I  Q  P  A  E  D

   CCAGATGATTATGATGATGGCTTTTCAATGAAGCATACAGCCACCGCCCGTTTCCAGAGA   540
161 P  D  D  Y  D  D  G  F  S  M  K  H  T  A  T  A  R  F  Q  R

   AACCACCGCCTCATCAGTGAAATTCTTAGTGAGAGTGTGGTGCCAGACGTTCGGTCAGTT   600
181 N  H  R  L  I  S  E  I  L  S  E  S  V  V  P  D  V  R  S  V

   GTCACAACAGCTAGAATGCAGGTCCTCAAACGGCAGGTCCAGTCCTTAATGGTTCATCAG   660
201 V  T  T  A  R  M  Q  V  L  K  R  Q  V  Q  S  L  M  V  H  Q

   CGAAAACTAGAAGCTGAACTTCTTCAAATAGAGGAACGACACCAGGAGAAGAAGAGGAAA   720
221 R  K  L  E  A  E  L  L  Q  I  E  E  R  H  Q  E  K  K  R  K

   TTCCTGGAAAGCACAGATTCATTTAACAATGAACTTAAAAGGTTGTGCGGTCTGAAAGTA   780
241 F  L  E  S  T  D  S  F  N  N  E  L  K  R  L  C  G  L  K  V

   GAAGTGGATATGGAGAAAATTGCAGCTGAGATTGCACAGGCAGAGGAACAGGCCCGCAAA   840
261 E  V  D  M  E  K  I  A  A  E  I  A  Q  A  E  E  Q  A  R  K

   AGGCAGGAGGAAAGGAGAAGGAGGCCGCAGAGCAAGCTGAGCGCAGTCAGAGCAGCATCG   900
281 R  Q  E  E  R  R  R  R  P  Q  S  K  L  S  A  V  R  A  A  S

   TTCCTGAGGAAGAACAAGCAGCTAACAAAGGCGAGGAGAAGAAAGACGACGAGAACATTC   960
301 F  L  R  K  N  K  Q  L  T  K  A  R  R  R  K  T  T  R  T  F
   CGATGGAGACAGAGGAGACACACCTTGAAGAAACAACAGAGAGCCAACAGAATGGTGAAG  1020
321 R  W  R  Q  R  R  H  T  L  K  K  Q  Q  R  A  N  R  M  V  K

   AAGGCACGTCTACTCCTGAGGACAAGGAGAGTGGGCAGGAGGGGGTCGACAGTATGGCAG  1080
341 K  A  R  L  L  L  R  T  R  R  V  G  R  R  G  S  T  V  W  Q

   AGGAAGGAACCAGTGATAGTAACACTGGCTCGGAGAGCAACAGTGCAACAGTGGAGGAGC  1140
361 R  K  E  P  V  I  V  T  L  A  R  R  A  T  V  Q  Q  W  R  S
```

```
    CACCAACAGATCCCATACCAGAAGATGAGAAAAAAGAATAAGTGTTGCCTTGTTTTGTGT 1200
381  H  Q  Q  I  P  Y  Q  K  M  R  K  K  N  K  C  C  L  V  L  C

    GTTCTAAATACTTTTTTTAATGAAAAAATGTTTTTTGGTTTTAATGGTGTTACGTGGTTT 1260
401  V  L  N  T  F  F  N  E  K  M  F  F  G  F  N  G  V  T  W  F

    GTGTATTAATTTTTTTTCTTGTCCATATCATACCACCAAAGGCTTTTGGACCATTTAGCA 1320
421  V  Y  *

   TCATGAGCCTAATGGCTCAGTCAGTCACCTTTCTTAAGTGTTGTGAAGATGGCTCTTTTC  1380
   TTTGGATCTTGTTTCTAGCCCTCAACTGCTGAAAGCCTCAGAATTTAGATTAATTGAGAA  1440
   AACACCCACCTCTTTTAGAGAATTATCCTTTGATGCTGCAGAATCTACTCTTACAATGCC  1500
   TTCCTACAGCTCACTGGGGTGCCCCAAAGCCATAGCTTTAAACCTTCCCAGTCCCCATCA  1560
   GAGCTTCCTGAAAGTCTCCTCTCTTGTTTACTTCTGCAAAGGGTAGCTTCTTAAAAACGT  1620
   GATCATGTATGAGTATGTATTTGTTCACTTACCCTTTTTTTACTTTTAATCAATGTCAGA  1680
   TACCAAGAGTTGTGTTAAGCTGAGTGTAGTATGTAACTAACTACACTTGGATCTTACTGA  1740
   TCCAGAAATAGTCCCCATAGTTAGAGTAGTTACTTATGAAGTGGTTATTAAAGTGAACAA  1800
   CAGCACATATACATTATCTATACTGCTTTTTGTTATGATTAATACGGGTGTTCTTAATAT  1860
   TCTTGTAATGGACTAGATAATTCAAACTGATTAGCCCATTCCAGAAGAAAAACAGCTGGG  1920
   AATTAAGTTAATCCACTTGAAATTGTTTTACAATAATCAGACATCCAAACCTCAAGGCTC  1980
   AGGATCCCATAGACCAGAGTTTTCCTTTTTGATAAACTTAGTAAAGTCTTGGAGACTAGA  2040
   AGCAAGATAGTTTGTGACACATAAGCTTCCTCGAAAAACTAGAATAGATTTTTACTGATA  2100
   GTGGTATATCTGATGGTATATGTTTCTTAAGGTCCGGAATTCCGGA                2146
```

(B)



HeLa cells detected by both cDNAs λ8.1 and λ17.1 was about 2.5 kb as determined by Northern blot analysis (see section 3.11). This indicates that the combined length of cDNAs λ8.1 and λ17.1 may be 300-400 bp smaller than the full-length polyadenylated message.

3.10 CHARACTERISTICS OF THE 7Eb PROTEIN

According to the deduced amino acid sequences, the combined sequences from λ8.1 and λ17.1 cDNA inserts are capable of encoding a polypeptide having 422 amino acid residues with a molecular weight of 49.8 kDa and an isoelectric point of 10.8.

A sequence identity search of the Swiss protein database (Release 20) with the deduced amino acid sequence revealed significant sequence similarity with several HMG Box domain containing proteins. These included a nuclear transcription factor (hUBF, 20% identity in 80 amino acid overlap), yeast nonhistone chromosomal protein 6B (33.7% identity in 77 amino acid overlap), yeast nonhistone chromosomal protein 6A (33.7% identity in 77 amino acid overlap), nonhistone chromosomal protein LG-1 (23.8% identity in 80 amino acid overlap), HMG-1 protein from rat (21.3% identity in 80 amino acid overlap), bovine (22.5% identity in 80 amino acid overlap), and human (22.5% identity in 80 amino acid overlap). The predicted protein also showed similarity with various cytoskeletal proteins. These include:

myosin heavy chain (20.3% identity in 192 amino acid overlap), dystrophin (14.2% identity in 218 amino acid overlap), bovine cardiac muscle troponin T isoform 2 (18.1% identity in 226 amino acid overlap), bovine troponin T isoform 1 (20.1% identity in 219 amino acid overlap), chicken neurofilament M (17.8% identity in 163 amino acid overlap), rat neurofilament M (17.0% identity in 165 amino acid overlap), human neurofilament M (17.1% identity in 158 amino acid overlap) and sheep trichohyalin (18.3% identity in 338 amino acid overlap).

The predicted protein contains many possible sites of posttranslational modification. One putative glycosylation site was identified at residue 164. Several potential phosphorylation sites were identified (figure 3.15), including those for cAMP- and cGMP-dependent protein kinase (at residues 316, 317, 328, 356, and 374) protein kinase C (at residues 11, 59, 169, 175, 203, 316, 319, 328, and 348) and casein kinase II (at residues 69, 90, 129, and 144). Moreover, a consensus sequence for a nuclear localization signal was also found at amino acid residues 390-394.

3.11 NORTHERN BLOT ANALYSIS

To investigate the mRNA species corresponding to cDNAs isolated in this study, poly $(A)^+$ mRNA from HeLa cells was prepared as described in section 2.2.25, fractionated in a denaturing formaldehyde agarose gel, and blotted on to a Hybond-N membrane. The blot was subsequently hybridized with cDNA inserts radioactively-

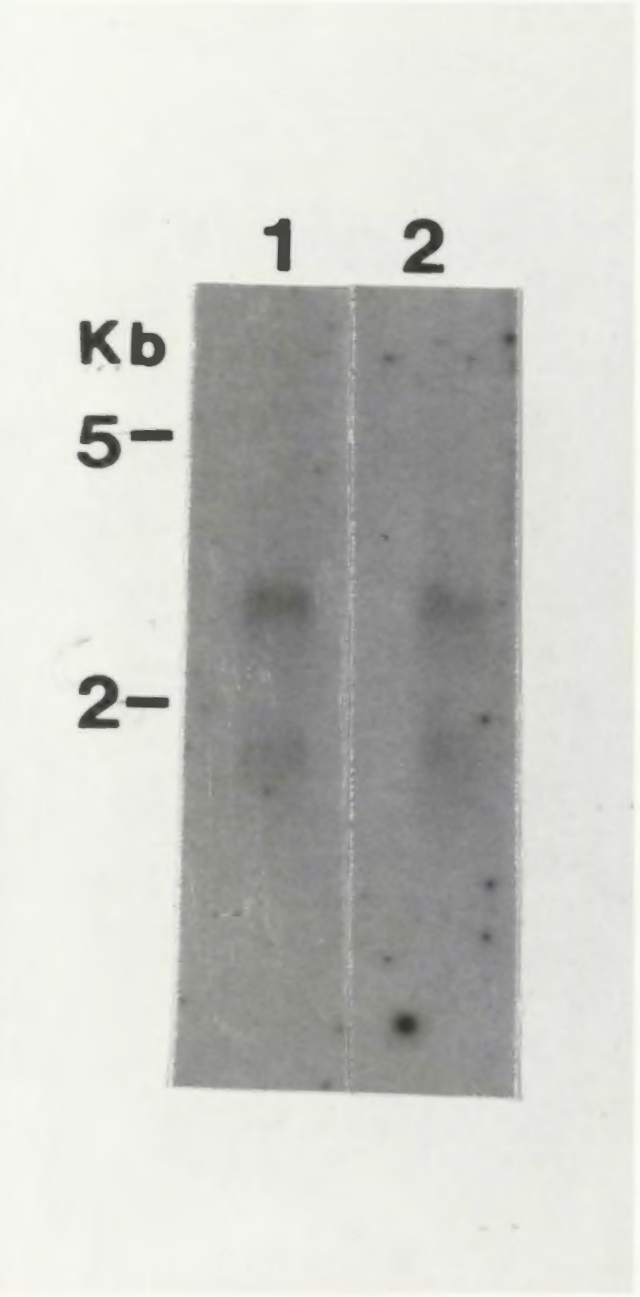
labelled by random priming (figure 3.16). The λ8.1 cDNA insert hybridized with a major band (lane 1), and compared with the positions of 18S and 28S rRNA, this mRNA migrated with an apparent size of approximately 2.5 kb. In addition, a minor band with an apparent size of 1.7 kb was also observed. Identical results were obtained when the blot was probed with either λ17.1 (lane 2) or 7Eb DNA (not shown). These results suggested that the full-length message which encodes the 7Eb protein (identified by immunoscreening) contained approximately 300-400 additional nucleotides (not present in composite sequence; figure 3.15) representing upstream and/or downstream untranslated sequences including a poly-A tail. The minor species of mRNA observed may represent an alternately spliced mRNA, a mRNA representing a related gene, or less likely a specific degradation product.

3.12 Primer extension analysis

Primer extension analysis was used to determine the length of 5' untranslated region of 7Eb protein. An oligonucleotide primer, p31, was synthesized. This primer was complementary to the nucleotide sequence of the cDNA clone λ8.1 containing the 5' most sequence of the 7Eb protein (see figure 3.15). For the oligonucleotide sequence, nucleotide -14 to +6 were selected. When oligonucleotide p31 was used in the primer extension reaction of HeLa cell poly $(A)^+$ and total RNA, two major bands of extended products were detected. The sizes of the major

Fig. 3.16 Northern blot analysis of HeLa cells RNA.

Poly(A)$^+$ RNA from HeLa cells was electrophoresed on a 1% formaldehyde agarose gel, blotted on to a nylon membrane (Hybon-N), and probed with radioactively-labelled insert from λ8.1 (lane 10) or with insert from λ17.1 (lane 2). Each lane contained 5 μg of poly(A)$^+$ RNA. The membrane was washed to a final stringency of 0.2 X SSC at 68°C. The numbers 5 and 2 Kb represent the positions of the 28S and 18S ribosomal RNA, respectively.


1
2
Kb
5–
2–

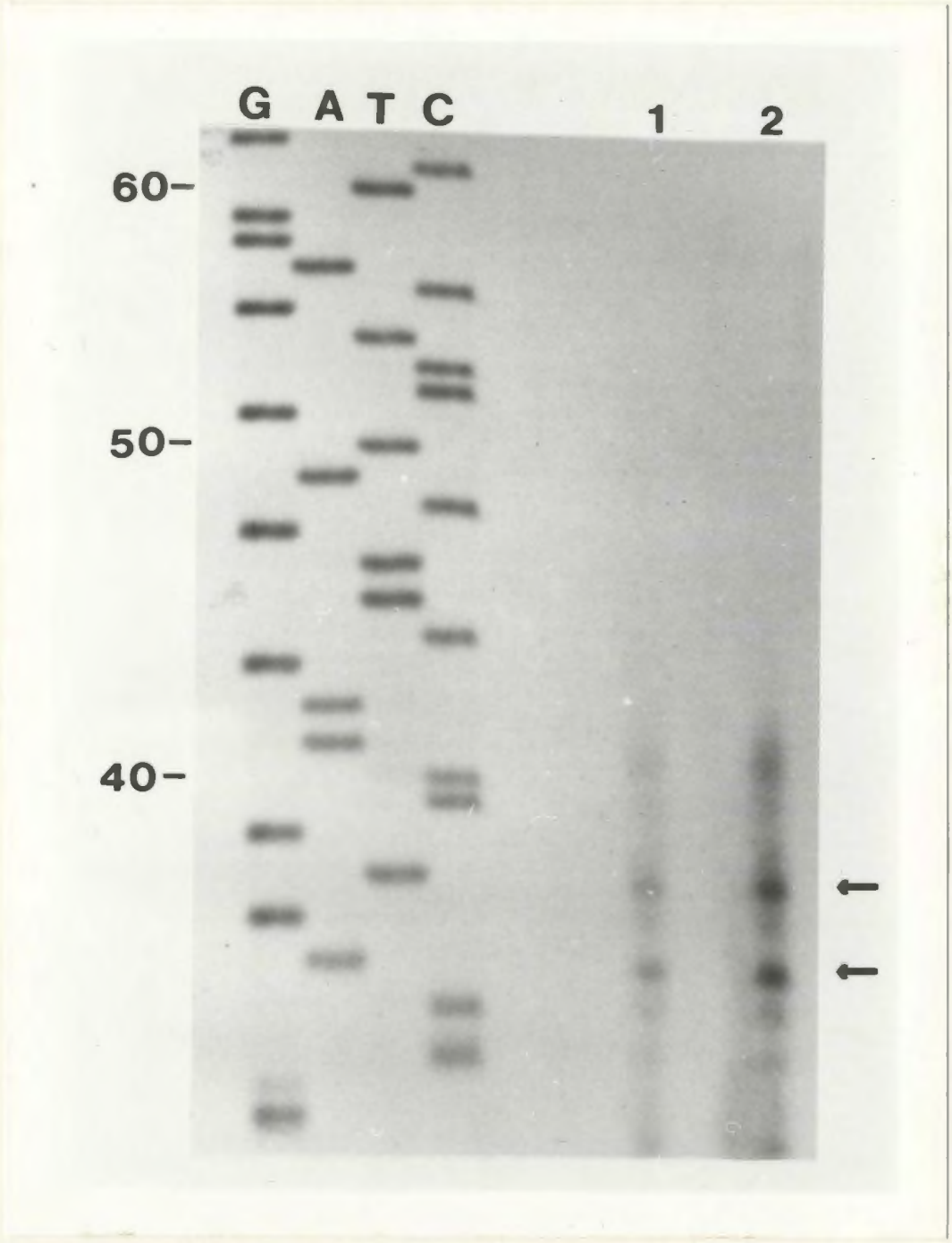
extended product was found to be 35 and 37 nucleotides (figure 3.17) by comparing with the M13 sequencing ladder run on the same gel. A faint band of about 41 nucleotides was also detected. Based upon the λ8.1 cDNA nucleotide sequence, the size of the extended product was expected to be at least 31 nucleotides long. This size of 31 nucleotides includes 20 nucleotides of the primer and 11 nucleotides of the cDNA present 5' to the sequence of the primer. Since the size of major extended products was 35 or 37 nucleotides, it is inferred that I have isolated a cDNA clone which contains most of the 5' end of the mRNA. The nucleotide sequence depicted in figure 3.15 lacks only 4 or 6 nucleotide from the 5' end of the mRNA encoding the 7Eb protein. If the faint band of 41 nucleotides is considered as a true band representing fully extended product, then the cDNA (figure 3.15) lacks 10 nucleotides. The bands in the lower part of the gel probably represent the often observed premature termination of AMV reverse transcriptase *in vitro*. The other less likely possiblity is that there are multiple start sites of the gene encoding the mRNA for 7Eb protein.

3.13 GENOMIC SOUTHERN BLOT ANALYSIS

To learn about the genomic organization and complexity of the gene encoding the 7Eb protein, a genomic Southern blot analysis was performed as described in section 2.2.26. High molecular weight human DNA was isolated and digested separately

Fig. 3.17 Primer extension analysis of HeLa cells RNA

Oligonucleotide p31 was used as a primer to extend HeLa cells 50 μg total (lane 1) or 5 μg poly(A)$^+$ RNA (lane 2). The reaction products were analyzed on a 6% polyacrylamide gel containing 7 M urea. The dried gel was exposed to X-ray film for 2 days. The numbers on the left of lane indicate the positions that products containing 30 and 60 bases would be found based on sequencing reactions (lanes G, A, T, and C) which used a M13 single-stranded DNA. The suggested positions of the full-length, primer-extended product are indicated by arrows.


G
A
T
C
1
2
60–
50–
40–

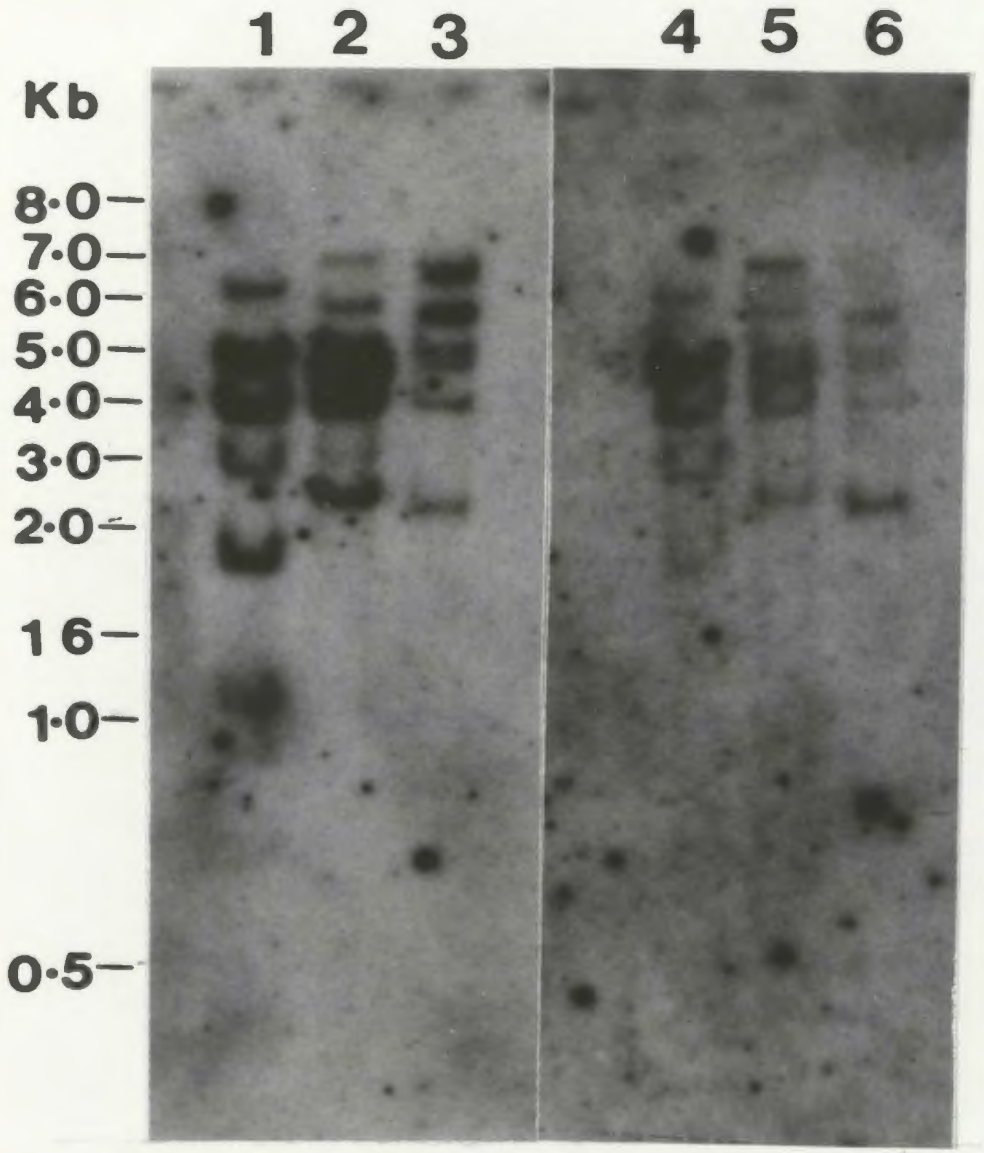
with three restriction endonucleases, *Eco*RI, *Bam*HI, and *Hind*III. Restriction endonuclease digested human genomic DNA was resolved on a 1% agarose gel in duplicate and then blotted on to a Hybond-N membrane. The blot containing one set of digested genomic DNA was hybridized with a radioactively-labelled λ8.1 cDNA insert and the duplicate blot probed with the λ17.1 cDNA insert. After 24 h of hybridization at 42°C, the blots were washed at high stringency (0.2 X SSC at 68°C), and exposed to X-ray film for 24 h. As shown in figure 3.18, hybrization of λ8.1 to restriction endonuclease digests of human DNA revealed a small number of bands. With *Hind*III digestion, 5 bands were seen (lane 1) ranging from ~1.8 to 6.5 kbp in size. *Eco*RI digestion generated 7 bands (Lane 2) ranging from ~ 2.5 to 7.0 kbp in size. With *Bam*HI digestion, 5 bands were seen (lane 3) in the size of between ~2.0 - 7.0 kbp. Almost identical results were obtained when the blot was probed with the λ17.1 cDNA insert. The combined sizes of these fragments amounts to about 25-30 kbp. This suggests that the cDNA is encoded either by a single gene or else it is a member of a small gene family.

3.14 Expression of protein 7Eb in different tumor cell lines

To investigate the expression of protein 7Eb in various cell lines, total (human epidermoid carcinoma A-431, breast carcinoma cells MDA MB-330, lung carcinoma BEN, and ovarian adenocarcinoma HEY) or poly $(A)^+$ RNA (cervical carcinoma

Fig. 3.18 Genomic Southern blot analysis.

Human genomic DNA was digested with restriction endonulceases and was then electrophoresed on a 1% agarose gel, blotted on to a nylon membrane (Hybond-N), and probed with radioactively-labelled insert from λ8.1 (lane 1-3) or with insert from λ17.1 (lane 4-6). Lane: 1 and 4 , *Bam*HI; 2 and 5, *Eco*RI; 3 and 6, *Hind*III. The membrane was washed to a final stringency of 0.2 X SSC, at 65°C and autoradiograph was exposed for 72 h.


1 2 3 4 5 6
Kb
8·0
7·0
6·0
5·0
4·0
3·0
2·0
1 6
1·0
0·5

HeLa S3, T-cell leukemia Jurkat, colon adenocarcinoma SW-620) was analysed by Northern blot analysis (section 2.2.25). The cDNA insert from clone λ8.1 was used as a probe for this analysis. Probing of the Northern blot of the above mentioned cell line RNA detected two bands of ~ sizes 2.5 and 1.7 kb; these sizes are similar to those detected in HeLa cells (section 3.11; figure 3.16, lane 1). In all of the cell lines except the colon tumor line, SW-620 (figure 3.19, lane 7) the 2.5 kb band was darker than the 1.7 kb band. Detection of mRNA of sizes similar to those obtained in HeLa cells suggest that the gene for protein 7Eb is expressed in these cell lines. To investigate this further, a western blot analysis was performed. Total nuclear proteins prepared from seven different cell lines were separated by SDS polyacrylamide gel electrophoresis. A western blot was prepared and probed with affinity-purified anti 7E β gal-fusion protein rabbit antibodies (RA-7E) (section 2.2.25). A single immunoreactive band was detected in each lane (figure 3.20). The approximate molecular weight of the immunoreactive protein was calculated to be 52 kDa, which was calculated by comparing the mobility of size markers ran in a parallel lane on the same gel. A band of the same size was also detected in the lane containing HeLa cells nuclear protein fraction. The detection of a similar size protein in different cell lines representing distinct tissue types suggests that the protein 7Eb is probably ubiquitous in nature and expressed by many different tissue types.

Figure 3.19 Northern blot analysis of different cell lines RNA.

Total (epidermoid carcinoma A-431, breast carcinoma MDA MB-330, lung carcinoma BEN, ovarian adenocarcinoma HEY; lane 2, 3, 5, and 6, respectively) or poly $(A)^+$ RNA (cervical carcinoma HeLa S3, T-cell leukaemia Jurkat, colon adenocarcinoma SW-620, lane 1, 4, and 7, respectively) were electrophoresed on a 0.7% formaladehyde agarose gel. The gel was blotted on to a nylon membrane (Zetaprobe), and probed with probed with radioactively-labelled insert from λ8.1. Each lane contained either 25 μg (total) or 5 μg (poly$(A)^+$ RNA). The membrane was washed to a final stringency of 0.2 X SSC at 62°C. The numbers 5 and 2 Kb represent the positions of the 28S and 18S ribosomal RNA, respectively.

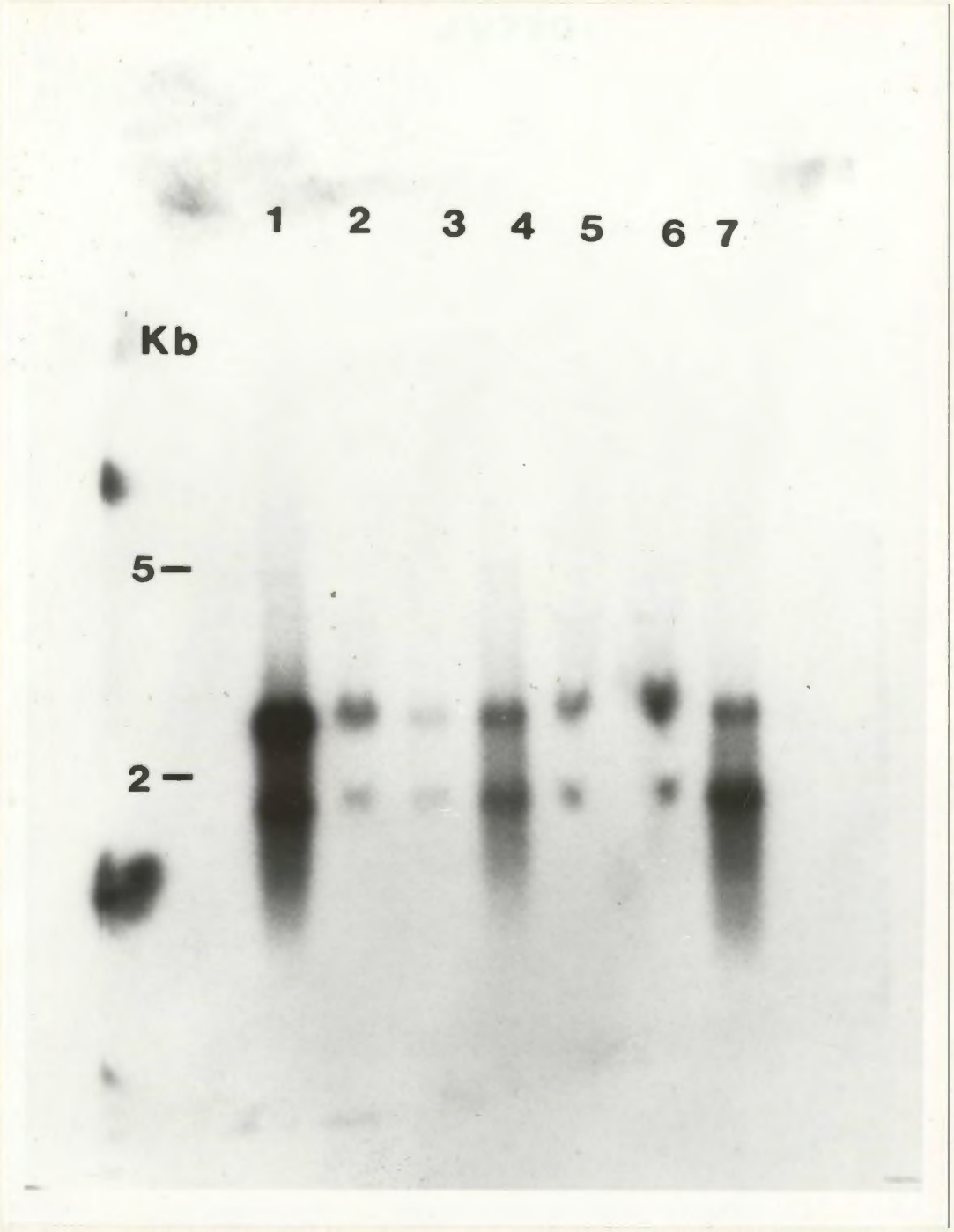
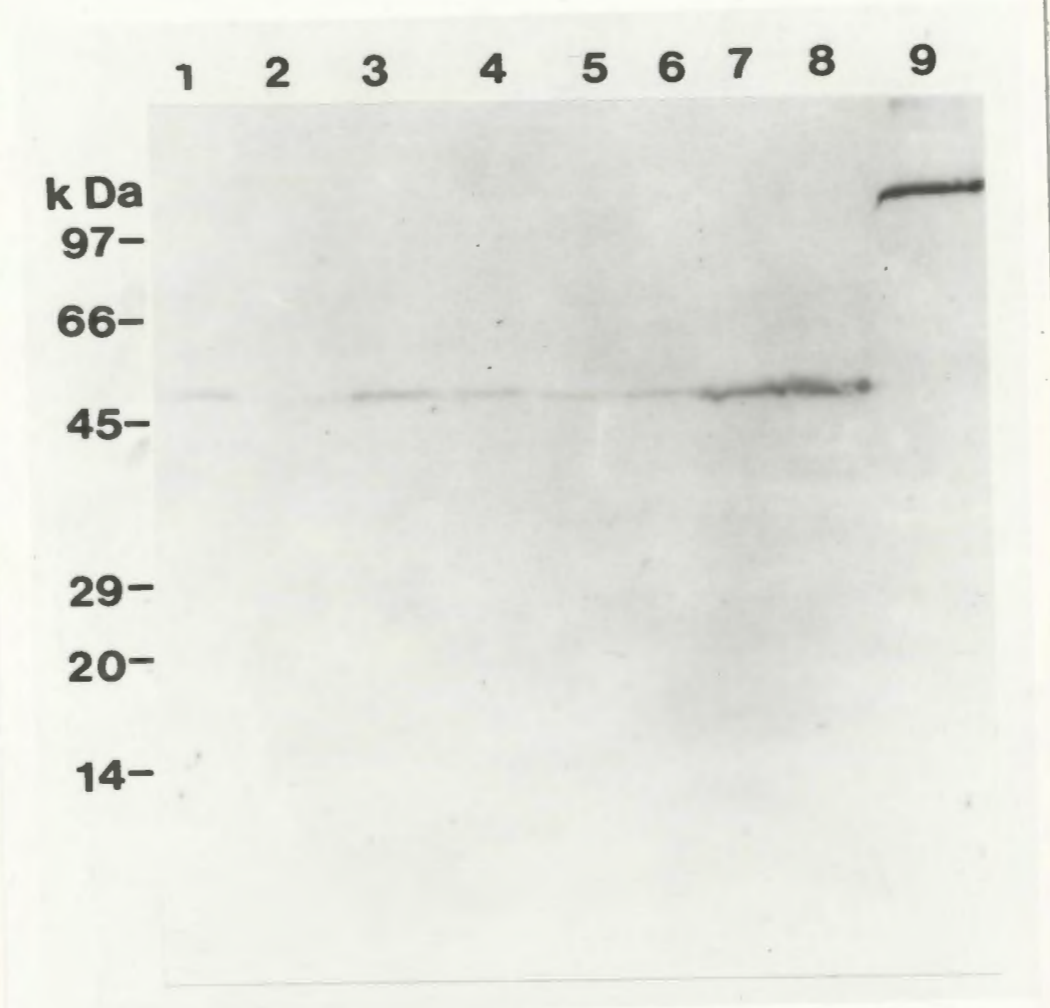

1 2 3 4 5 6 7
Kb
5—
2—

Fig. 3.20 Expression of 7Eb protein in different cell lines.

Nuclear proteins prepared from HeLa S3 (lane 1), A-431 (lane 2), MDA MB-330 (lane 3), Jurkat (lane 4), BEN (lane 5), HEY (lane 6), SW-620 (lane 7), and SW-480 (lane 8), cells and resolved on a 10% SDS polyacrylamide gel. The separated proteins were electrophoretically transferred to a nitrocellulose membrane and probed with the RA-7E. Lane 9 contains affinity purified 7E b gal-fusion protein. The numbers on the left of the lane represent approximate positions of molecular weight standards.


1
2
3
4
5
6
7
8
9
k Da
97–
66–
45–
29–
20–
14–

CHAPTER 4

DISCUSSION

An immunological approach was used to identify cDNA clones encoding protein components of the nuclear matrix. A nuclear matrix fraction was obtained from HeLa cells and a polyclonal antiserum was generated in a rabbit (section 3.1.2). This polyclonal antiserum recognized several bands on an immunoblot that were only present in the nuclear matrix fraction and were absent from the cytoplasmic fraction. This complex rabbit antiserum was used to screen a human fibroblast cDNA expression library in λgt11, which resulted in the isolation of 12 positive λgt11 clones reactive with rabbit antiserum (section 3.1.3). To identify the proteins encoded by the cDNA clones, rabbits were immunized with five individual β gal-fusion proteins. Three rabbits produced antibodies which recognized the β gal-fusion protein on a western blot, the remaining two did not produce antibodies recognizing β gal-fusion protein (section 3.1.4). To determine the location of the protein encoded by cDNAs in the cell, immunofluorescent staining of HeLa cell was carried out; despite many attempts, results of these experiments remained inconclusive. Hence these λgt11 clones were not characterized further. It is possible that these λgt11 clones may still be valuable as nuclear matrix protein expressing cDNAs since they

were initially selected based on reactivity with rabbit antiserum raised against nuclear matrix proteins, hence a further improvement of immunofluorescent staining technique may provide conclusive evidence regarding the nuclear matrix origin of the proteins, encoded by the cDNAs isolated in this study.

As a second attempt to isolate a cDNA for a nuclear matrix protein, the same cDNA library was probed with another polyclonal antibody (CRLM). Screening of the library with CRLM resulted in the isolation of 13 positive λgt11 clones expressing proteins recognized by CRLM (section 3.2.2). Four strongly positive immunoreactive clones, λ7A, λ7E, λ8D, and λ8H were selected for further study. Antisera raised against the β gal-fusion proteins produced by the four clones which reacted with their respective fusion proteins in an immunoblot assay (section 3.2.3.2), showed that the fusion protein from λ7E reacted strongly with the antiserum raised against it. Therefore, of these four clones, I have characterized the λ7E cDNA clone in detail.

4.1 IDENTIFICATION OF PROTEIN ENCODED BY λ7E

For the validation of λ7E as nuclear matrix protein expressing clone, antiserum to 7E β gal-fusion protein was raised in a rabbit and used to probe a western blot of HeLa cell proteins (section 3.2.4.2 and 3.2.4.3). The affinity purified 7E β gal-

fusion protein specific antibody (RA-7E) recognized a protein in the nuclear fraction (section 3.2.4.3), which was undetectable in the cytoplasmic fraction. This suggested that the λ7E clone contained a part of a cDNA for a nuclear protein. This conclusion was supported by immunofluorescent staining of HeLa cells (section 3.2.4.3). The same preparation of antiserum stained only the nuclei, excluding cytoplasm. The approximate molecular weight of the immunoreactive protein on a western blot was calculated to be 52 kDa. Immunoblot analysis using various subcellular fractions that were obtained during nuclear matrix preparation from HeLa cells showed that the 52 kDa protein identified by the antiserum against 7E β gal-fusion protein is a component of the nuclear matrix fraction (section 3.2.4.6). The identification of the protein encoded by this cDNA as a nuclear protein was further substantiated by the sequence data (see below), which identified a nuclear localization signal sequence and a DNA binding consensus sequence.

4.2 ISOLATION AND CHARACTERIZATION OF LONGER cDNAs

The λ7E clone contained two *Eco*RI fragments, 7Ea and 7Eb of 948 bp and 360 bp (section 3.3). 7Eb had one open reading frame which was adjacent to and in phase with the β galactosidase coding sequence in λgt11. 7Ea contained several stop codons. The open reading frame in λ7E could not possibly code for a 52 kDa

protein (it was only 360 bases long). Thus, a further round of screening of the HeLa cell λgt11 library was carried out using 7Eb DNA as a probe (section 3.4). This screening resulted in the isolation of five clones containing longer cDNA inserts. Based upon sequence data, the cDNAs clones were classified into two types: one type contained almost the entire 7Eb sequence except 27 bp at the 5'end of the 7Eb sequence, the other type had only 276 bp of sequence similar to 7Eb DNA (refer to section 3.5, figure 3.10). Outside the region of sequence identity with 7Eb, the rest of the sequence did not have sequence identity to other cDNAs isolated in this round of screening (refer to figure 3.10). Neither of the two types of cDNAs described above showed any sequence similarity to 7Ea DNA. These results supported the earlier suggestion that the 7Ea and 7Eb are two independent cDNA clones ligated into a single λgt11 molecule. It is possible that the extra 27 bases at the 5' end of 7Eb DNA, not found in many other clones I have isolated so far, could have originated from a cloning artifact. It is also possible that those extra bases represent alternately spliced mRNA or an intron sequence which was not excised from the mRNA. These possibilities could be resolved by a detailed study of genomic clones corresponding to this mRNA.

The largest insert-containing clone isolated in this round of screening was λ1B. It contained 947 bp and a single open reading frame of 855 nucleotides encoding 285 amino acid residues. The open reading frame began at nucleotide 93 with an initiation codon ATG, and remained open throughout the λ1B insert DNA. The 7Eb sequence

was located in λ1B between nucleotide 634 and 910.

As mentioned earlier (section 3.6), while the sequencing of λ1B was in progress, I rescreened the HeLa cell λgt11 library using a 242 bp *Eco*RI-*Xho*II fragment isolated from the 5' end of λ1B DNA. This screening resulted in the isolation of clones containing longer cDNA inserts. A partial nucleotide sequence was determined for each clone. A comparison of λ1B with these sequences revealed sequence identity of 64 bp at the 5' end of each clone with nucleotides 12 to 75 of λ1B. The remainder of the DNA sequence showed no identity to λ1B. It was concluded that the clones isolated in this latest round of screening were unrelated to λ1B and hence were not characterized further. Recently a sequence identity search performed on EMBL DNA sequence data bank (release 35.0) revealed complete identity between one of these clones, λ5h and the human mitochondrial DNA. Furthermore, all sequences determined on this series of cDNA clones (λ5a to λ5h) were also found to be almost identical to human mitochondrial DNA.

Screening of another human cDNA library (HeLa cell λZap II, section 3.8) with a fragment isolated from λ6D (a 7Eb selected clone containing a 318 bp region identical to 7Eb DNA (section 3.4 and 3.5) resulted in the isolation of longer cDNAs. The nucleotide sequence for two overlapping cDNA clones, λ8.1 and λ17.1 containing 1263 and 1187 bp, respectively (appendices 12 and 13) spaning a total of 2,171 bp was determined completely (figure 3.15). The composite sequence obtained from the λ8.1 and λ17.1 cDNA clones revealed a large open reading frame encoding

a polypeptide of 422 amino acids. The open reading frame beginning at nucleotide 26 of λ8.1 continued to a stop codon, TAA, at nucleotide 1291 of λ17.1 (figure 3.15). In this sequence, nucleotides between 901 to 1218 (amino acid residues 293 to 398) are identical to the 7Eb sequence. Similarly, nucleotides 339 to 1177 (amino acid residues 112 to 384) are the same as in the λ1B. Since the open reading frame was not preceded by an in frame stop codon, the actual initiation site for this protein is not known. However, the calculated molecular weight of 49.8 kDa for the polypeptide deduced from the sequence data was very close to the molecular size (52 kDa) for the 7Eb protein identified by the western blot analysis. This suggested that I have isolated a full-length or very nearly full-length, cDNA for the 7Eb protein. Further evidence supporting the assumption that the composite sequence of λ8.1 and λ17.1 represents nearly the full coding region comes from primer extension analysis (section 3.12). This analysis revealed that only about 7 nucleotides of mRNA were missing from the λ8.1 sequence. Therefore, based upon the evidence discussed above it is concluded that I have isolated cDNA clones representing nearly the full-length of the coding sequence for 7Eb protein.

4.3 CHARACTERISTICS OF THE 7Eb PROTEIN

Examination of the predicted amino acid sequence of 7Eb protein revealed some interesting features of the protein. The predicted protein is rich in basic amino acids

of a total of 422 amino acids, 93 residues (22%) are basic. The amino-terminal half of the protein has an equal distribution of charged residues, whereas the carboxyl-terminal half contains a high percentage of basic residues (53 basic and 15 acidic). Therefore, the carboxyl-terminal half of this protein has the potential to bind to DNA, RNA, or to acidic regions of other proteins. A computer search for sequence identity to the deduced amino acid sequence of 7Eb protein revealed no similarity in the PROSITE data bank (PC/GENE, a DNA and Protein sequence analysis program from Intelligenetics Inc.), which contains protein signature and common motifs found in various transcription factors. The primary sequence of 7Eb protein exhibited none of the amino acid motifs (helix-turn-helix, helix-loop-helix, leucine-zipper, or zinc-finger) that have been observed in many recently identified DNA binding proteins including many of the eukaryotic transcription factors (Johnson & McKnight, 1989).

The predicted amino acid sequence of 7Eb protein contains a nuclear targeting signal (Bairch, 1991) between the residues 390 and 394, which further indicates that the cDNAs I have isolated encode a nuclear protein.

The predicted 7Eb protein also showed multiple potential phosphorylation sites of various protein kinases as well as glycosylation consensus sequence. For example, protein kinase C phosphorylation sites at residues 11 and 59, and casein kinase II phosphorylation sites at residues 69, 90, and 129. Therefore, there is considerable potential for regulation of this protein's function through phosphorylation (see section 1.2.2), and it is tempting to speculate that these sites may serve as regulatory sites.

However, for these and all other potential phosphorylation sites, it still has to be determined experimentally whether they are actually phosphorylated *in vivo*.

4.3.1 Similarity of 7Eb protein to HMG box containing proteins

The predicted 7Eb protein sequence was compared with the amino acid sequences in the Swiss Protein database using the FASTA program (Pearson and Lipman, 1988), but this comparision did not show any identity to previously cloned nuclear matrix proteins including lamins. However, a significant similarity between the 7Eb amino acid sequence and a number of proteins, ribosomal promoter upstream-binding protein (hUBF; Jantzen *et al.*, 1990), yeast nuclear proteins (nonhistone binding proteins 6A and 6B; NHP6A and NHP6B, respectively; Kolodrubetz & Burgum, 1990), *Tetrahymena* macronucleus-associated basic protein (LG-1; Schulman *et al.*, 1987), and HMG-1 proteins from human, bovine, and rat. Alignment of all of these protein sequences with the predicted amino acid sequence of 7Eb revealed a common region of similarity among these proteins. The nuclear proteins hUBF, NHP6A, NHP6B, and LG-1, all contain a recently identified novel DNA binding domain known as an HMG box which is the region of amino acid sequence similarity found in all of these proteins. This region contains an amino acid sequence which is similar to nuclear protein HMG-1. The region of similarity which comprises about 80 amino acids, has been termed the HMG box (Jantzen *et al.*, 1990). At least 29 residues

within the 80-amino acid HMG box are identical or highly conserved among these proteins. It is this region which has a similarity to the predicted 7Eb protein sequence described here (figure 4.1). Overall, the predicted 7Eb protein contains 30 of the 42 shared residues and it demonstrates, the most similarity to yeast Nh6B, Nhp6A, and hUBF.

Lack of a clearly defined large consensus sequence among the HMG box domains in a variety of proteins (figure 4.1) may indicate either that such proteins recognize different DNA structures or that their different shapes are capable of recognizing similar DNA structures. The other members of the HMG box containing protein family include, the mitochondrial DNA-binding protein mtTF1 (Parisi & Clayton, 1991), the T-cell factor 1 (TCF-1; Oosterwegel *et al.*; Van de Wetering *et al.*, 1991), the lymphoid enhancer-binding factor 1 (LEF-1; Travis *et al.*, 1991), genetically defined mammalian testis-determining factor SRY (Gubbay *et al.*, 1990; Sinclair *et al.*, 1991), the fungal mating-type proteins Mat Mc and Mt al (Kelly *et al.*, 1988; Staben & Yanofsky, 1990), and the structure specific recognition protein (SSRP1; Bruhn *et al.*, 1992). Experiments aimed at delineating the DNA binding domains of hUBF, LEF-1 and TCF-1 indicated that a region containing about 200 amino acids which also includes the HMG box is necessary for DNA binding (Jantzen *et al.*, 1990; Oosterwegel *et al.*, 1991; Waterman *et al.*, 1991).

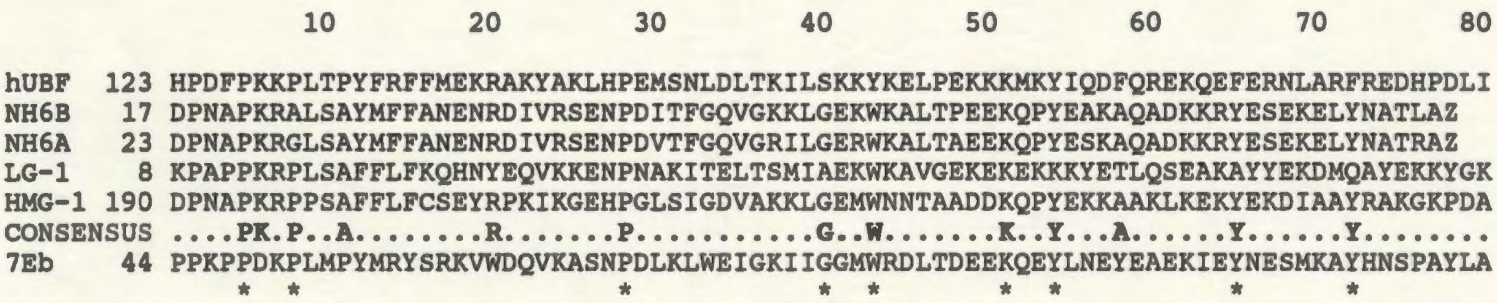
Except for HMG-1, all of the HMG box containing proteins which have been tested so far are sequence specific binding proteins. HMG-1 binds DNA in a

Fig. 4.1 Consensus sequences for HMG box domain.

Alignment of HMG box sequences.

The predicted protein (7Eb) sequence is aligned with other HMG box sequences. The numbers on the left refer to the amino acid positions in the proteins. The consensus sequence is from Clevers et al., 1993. * indicates 7Eb residues matching the consensus sequence.

The sequences are from: hUBF (Jantzen *et al.*, 1990), NH6A and NH6B (Kolodrubetz & Burgum, 1990), LG-1 (Schulman *et al.*, 1987), HMG-1 (Wen *et al.*, 1989)


10 20 30 40 50 60 70 80
hUBF 123 HPDFPKKPLTPYFRFFMEKRAKYAKLHPEMSNLDLTKILSKKYKELPEKKKMKYIQDFQREKQEFERNLARFREDHPDLI
NH6B 17 DPNAPKRALSAYMFFANENRDIVRSENPDITFGQVGKKLGEKWKALTPEEKQPYEAKAQADKKRYESEKELYNATLAZ
NH6A 23 DPNAPKRGLSAYMFFANENRDIVRSENPDVTFGQVGRILGERWKALTAEEKQPYESKAQADKKRYESEKELYNATRAZ
LG-1 8 KPAPPKRPLSAFFLFKQHNYEQVKKENPNAKITELTSMIAEKWKAVGEKEKEKKKYETLQSEAKAYYEKDMQAYEKKYGK
HMG-1 190 DPNAPKRPPSAFFLFCSEYRPKIKGEHPGLSIGDVAKKLGEMWNNTAADDKQPYEKKAAKLKEKYEKDIAAYRAKGKPDA
CONSENSUS ....PK.P..A.....R.....P.....G..W.....K..Y...A.....Y.....Y.....
7Eb 44 PPKPPDKPLMPYMRYSRKVWDQVKASNPDLKLWEIGKIIGGMWRDLTDEEKQEYLNEYEAEKIEYNESMKAYHNSPAYLA

nonspecific manner (Van Holde, 1989), whereas hUBF, mtTF-1, LEF-1, and TCF-1 bind specific nucleotide sequences (Jantzen *et al.*, 1990; Parisi and Clayton, 1991; Travis *et al.*, 1991; Van de Wetering *et al.*, 1991; Waterman *et al.*, 1991). Furthermore, *in vitro* binding studies have shown that HMG-1 binds preferentially to A+T-rich sequences (Wright & Dixon, 1988; Reeves & Nissen, 1990), single-stranded DNA (Isackson *et al.*, 1979), B-Z junctions (Hamada & Bustin, 1985) and cruciform structures (Bianchi *et al.*, 1989). These properties of HMG-1 are consistent with the suggestion that HMG-1 recognizes DNA structure rather than the sequence.

Western blot analysis (section 3.2.4.6) suggested that the HMG box containing 7Eb protein identified in this study is a component of the nuclear matrix fraction. As mentioned earlier, the nuclear matrix associated DNAs (MARs) are usually A+T-rich and some of the MARs have potential for extensive unpairing (or unwinding) when subjected to superhelical strain (Bode *et al.*, 1992). The identification of a nuclear matrix protein containing a putative DNA binding domain (HMG box), which may have the potential to bind an A+T-rich sequence and have higher affinity for single-stranded DNA over double-stranded (as is the case of HMG-1), would suggest a possible role for 7Eb protein in the organization of nuclear matrix associated DNA (refer to section 1.2.3).

The functional significance of the HMG box in this protein remains to be determined. An *in vitro* MAR protein binding study would be relevant to evaluate

the candidacy of the 7Eb protein as a DNA binding protein. To produce a sufficient amount of 7Eb protein that can be used for binding studies, it would be necessary to express this protein in bacteria. Further studies could then be directed towards the functional aspects of the protein. Once the DNA binding specificity of the 7Eb protein is established, further experiments could be performed to determine the specific region of the protein involved in the DNA-protein interaction. By creating deletions in the protein coding region of the cDNAs that I have isolated in this study, truncated forms of 7Eb protein can be expressed in bacteria. These truncated proteins can be used in a DNA protein binding assay to determine the region/s of the 7Eb protein involved. Furthermore, once the specificity of the DNA, which binds with this protein is established, further experiments can be designed to elucidate the mechanism of DNA protein interaction with this protein.

4.3.2. Similarity of 7Eb protein to cytoskeletal proteins

In addition to sequence similarity to HMG box containing proteins, the predicted 7Eb protein also revealed a significant sequence similarity to some non-HMG related proteins. These include, sheep trichohyalin, myosin heavy chain, dystrophin, bovine cardiac muscle troponin T isoform 1 and 2, and human, rat, and chicken neurofilament M. All of these are cytoskeletal proteins, and it is quite significant that the protein I have identified in this study shows amino acid sequence similarity to

some of the structural proteins of cells. The presence of amino acid sequences similar to those found in cytoskeletal protein strongly suggests that the protein identified in this study is a structural component of the nuclear matrix. However, further studies are needed to clearly define its role in the cell (nucleus).

An immunological approach can be adopted to learn more about 7Eb protein. It would be appropriate to express full-length 7Eb protein, and use that as an immunogen to raise a polyclonal antiserum against the whole protein. This preparation of antibody could be very useful to gain insight into the protein's *in vivo* function(s). In combination with cell fractionation, immunoblotting and immunofluorescent microscopy, the distribution of this protein during different phases of the cell cycle could be studied. Secondly, the effect of anti 7Eb antibody on cells as they progress through cell cycle can be determined by microinjection of the antibody into the nuclei (Kreis & Birchmeier, 1982). Microinjection can be a powerful way to analyse the function of an unknown protein such as 7Eb. Furthermore, with the availability of the cDNAs for 7Eb protein, it is possible to create a recombinant expression plasmid with an inducible promoter upstream of an antisense-oriented cDNA. The antisense cDNA can then be induced to produce anti-sense RNA in cells transfected with this plasmid. In this way the effect of depletion of endogenous mRNA corresponding to this protein by anti-sense RNA can be studied.

4.4 NORTHERN BLOT ANALYSIS

Northern blot analysis of HeLa cell mRNA has shown two bands with the cDNA probe. One major band with an apparent size of 2.5 kb was detected. This would suggest that the full-length message encoding the 7Eb protein contains approximately 300-400 bases of upstream and/or downstream untranslated sequence including a poly-A tail. The same sizes of mRNA were also observed when RNAs from various cell lines derived from various tumor tissues were probed with a cDNA fragment (see sections 3.14 and 4.6).

4.5 PRIMER EXTENSION ANALYSIS

To determine the length of the 5' end of the mRNA, a primer extension experiment was performed on HeLa cell total and poly(A)$^+$ RNA. For this analysis a primer (p31) corresponding to the extreme 5'end of the λ8.1 cDNA was used (section 3.12). The results of this analysis showed that about five to seven nucleotides are missing from the composite cDNA sequence presented in figure 3.17.

Hence the primer extension analysis indicated the length of the mRNA upstream of the ATG codon at +1 is about 29 or 31 nucleotide (section 3.12). Therefore, the primer extension analysis provided evidence that the length of the 5' untranslated region in the mRNA encoding 7Eb protein is only about 25 to 30 nucleotides. Thus

most of the 300-400 nucleotides missing from the composite cDNA presented in figure 3.15, are from the 3' untranslated region. To determine the sequence of the 5' end and 3' end of full-length mRNA, an alternate cloning strategy that utilizes a RACE (Rapid Amplification of cDNA Ends; Frohman *et al.*, 1988) PCR strategy can be employed.

4.6 EXPRESSION OF 7Eb PROTEIN IN DIFFERENT TUMOR CELL LINES

Northern blot analysis of RNA from several other cell lines also detected two bands. One major band of ~ 2.5 kb (in all the cell lines tested, except SW-620) and a minor band of ~ 1.7 kb (in all the cell lines tested, except SW-620). The minor band may represent an alternately spliced mRNA, a related message from another member of 7Eb gene family, or less likely a degradation product. To resolve these possibilities, a series of northern blotting experiments could be performed with HeLa cells mRNA. Using several restriction fragments or oligonucleotides complementary to various regions of the λ8.1 and λ17.1 cDNA clones as probes, one can determine how much these two mRNAs which were detected by northern blot analysis are related to each other. But in order to determine complete nucleotide sequence, it would be necessary to isolate a full-length cDNA corresponding to the 1.7 kb mRNA. For this, a small cDNA library can be made with size fractionated HeLa cell mRNA.

The resulting library could be screened with a DNA probe, selected on the basis of information obtained from northern blot analysis. Once the nucleotide sequence for a cDNA corresponding to 1.7 kb mRNA is established, further experiments can be directed towards the genomic mapping of these two mRNAs.

Detection of mRNA sizes similar to those obtained in HeLa cell suggested that the 7Eb protein is expressed in these cell lines. To investigate this further, a western blot analysis was performed. Total nuclear proteins prepared from different tumor cell lines were transferred to a blotting membrane, which was subsequently probed with RA-7E (section 3.2.4.5). A single immunoreactive band was detected in the nuclear protein fraction of all cell lines. Based on the size of the immunoreactive band which was calculated as ~ 52 kDa, it is probably the same protein (7Eb Protein) identified in HeLa cells. The detection of 7Eb protein in cell lines which represents distinct tissue types, therefore, suggest that this protein is probably ubiquitous in nature.

4.7 GENOMIC SOUTHERN BLOT ANALYSIS

To learn about the genomic organization of the 7Eb protein identified in this study, I have probed a genomic Southern with λ8.1 and λ17.1 cDNA inserts. Although a simple and almost identical band pattern was observed with both probes, multiple bands were detected and the combined sizes of these fragments amounts to

about 25-30 kbp. This suggests that the gene encoding the 7Eb protein has multiple introns. Another possiblity which is not necessarily in conflict with the first, is that the gene is a member of a small gene family. These possibilities can only be resolved by a detailed analysis of genomic clones encoding the 7Eb protein.

4.8 CONCLUSION

In conclusion, I have isolated and characterized cDNAs for a 52 kDa nuclear matrix protein (7Eb protein). The predicted amino acid sequence strongly suggests that the 7Eb protein I have identified, is a new member of the HMG box containing DNA binding proteins. Furthermore, on the basis of amino acid similarity to the HMG box domain as well as similarity to cytoskeletal proteins, a nuclear targeting signal in the predicted amino acid sequence, and a nuclear matrix localization of protein by cell fractionation studies, I propose, that the 7Eb protein I have identified is a nuclear matrix, DNA binding protein.

REFERENCES

Adachi, V., Kas, E., & Laemmli, U.K. (1989). Preferential cooperative binding of DNA topoisomerase II to scaffold associated regions. EMBO J. **8**, 3997-4006.

Aebi, U., Cohn, J., Buhle, L., & Gerace, L. (1986). The nuclear lamina is a meshwork of intermediate-type filaments. Nature **323**, 560-564.

Bairoch, A. (1991). Prosite: a dictionary of sites and patterns in proteins. Nucl. Acids Res. **19**, 2241-2245.

Belgrader, P., Siegel, A.J., & Berezney, R. (1991). A comprehensive study on the isolation and characterization of the HeLa S3 nuclear matrix. J. Cell Sci. **98**, 281-291.

Benyajati, C., & Worcell, A. (1976). Isolation, characterization, and structure of the folded interphase genome of *Drosophila melanogaster*. Cell **9**, 393-407.

Berezney, R. (1984). Organization and function of the nuclear matrix. In: Chromosomal Non histone Proteins (Hnilicia, L.S., ed.) vol.**4**, pp. 119-180, CRC Press, New York.

Berezney, R., & Coffey, D.S. (1974). Identification of a nuclear protein matrix. Biochem. Biophys. Res. Commun. **60**, 1410-1417.

Berezney, R., & Coffey, D.S. (1975). Nuclear protein matrix: association with newly synthesized DNA. Science **189**, 291-292.

Berezney, R., & Coffey, D.S. (1977). Nuclear matrix: isolation and characterization of a framework structure from rat liver nuclei. J. Cell Biol. **73**, 616-637.

Berrios, M., Osherhoff, N. & Fisher, P.A. (1985). *In situ* localization of DNA topoisomerase II, a major polypeptide component of the *Drosophila* nuclear matrix. Proc. Natl. Acad. Sci. USA. **82**, 4142-4146.

Bianchi, M.E., & Paonessa, G. (1989). Specific recognition of cruciform DNA by nuclear protein HMG-1. Science **243**, 1056-1059.

Birnboim, H.C. (1983). A rapid alkaline extraction method for the isolation of plasmid DNA. Methods Enzymol. **100**, 243-255.

Bode, J., Kohwi, Y., Dickinson, L., Joh, T., Klehr, D., Mielke, C., & Kohwi-Shigematsu, T. (1992). Biological significance of unwinding capability of nuclear matrix-associating DNAs. Science **255**, 195-197.

Bouvier, D., Hubert, J., Seve, A., & Bouteille, M. (1982). RNA is responsible for the three-dimensional organization of nuclear matrix protein in HeLa cells. Biol. Cell **43**, 143-146.

Brasch, K. (1982). Fine structure and localization of the nuclear matrix *in situ*. Exp. Cell Res. **140**, 1161-1171.

Bruhn, S., Pil, P.M., Essigmann, J.M., & Housman, D.E. (1992). Isolation and characterization of human cDNA clones encoding a high mobility group box protein that recognizes structural distortions to DNA caused by binding of the anticancer agent cisplatin. Proc. Natl. Acad. Sci. USA. **89**, 2307-2311.

Buik, R.N., Pullano, R., & Trent, J.M. (1985). Comparative properties of five human ovarian cell lines. Cancer Res. **45**, 3668-3676.

Burke, B., & Gerace, L. (1986). A cell free system to study reassembly of the nuclear envelope at the end of mitosis. Cell **44**, 639-652.

Cailleau, R. M, Brinkley, B.R., Beall, P.T., Wible, L.J. Mace, M.L., & Turner, D.S. (1980). Variations in cell form and cytoskeleton in human breast carcinoma cells *in vitro*. Cancer Res. **40**, 3118-3129.

Capco, D.G., & Penman, S. (1983). Mitotic architecture of the cell: the filament networks of the nucleus and cytoplasm. J. Cell Biol. **96**, 896-906.

Capco, D.G., Wan, K.M., & Penman, S. (1982). The nuclear matrix: Three-dimensional architecture and protein composition. Cell **29**, 847-858.

Chomczynski, P., & Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162**, 156-159.

Ciejeck, E.M., Nordstrom, J.L., Tsai, M.J., & O'Malley, B. (1982). Ribonucleic acid precursors are associated with the chick oviduct nuclear matrix. Biochemistry **21**, 4945-4953.

Ciejeck, E.M., Tsai, M.J., & O'Malley, B. (1983). Actively transcribed genes are associated with the nuclear matrix. Nature **306**, 607-609.

Clevers, H., Stehelin, D., & Laudet, V. (1993). Ancestry and diversty of the HMG box superfamily. Nucl. Acids Res. **21**, 2493-2501.

Cockerill, P.N., & Garrad, W.T. (1986a). Chromosomal loop anchorage of the kappa immunoglobulin gene occurs next to the enhancer in a region containing topoisomerase II sites. Cell **44**, 273-282.

Cockerill, P.N., & Garrad, W.T. (1986b). Chromosomal loop anchorage sites appear to be evolutionary conserved. FEBS Lett. **205**, 5-7.

Cockerill, P.N., Yuen, M., & Garrad, W.T. (1987). The enhancer of the immunoglobulin heavy chain locus is flanked by presumptive chromosomal loop anchorage elements. J. Biol. Chem. **262**, 5394-5397.

Cook, P.R., & Brazell, I.A. (1975). Supercoils in human DNA. J.Cell Sci. **19**, 261-279.

Cook, P.R., & Brazell, I.A. (1976). Confirmational constraints in nuclear DNA. J. Cell Sci. **22**, 287-302.

Dalton, S., Younghusband, H.B., & Wells, R.E. (1986). Chicken histone genes retains nuclear matrix association throughout the cell cycle. Nucl. Acids Res. **14**, 6507-6523.

Dente, L., Cesareni, G., & Cortese, R. (1983). EMBL: A new family of single-stranded plasmids. Nucl. Acids Res. **11**, 1645-1655.

Dickinson, L.A., Joh, T., Kohwi, Y., & Kowi-Shigematsu, K. (1992). A tissue-specific MAR/SAR DNA-binding protein with unusual binding site recognition. Cell **70**, 631-645.

Dijkwel, P.A., Wenink, P.V., & Poddighe, J. (1986). Permanent attachment of replication origins to the nuclear matrix in BHK cells. Nucl. Acids Res. **14**, 3241-3249.

Dworetzky, S.I., Fey, E.G., Penman, S., Lian, J.B., Stein, J.L., & Stein, G.S. (1990). Progressive changes in the protein composition of the nuclear matrix during rat osteoblast differentiation. Proc. Natl. Acad. Sci. USA. **87**, 4605-4609.

Earnshaw, W.C., Halligan, N., Cooke, C., Heck, M.M.S., & Liu, L.F. (1985). Topoisomerase II is a structural component of mitotic chromosome scaffolds. J. Cell Biol. **100**, 1706-1715.

Earnshaw, W.C., & Heck, M.M.S. (1985). Localization of topoisomerase II in mitotic chromosomes. J. Cell Biol. **100**, 1716-1725.

Feinberg, A.P., & Vogelstein, B. (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**, 6-13.

Feinberg, A.P., & Vogelstein, B. (1984). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **137**, 266-238.

Fey, E.G., Krochmalnic, G., & Penman, S. (1986). The nonchromatin substructures of the nucleus: the RNP-containing and RNP-depleted matrices analyzed by sequential fractionation and resinless electron microscopy. J. Cell Biol. **102**, 1554-1665.

Fey, E.G., & Penman, S. (1988). Nuclear matrix proteins reflect cell types origin in cultured human cells. Proc. Natl. Acad. Sci. USA. **85**, 121-125.

Fields, A.P., Kaufmann, S.H., & Shaper, J.H. (1986). Analysis of the internal nuclear matrix: oligomers of a 38 kD nucleolar polypeptide stability by disulphide bonds. Exp. Cell Res. **164**, 139-153.

Fisher, D.Z., Chaudhury, N., & Blobel, G. (1982). Isolation and characterization of a proteinaceous subnuclear fraction composed of nuclear matrix peripheral lamina, and nuclear pore complex from embryos of *Drosophila melanogaster*. J. Cell Biol. **92**, 674-686.

Franke, W.W. (1987). Nuclear lamins and cytoplasmic intermediate filament proteins: a growing multigene family. Cell **18**, 33-44.

Frohman, M.A., Dush, M.K., & Martin, G.R. (1988). Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA. **85**, 8998-9002.

Giard, D.J., Aaronson, S.T., Todaro, G.J., Arnstein, P., Dosik, H., & Parks, W.P. (1973). *In vitro* cultivation of human tumors: Establishment of cell lines dervied from a series of solid tumors. J. Natl. Cacer Inst. **51**, 1417-1423.

Gasser, S.M., & Laemmli, U.K. (1986a). The organization of chromatin loops; characterization of a scaffold attachment site. EMBO J. **5**, 511-517.

Gasser, S.M., & Laemmli, U.K. (1986b). Cohabitation of scaffold binding regions with upstream/enhancer elements of three developmentally regulated genes of *D. melanogaster*. Cell **46**, 521-530.

Gasser, S.M., & Laemmli, U.K. (1987). A glimpse at chromosomal order. Trends in Genet. **3**, 16-22.

Gasser, S.M., Laroche, T., Falques, J., Boy de La Tour, E., & Laemmli, U.K. (1986). Metaphase chromosome structure: Involvement of topoisomerase II. J. Mol. Biol. **188**, 613-629.

Georgatos, S.D., & Blobel, G. (1987). Lamin B constitutes an intermediate filament attachment site at the nuclear envelope. J. Cell Biol. **105**, 117-125.

Georgatos, S.D., Stournaras, C., & Blobel, G. (1988). Heterotypic and homotypic associations between the nuclear lamins: site-specificity and control by phosphorylation. Proc. Natl. Acad. Sci. USA. **85**, 4325-4329.

Germino, J., Gray, J.G., Charbonneau, H., Vanaman, T. & Bastia, D. (1983). Use of gene fusions and protein-protein interaction in the isolation of a biologically active regulatory protein: The replication initiator protein pf plasmid R6K. Proc. Natl. Acad. Sci. USA. **80**, 6848-6852.

Gerrace, L., & Blobel, G. (1980). The nuclear envelope lamina is reversibly depolymerized during mitosis. Cell **19**, 277-288.

Gerrace, L., Ottaviano, Y., & Konodor-Kach, C. (1982). Identification of a major polypeptide of the nuclear pore complex. J. Cell Biol. **95**, 826-837.

Goldman, M.A. (1988). The chromatin domain as a unit of gene regulation. BioEssays **9**, 50-55.

Goldman, A.E., Maul, G., Steinert, P.M., Yang, H., & Goldman, R.D. (1986). Keratin-like proteins that coisolate with intermediate filaments of BHK-21 cells are nuclear lamins. Proc. Natl. Acad. Sci. USA. **83**, 3839-3843.

Gubbay, J., Callignon, J., Koopman, P., Capel, B., Economou, A., Munsterberg, A., Vivian, N., Goodfellow, P., & Lovell-Badge, R. (1990). A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. Nature **346**, 245-250.

Hakes, D.L., & Berezney, R. (1991). DNA binding properties of the nuclear matrix and individual matrix proteins. Evidence for salt-resistant DNA binding sites. J. Biol. Chem. **266**, 11131-11140.

Hamada, H., & Bustin, M. (1985). Hierarchy of binding sites for chromosomal proteins HMG-1 and 2 in supercoiled deoxyribonucleic acid. Biochemistry **24**, 1428-1433.

Hanover, J.A., Cohen, C.K., Willingham, M.C., & Park, M.K. (1987). O-linked N-acetylglucosamine is attached to proteins of the nuclear pore. Evidence for cytoplasmic and nucleoplasmic glycoproteins. J. Biol. Chem. **262**, 9887-9894.

Hattori, M., & Sakaki, Y. (1986). Dideoxy sequencing method using denatured plasmid templates. Anal. Biochem. **152**, 232-238.

He, D., Nickerson, J.A. & Penman, S. (1990). Core filaments of the nuclear matrix. J.Cell Biol. **110**, 569-580.

Herman, R., Weymouth, L., & Penman, S. (1978). Heterogenous nuclear RNA-protein fibers in chromatin depleted nuclei. J. Cell Biol. **78**, 663-674.

Hofmann, J.F.X. & Gasser, S.M. (1991). Identification and purification of a protein that binds the yeast ARS consensus sequence. Cell **64**, 951-960.

Hofmann, J.F.X., Laroche, T., Brand, A.H., & Gasser, S.M. (1989). RAP-1 loop formation is necessary for DNA loop formation in vitro at the silent mating type locus HML. Cell **57**, 725-737.

Huynh, T., Young, R., & Davis, R. (1985). Construction and screening of cDNA libraries in lambda gt10 and lambda gt11. In: DNA cloning : A Practical Approach. (Glover, D., ed.) pp. 17-43, IRL press, Oxford.

Isackson, P.J, Fishback, J.L., Bidney, D.L., & Reek, G.R. (1979). Preferential affinity of high molecular weight high mobility group non-histone chromatin proteins for single-stranded. J. Biol. Chem. **254**, 5569-5572.

Jackson, D.A., & Cook, P.R. (1985a). A general method for preparing chromatin containing intact DNA. EMBO J. **4**, 913-918.

Jackson, D.A., & Cook, P.R. (1985b). Transcription occurs at a nucleoskeleton. EMBO J. **5**, 1403-1410.

Jackson, D.A., & Cook, P.R. (1988). Visualization of a filamentous nucleoskeleton with a 23 nm axial repeat. EMBO J. **7**, 3667-3677.

Jackson, D.A., Dikinson, P., & Cook, P.R. (1990). Attachment of DNA to the nucleoskeleton of HeLa cells examined using physiological conditions. Nucl. Acids Res. **18**, 4385-4393.

Jackson, D.A., McCready, S.J. & Cook, P.R. (1981). RNA is synthesized at the nuclear cage. Nature **292**, 552-555.

Jackson, D.A., McCready, S.J., & Cook, P.R. (1982). The ovalbumin gene is associated with the nuclear matrix of chicken oviduct cells. Cell **28**, 99-106.

Jantzen, H.M, Admon, A., Bell, P.B., & Tjian, R. (1990). Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG proteins. Nature **344**, 830-836.

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

Jost, J., & Seldran, M. (1984). Association of transcriptionally active vitelgenin II gene with the nuclear matrix of chicken liver. EMBO J. **3**, 2005-2008.

Kas, E., & Chasin, L.A. (1987). Anchorage of the chinese hamster dihydrofolate reductase gene to the nuclear scaffold occurs in a intragenic region. J. Mol. Biol. **198**, 677-692.

Kas, E., Izaurralde, E., & Laemmli, U.K. (1989). Specific inhibition of DNA binding to nuclear scaffolds and histone H1 by distamycin. The role of oligo(dA).oligo(dT) tracts. J. Mol. Biol. **210**, 587-599.

Kaufmann, S.H., Coffey, D.S. & Shaper, J.H. (1981). Consideration in the isolation of rat liver nuclear matrix, nuclear envelope and pore complex lamina. Exp. Cell Res. **132**, 104-123.

Kaufmann, S.H., Feilds, A.P., & Shaper, J.H. (1986). The nuclear matrix: Current concepts and unanswered questions. In Meth. Achiev. Exp. Path. Nuclear Electron Microscopy, vol. **12** (Jasmin, G. & Simard, R. ed.), pp. 141-171, Basel.

Kaufmann, S.H., Gibson, W., & Shaper, J.H. (1983). Characterization of the major polypeptides of the rat liver nuclear envelope. J. Biol. Chem. **258**, 2710-2719.

Kaufmann, S.H., & Shaper, J.H. (1984). A subset of non-histone nuclear proteins reversibly stabilized by the sulfhydryl cross-linking reagent tetrathinate. Exp. Cell Res. **155**, 477-495.

Kaufmann, S.H., & Shaper, J.H. (1991). Association of topoisomerase II with the hepatoma cell nuclear matrix: the role of intermolecular disulphide bond formation. Exp. Cell Res. **192**, 511-523.

Kelly, M., Burke, J., Smith, M., Klar, A., & Beach, D. (1988). Four mating -type genes control sexual differentiation in the fission yeast. EMBO J. **7**, 1537-1547.

Kohwi-Shigematsu, T., & Kohwi, Y. (1990). Torsional stress stabilizes immunoglobulin heavy chain enhancer. Biochemistry **29**, 9551-9560.

Kolodrubetz, D., & Burgum (1990). Duplicated NHP6 genes of *Saccharomyces cerevisae* encode proteins homologous to bovine high mobility group protein 1. J. Biol. Chem. **265**, 3234-3239.

Kreis, T.E., & Birchmeier, W. (1982). Microinjection of fluorescently labeled proteins into living cells with emphasis on cytoskeletal proteins. Intl. Rev. Cytol. **75**, 209-227.

Krohne, G., & Benavente, R. (1986). The nuclear lamins: a multigene family of proteins in evolution and differentiation. Exp. Cell Res. **162**, 1-10.

Krohne, G., Wolin, S.L., McKeon, F.D., Franke, W.W., & Kirschner, M.W. (1987). Nuclear lamin L_I of *Xenopus laevis*: cDNA cloning, amino acid sequence and binding specificity of a member of lamin B subfamily. EMBO J. **6**, 3801-3808.

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

Laemmli, U.K., Cheng, S.M., Adolph, K.W., Paulson, J.R., Brown, J.A., & Baumbach, W.R. (1978). Metaphase chromosome structure: the role of nonhistone proteins. Cold Spring Harbor Symp. Quant. Biol. **42**, 109-118.

Laliberte, J.F, Dagenais, A., Fillion, M., Bibor-Hardy, V., Simard, R., & Royal, A., (1984). Identification of distinct messenger RNAs for nuclear lamin C and a putative precursor of nuclear lamin A. J. Cell Biol. **98**, 980-985.

Lau, Y.F., & Arrighi, F.E. (1981). Studies of mammalian chromosome replication II. Evidence for the existence of define chromosome replicating units. Chromosoma **83**, 721-741.

Lawson, G.M., Knoll, B.J., March, C.J., Woo, S.L.C., Tsai, M.J., & O'Malley, B.W. (1982). Definition of 5' and 3' structural boundaries of the chromatin domain containing the ovalbumin multigene family. J. Biol. Chem. **257**, 1501-1507.

Leibovitz A., Stinson, J.C., McCombs, W.B., McCoy, C.E., Mazur, K. C., & Mabry, N., (1976). Classification of human colorectal adenocarcinoma cell lines. Cancer Res. **36**, 4562-4569.

Lebel, S., & Raymond, Y. (1987). Lamins A, B, and C share an epitope with the common domain of intermediate filament proteins. Exp. Cell Res. **169**, 560-565.

Lewis, C.D., & Laemmli, U.K. (1982). Higher order metaphase chromosome structure: evidence for metalloprotein interactions. Cell **29**, 171-181.

Maniatis, T., Fritsch, E.F., & Sambrook, J. (1982). Molecular Cloning : A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Martin, T.J., Moseley, J.M., Kubota, M., Jagger, D.H., Wettenhall, R.E., Kemp, B.E., Suva, L.J., Rodda, C.P., Ebelling, P.R., Hudson, P.J., & Zajac, J.D. (1987). Parathyroid hormone-related protein purified from a human lung cancer cell line. Proc. Natl. Acad. Sci. USA. **84**, 5048- .

Maundrell, K., Maxwell, E.S., Puvion, E., & Scherrer, K. (1981). The nuclear matrix of duck erythroblasts is associated with globin mRNA transcripts but not with the major proteins of 40S nuclear RNP. Exp. Cell Res. **136**, 435-445.

McKeon, F.D., Kirschner, M.W., & Caput, D. (1986). Homologies in both primary and secondary structure between nuclear envelope and intermediate filament proteins. Nature **319**, 463-468.

Miller, T.E., Huang, C., & Pogo, A.O. (1978). Rat liver nuclear skeleton and ribonucleoprotein complexes containing hnRNA. J.Cell Biol. **76**, 675-691.

Mirkovitch, J., Gasser, S.M., & Laemmli, U.K. (1988). Scaffold attachment of DNA loops in metaphase chromosomes. J.Mol. Biol. **200**, 100-109.

Mirkovitch, J., Mirault, M.-E., & Laemmli, U.K. (1984). Organization of the higher-order chromatin loop: Specific DNA attachment sites on nuclear scaffold. Cell **39**, 223-232.

Nakamura, H., Morita, T., & Sato, C. (1986). Structural organization of replicon domains during the DNA synthetic phase in the mammalian nucleus. Exp. Cell Res. **165**, 291-297.

Nakayasu, H., & Berezney, R. (1989). Mapping replicational sites in the eukaryotic cell nucleus. J.Cell Biol. **108**, 1-11.

Nakayasu, H., & Berezney, R. (1991). Nuclear matrins: Identification of the major nuclear matrix proteins. Proc. Natl. Acad. Sci. USA **88**, 10312-10316.

Nelkin, B.D., Pardoll, D.M., & Vogelstein, B. (1980). Localization of SV40 genes within supercoiled loop domains. Nucl. Acids Res. **8**, 5623-5633.

Newport, J.W., & Forbes, D.J. (1987). The nucleus structure, function, and dynamics. Ann. Rev. Biochem. 56, 535-565.

Newport, J., & Spann, T. (1987). Disassembly of the nucleus in mitotic extract: membrane vesicularization, lamin disassembly and chromosome condensation are independent process. Cell **48**, 219-230.

Nickerson, J.A., Krochmalnic, G., Wan, K.M., & Penman, S. (1989). Chromatin architecture and nuclear RNA. Proc. Natl. Acad. Sci. USA. **86**, 177-181.

Olmsted, J.B. (1981). Affinity purification of antibodies from diazotized paper blots of heterogeneous protein samples. J. Biol. Chem. **256**, 11955-11957.

Oosterwegel, M., Van de Wetering, M., Dooijes, D., Klomp, L., Winoto, A., Georgopoulos, K., Meijlink, F., & Clevers, H. (1991). Cloning of murine TCF-1, a T cell-specific transcription factor interacting with functional motifs in the CD3-Σ and TCR-α enhancers. J. Exp. Med. **173**, 1133-1142.

Osborn, M., & Weber, K. (1987). Cytoplasmic intermediate filament proteins and the nuclear lamins A, B, and C share the IFA epitope. Exp. Cell Res. **170**, 195-203.

Pardoll, D.M., & Vogelstein, B. (1980). Sequence analysis of nuclear matrix associated DNA from rat liver. Exp. Cell Res. **128**, 466-470.

Parisi, M.A., & Clayton, D.A. (1991). Similarity of human mitochondrial transcription factor 1 to high mobility group proteins. Science **252**, 965-969.

Pearson, W.R., & Lipman, D.L. (1988). Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA. **85**, 2444-2448.

Puk, T.T., & Fisher H.W. (1956). Demonstration of the existence of mutants with different growth requirements in a human cancer cell strain (HeLa). Exp. Med. **104**, 427-434.

Reichelt, R., Holzenburg, A., Buhle, E.T.J., Jarnik, M., Engel, A., & Aebi, U. (1990). Correlation between structural and mass distribution of the nuclear pore complex and of distinct pore complex components. J. Cell Biol. **110**, 883-894.

Robinson, S.I., Small, D., Rejeau, I., McKnight, S.G., & Vogelstein, B. (1983). The association of transcriptionally active genes with the nuclear matrix of the chicken oviduct. Nucl. Acids Res. **11**, 5113-5131.

Romig, H., Fackelmayer, F.O., Renz, A., Ramsperger, U., & Richter, A. (1992). Characterization of SAF-A, a novel nuclear DNA binding protein from HeLa cells with high affinity for nuclear matrix/scaffold attachment DNA elements. EMBO J. **11**, 3431-3440.

Roth, S.Y., Schulman, I.G., Cook, R.G., & Allis, C.D. (1987). The complete amino acid sequence of an HMG-like protein isolated from the macronucleus of *Tetrahymena*. Nucl. Acids Res. **15**, 8112.

Ruby, S.W., & Abelson, J. (1991). Pre-mRNA splicing in yeast. Trends Genet. **7**, 3710-3719.

Sambrook, J., Fritsch, E.F., & Maniatis, T. (1989). Molecular cloning : A L Laboratory Mananual. Cold Spring Harbor Laboratory; Cold Spring Harbor, New York.

Sanger, F., Nicklen, S., & Coulson, A.R. (1977). DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA. **74**, 5463-5467.

Sarnow, P., Mearing, P., Anderson, C.W., Reich, N., & Levine, A.J. (1982). Identification and characterization of an immunoloically conserved adenovirus early region 11,000 M_r protein and its association with the nuclear matrix. J. Mol. Biol. **162**, 565-583.

Schroder, H.C., Trolltsch, R., Wenger, M., Bachmann, B., Diehl-Seifest, & Muller, W.E.G., (1987). Cytochalasin B selectively releases ovalbumin mRNA precursors but not the mature ovalbumin mRNA from hen oviduct nuclear matrix. Eur. J. Biochem. **167**, 239-245.

Sinclair, A.H, Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischaux, A.-M., Lovell-Badge, R., & Goodfellow, P.N. (1990). A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motifs. Nature **346**, 240-244.

Small, D., Nelkin, B. & Vogelstein, B. (1985). The association of transcribed genes with the nuclear matrix of *Drosophila* cells during heat shock. Nucl. Acids Res. **13**, 2413-2431.

Small, D., & Vogelstein, B. (1985). The anatomy of supercoiled loops in the *Drosophila* 7F locus. Nucl. Acids Res. **13**, 7703-7713.

Smith, H.C., & Berezney, R. (1980). DNA polymerase is tightly bound to the nuclear matrix of actively replicating liver. Biochem. Biophys. Res. Commun. **104**, 548-556.

Smith, H.C., & Berezney, R. (1982). Nuclear matrix-bound deoxynucleic acid synthesis: an in vitro system. Biochemistry **21**, 6751-6761.

Smith, H.C., Ochs, R.L., Fernandez, E.A., & Spector, D.L. (1986). Macromelecular domains containing nuclear protein p107 and U-snRNP protein p28: further evidence for an *in situ* nuclear matrix. Mol. Cell. Biochem. **70**, 151-168.

Smith, H.C., Puvion, E., Buchholtz, A., & Berezney, R. (1984). Spatial distribution of DNA loop attachment and replicational sites in the nuclear matrix. J. Cell Biol. **99**, 1794-1802.

Smith, H.C., Spector, D.L., Woodcock, C.L.F., Ochs, R.L., & Bhorjee, J. (1985). Alteration in chromatin confirmation are accompanied by reorganization of non chromatin domains that contain U-snRNP protein p28 and nuclear protein p107. J. Cell Biol. **101**, 560-567.

Snow, C.M., Senior, A., & Gerrace, L. (1987). Monoclonal antibodies identify a group of nuclear pore complex glycoproteins. J. Cell Biol. **104**, 1143-1156.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**, 503-517.

Sperry, A.O., Blasquez, V.C., & Garrad, W.T. (1989). Dysfuction of chromosomal loop attachment sites: Illegitimate recombination linked to matrix association regions and topoisomerase II. Proc. Natl. Acad. Sci. USA. **86**, 5497-5501.

Staben, C., & Yanofsky, C. (1990). *Neurpspora crassa* mating type-region. Proc. Natl. Acad. Sci. USA. **87**, 4917-4921.

Stalder, J., Larsen, A., Engel, J.D., Polan, M., Groudine, M., & Weintraub, H. (1980). Tissue-specific DNA cleavage in the globin chromatin domain introduced by DNase I. Cell **20**, 451-460.

Staufenbiel, M., & Depper, W. (1983). Different structural systems of the nucleus are targets for SV40 Large T antigen. Cell **33**, 173-181.

Staufenbiel, M. & Depper, W. (1984). Preparation of nuclear matrices from cultured cells: subfractionation of nuclei *in situ*. J. Cell Biol. **98**, 1886-1894.

Stief, A., Winter, D.M., Stratling, W.H., & Sipperl, A.E. (1989). A nuclear DNA attachment element mediates elevated and position-independent gene activity. Nature **341**, 343-345.

Stuurman, N., Meijne, M.L., Van der pol, A.J., Se Jong, L., Van Driel, R., & Van Renswoude, J. (1990). The nuclear matrix from cells of different origin: Evidence for a common set of matrix proteins. J. Biol. Chem. **265**, 5460-5465.

Sykes, R. C., Lin, D., Hwang, S.J., Framson, P.E., & Chinault, A.C. (1988). Yeast ARS function and nuclear matrix association coincide in a short sequence from the human HPRT locus. Mol. Gen. Genet. **212**, 301-309.

Travis, A., Amsterdam, A., Belanger, C., & Grosschedl, R. (1991). LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor alpha enhancer function. Genes Dev. **5**, 880-894.

Tubo, R.A., & Berezney, R. (1987a). Pre-replicative association of multiple replicative enzyme activities with the nuclear matrix during rat liver regeneration. J. Biol. Chem. **262**, 1148-1154.

Tubo, R.A., & Berezney, R. (1987b). Identification of 100 and 150 S DNA polymerase α-primase megacomplexes solubilized from the nuclear matrix of regenerating rat liver. J. Biol. Chem. **262**, 5857-5865.

Tubo, R.A., Smith, H.C., & Berezney, R. (1985). The nuclear matrix continues DNA synthesis at *in vivo* replication forks. Biochim. Biophys. Acta **825**, 326-334.

Van der Velden, H.M.W., Van Willigen, G., Wetzels, R.H.W., & Wanka, F. (1984). Attachment of origins of replication to the nuclear matrix and chromosomal scaffold. FEBS Lett. **171**, 13-16.

Van de Wetering, M., Oosterwegel, M., Dooijes, D., & Clevers, H. (1991). Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box. EMBO J. **10**, 123-132.

Van Eekelen, C.A.G., Salden, M.H.L., Habets, W.J.A., Van de Putte, L.B.A., & Van Venrooij, W.J. (1982). On the existence of an internal nuclear protein structure in HeLa cells. Exp. Cell Res. **141**, 181-190.

Van Eekelen, C.A.G., & Van Venrooij. (1981). HnRNA and its attachment to a nuclear protein matrix. J. Cell Biol. **88**, 554-563.

Van Holde, K.E. (1989). Chromatin (Rich, A. ed.), Springer-Verlag Press, New York.

Verheijen, R., Kuijpers, H., Vooijs, P., Van Venrooij, W., & Ramaekers, F. (1986). Distribution of the 70K U1 RNA associated protein during interphase and mitosis. J.Cell Sci. **80**, 103-122.

Vogelstein, B., Pardoll, D.M., & Coffey, D.S. (1980). Supercoiled loops and eukaryotic DNA replication. Cell **22**, 79-85.

Von Kries, J.P., Buhrmester, H., & Stratling, W.H. (1991). A matrix scaffold attachment region binding protein: Identification, purification, and mode of binding. Cell **64**, 123-135.

Waterman, M.L., Fischer, W.H., & Jones, K.A. (1991). A thymus-specific member of the HMG protein family regulates the human T cell receptor alpha enhance. Genes Dev. **5**, 656-669.

Weisbrod, S. (1982). Active chromatin. Nature **297**, 289-295.

Wen, L., Huang, J.K., Johnson, B.H., & Reeck, G.R. (1989). A human placental cDNA clone that encodes nonhistone chromosomal protein HMG-1. Nucl. Acids Res. **17**, 1197-1214.

Wright, J.M. & Dixon, G.H. (1988). Induction by torsional stress of an altered DNA conformation 5' upstream of the gene for a high mobility group protein from trout and specific binding to flanking sequences by the gene product HMG-T. Biochemistry **27**, 576-581.

Younghusband, H.B. and Maundrell, K. (1982) Adenovirus DNA is associated with the nuclear matrix of infected cells. J. Virol. **43**, 705-713.

Zeitlin, S., Paren, A., Silverstein, S., & Efstratiadis, A. (1987). Pre-mRNA splicing and the nuclear matrix. Mol. Cell. Biol. **7**, 111-120.

Zeitlin, S., Wilson, R.C., & Efstratiadis, A. (1989). Autonomous splicing and complementation of *in vivo*-associated spliceosomes. J. Cell Biol. **108**, 765-777.

APPENDIX 1

Nucleotide sequence of clone λ3A

```
GGAAAGGAGAAGGAGGCCGCAGAGCAAGCTGAGCGCAGTCAGAGCAGCATCGTTCCTGAG   60
GAAGAACAAGCAGCTAACAAAGGCGAGGAGAAGAAAGACGACGAGAACATTCCGATGGAG  120
ACAGAGGAGACACACCTTGAAGAAACAACAGAGAGCCAACAGAATGGTGAAGAAGGCACG  180
TCTACTCCTGAGGACAAGGAGAGTGGGCAGGAGGGGGTCGACAGTATGGCAGAGGAAGGA  240
ACCAGTGATAGTAACACTGGCTCGGAGAGCAACAGTGCAACAGTGGAGGAGCCACCAACA  300
GATCCCATACCAGAAGATGAGAAAAAAGAATAAGTGTTGCCTTGTTTTGTGTGTTCTAAA  360
TACTTTTTTTAATGAAAAAATGTTTTTTGGTTTTAATGGTGTTACGTGGTTTGTGTATTA  420
ATTTTTTTTCTTGTCCATATCATACCACCAAAGGCTTTTGGACCATTTAGCATCATGAGC  480
CTAATGGCTCAGTCAGTCACCTTTCTT                                   527
```

APPENDIX 2

Nucleotide sequence of clone λ6D

```
AAAGGCAGGAGGAAAGGAGAAGGAGGCCGCAGAGCAAGCTGAGCGCAGTCAGAGCAGCAT   60
CGTTCCTGAGGAAGAACAAGCAGCTAACAAAGGCGAGGAGAAGAAAGACGACGAGAACAT  120
TCCGATGGAGACAGAGGAGACACACCTTGAAGAAACAACAGAGAGCCAACAGAATGGTGA  180
AGAAGGCACGTCTACTCCTGAGGACAAGGAGAGTGGGCAGGAGGGGGTCGACAGTATGGC  240
AGAGGAAGGAACCAGTGATAGTAACACTGGCTCGGAGAGCAACAGTGCAACAGTGGAGGA  300
GCCACCAACAGATCCCATACCAGAAGATGAGAAAAAAGAATAAGTGTTGCCTTGTTTTGT  360
GTGTTCTAAATACTTTTTTTAATGAAAAAATGTTTTTTGGTTTTAATGGTGTTACGTGGT  420
TTGTGTATTAATTTTTTTTCTTGTCCATATCATACCACCAAAGGCTTTTGGACCATTTAG  480
CATCATGAGCCTAATGGCTCAGTCAGTCACCTTTCTTAAGTGTTGTGAAGATGGCTCTTT  540
TCTTTGGATCTTGTTTCTAGCCCTCAACTGCTGAAAGCCTCAGAATTTAGATTAATTGAG  600
AAAACACCCACCTCTTTTAGAGAATTATCCTTTGATGCTGCAGAATCTACTCTTACAATG  660
CCTTCCTACAGCTCACTGGGGT                                         682
```

APPENDIX 3

Nucleotide sequence of clone λ1B

```
GAAAAAGTCATGGAGGCCATGGGGTTGGCTTGAAACCAGCTTTGGGGGGTTCGATTCCTT   60
CCTTTTTGTCTAGATAGAGTACAATGAATCTATGAAGGCCTATCATAATTCCCCCGCGTA  120
CCTTGCTTACATAAATGCAAAAAGTCGTGCAGAAGCTGCTTTAGAGGAAGAAAGTCGACA  180
GAGACAATCTCGCATGGAGAAAGGAGAACCGTACATGAGCATTCAGCCTGCTGAAGATCC  240
AGATGATTATGATGATGGCTTTTCAATGAAGCATACAGCCACCGCCCGTTTCCAGAGAAA  300
CCACCGCCTCATCAGTGAAATTCTTAGTGAGAGTGTGGTGCCAGACGTTCGGTCAGTTGT  360
CACAACAGCTAGAATGCAGGTCCTCAAACGGCAGGTCCAGTCCTTAATGGTTCATCAGCG  420
AAAACTAGAAGCTGAACTTCTTCAAATAGAGGAACGACACCAGGAGAAGAAGAGGAAATT  480
CCTGGAAAGCACAGATTCATTTAACAATGAACTTAAAAGGTTGTGCGGTCTGAAAGTAGA  540
AGTGGATATGGAGAAAATTGCAGCTGAGATTGCACAGGCAGAGGAACAGGCCCGCAAAAG  600
GCAGGAGGAAAGGAGAAGGAGGCCGCAGAGCAAGCTGAGCGCAGTCAGAGCAGCATCGTT  660
CCTGAGGAAGAACAAGCAGCTAACAAAGGCGAGGAGAAGAAAGACGACGAGAACATTCCG  720
ATGGAGACAGAGGAGACACACCTTGAAGAAACAACAGAGAGCCAACAGAATGGTGAAGAA  780
GGCACGTCTACTCCTGAGGACAAGGAGAGTGGGCAGGAGGGGGTCGACAGTATGGCAGAG  840
GAAGGAACCAGTGATAGTAACACTGGCTCGGAGAGCAACAGTGCAACAGTGGAGGAGCCA  900
CCAACAGATCGCGGCTACTAACCAAACTGCCAGATACAAAAGGGTCA               947
```

APPENDIX 4

Nucleotide sequence of clone λ5h

```
CTCCGGCGGGGTCGAAGAAGGTGGTGTTGAGGTTGCGGTCTGTTAGTAGTATTGTGATGC   60
CAGCAGCTAGGACTGGGAGAGATAGGAGAAGTAGGACTGCTGTGATTAGGACGGATCAGA  120
CGAAGAGGGGCGTTTGGTATTGGGTTATGGCAGGGGGTTTTATATTGATAATTGTTGTGA  180
TGAAATTGATGGCCCCTAAGATAGAGGAGACACCTGCTAGGTGTAAGGAGAAGATGGTTA  240
GGTCTACGGAGGCTCCAGGGTGGGAGTAGTTCCCTGCTAAGGAGGGTAGACTGTTCAACC  300
TGTTCCTGCTCCGGCCTCCACTATAGCAGATGCGAGCAGGAGTAGGAGAGAGGGAGGTAA  360
GAGTCAGAAGCTTATCTTGTTTATGCGGGGAAACGCCATATCGGGGGCACCGATTACCCA  420
GGGGAACTAGTGAGTTGCCAAAGCCTCCGATTATGATGGGTATTACTATGAAGAAGATTA  480
TTACAAATGCATGGGCTGTGACGATAACGTTGTAGATGTGGTCGTTACCTAGAAGGTTGC  540
CTGGCTGGGGGAGCTCGGCTCGNAATAAGGAGGCTTAGAGCTGTGCCTAGGACTCCAGCT  600
CATGCGCCGAATAATAGGTATAGTGTTCCAATGTCTTTGTGGTTTGTAGAGAATAGTCAA  660
CGGTCGGGCTTACGATCATCCAGCACCAGACGGCCTTTCACCTCGCCAGTTTTGGCATCT  720
ATCGCCG                                                        727
```

Nucleotide sequence of clone λ5h.1

```
GGAGGCCATGGGGTTGGCTTGAAACCAGCTTTGGGGGGTTCGATTCCTTCCTTTTTTGTC   60
TAGATTTTATGTATACGGGTTCTTCGAATGTGTGGTAGGGTGGGGGCATCCATATAGTCA  120
CTCCAGGTTTATGGAGGGTCAACTACTATTAGGACTTTTCCCTTCGAAAGCGAAGGCTTG  180
ACGTTTCAGACAA                                                  193
```

Nucleotide sequence of clone λ5h.2

```
GACAATCGAGTAGTACTCCGATTGAAGCCCCATTCGTATAATAATTACATCACAAGACGT   60
CTTGCACTCATGAGCTGTCCCCACATTAGGCTTAAAAACAGATGCAATTCCCGGACGTCT  120
AAACCAAACCACTTTCACCGCTACACGACCGGGGGTATACTACGGTCAATGCTCTGAAAT  180
CTGTGGAGCAAACCACAGTTTCATGCCCATCGTCCTAGAATTAATTCCCCTAAAAATCTT  240
TGAAATAGGACCCGTATTTACCCTATAGCACCCCCTCTACCCCCTCTAGAGCGGAATTCG  300
ATATCAAGCTTATCGAT                                              317
```

APPENDIX 5

Nucleotide sequence of clone λ5A

```
CTCTATCGTCGATGCGGAGGCCATGGGGTTGGCTTGAAACCAGCTTTGGGGGGTTCGATT   60
CCTTCCTTTTTGTCTAGATTTTATGTATACGGGTTCTTCGAATGTGTGGTAG          112
```

Appendix 6

Nucleotide sequence of clone λ5B

```
CCACTGATTTCCCCTATCTCAGGCTACACCCTAGACCAAACCTACGCCAAAATCCATTTC   60
ACTATCATATCATCGGCGTAAATCTAACTTTCTTGGAGGCCATGGGGTTGGCTTGAAACC  120
AGCTTTGGGGGGTTCGATTCCTTCCTTTTTGTCTAGAT                        158
```

Appendix 7

Nucleotide sequence of clone λ5C

```
ATGCGGAGGCCATGGGGTTGGCTTGAAACCAGCTTTGGGGGGTTCGATTCCTTCCTTTTT   60
TGTCTAGATTTTATGTATACGGGTTCTTCGAATGTGTGGTAGGGTGGGGGCATCCATATA  120
GTCACTCCAGGTTTATGG                                             138
```

Appendix 8

Nucleotide sequence of clone λ5D

```
CCACTGATTTCCCCTATCTCAGGCTACACCCTAGACCAAACCTACGCCAAAATCCATTTC   60
ACTATCATATCATCGGCGTAAATCTAACTTTCTTGGAGGCCATGGGGTTGGCTTGAAACC  120
AGCTTTGGGGGGTTCGATTCCTTCCTTTTTGTCTAGAT                        138
```

Appendix 9

Nucleotide sequence of clone λ5E

```
CTCTATCGTCGATGCGGAGGCCATGGGGTTGGCTTGAAACCAGCTTTGGGGGGTTCGATT   60
CCTTCCTTTTTGTCTAGATTTTATGTATACGGGTTCTTCGAATGTGTGGTAGG         113
```

Appendix 10

Nucleotide sequence of clone λ5F

```
CCATGGGGTTGGCTTGAAACCAGCTTTGGGGGGTTCGATTCCTTCCTTTTTTGTCTAGAT   60
TTTATGTATACGGGTTCTTCGAATGTGTGGTAGGGTGGGGGCATCCATATAGTCACTCCA  120
GGTT                                                           124
```

Appendix 11

Nucleotide sequence of clone λ5G

```
CGTCGATGCGGAGGCCATGGGGTTGGCTTGAAACCAGCTTTGGGGGGTTCGATTCCTTCC   60
TTTTTGTCTAGATTTTATGTATACGGGTTCTTCGAATGTGTGGTAG                104
```

APPENDIX 12

Nucleotide sequence of clone λ8.1

```
GGAATTCCCAGCTCCTGCAACACAAATGCCCAGCACACCAGGGTTTGTGGGATACAATCC   60
ATACAGTCATCTCGCCTACAACAACTACAGGCTGGGAGGGAACCCGGGCACCAACAGCCG  120
GGTCACCGGATCCTCTGGTATCACGATTCCAAAACCCCCAAAGCCACCAGATAAGCCGCT  180
GATGCCCTACATGAGGTACAGCAGAAAGGTCTGGGACCAAGTAAAGGCTTCCAACCCTGA  240
CCTAAAGTTGTGGGAGATTGGCAAGATTATTGGTGGCATGTGGCGAGATCTCACTGATGA  300
AGAAAAACAAGAATATTTAAACGAATACGAAGCAGAAAAGATAGAGTACAATGAATCTAT  360
GAAGGCCTATCATAATTCCCCCGCGTACCTTGCTTACATAAATGCAAAAAGTCGTGCAGA  420
AGCTGCTTTAGAGGAAGAAAGTCGACAGAGACAATCTCGCATGGAGAAAGGAGAACCGTA  480
CATGAGCATTCAGCCTGCTGAAGATCCAGATGATTATGATGATGGCTTTTCAATGAAGCA  540
TACAGCCACCGCCCGTTTCCAGAGAAACCACCGCCTCATCAGTGAAATTCTTAGTGAGAG  600
TGTGGTGCCAGACGTTCGGTCAGTTGTCACAACAGCTAGAATGCAGGTCCTCAAACGGCA  660
GGTCCAGTCCTTAATGGTTCATCAGCGAAAACTAGAAGCTGAACTTCTTCAAATAGAGGA  720
ACGACACCAGGAGAAGAAGAGGAAATTCCTGGAAAGCACAGATTCATTTAACAATGAACT  780
TAAAAGGTTGTGCGGTCTGAAAGTAGAAGTGGATATGGAGAAAATTGCAGCTGAGATTGC  840
ACAGGCAGAGGAACAGGCCCGCAAAAGGCAGGAGGAAAGGAGAAGGAGGCCGCAGAGCAA  900
GCTGAGCGCAGTCAGAGCAGCATCGTTCCTGAGGAAGAACAAGCAGCTAACAAAGGCGAG  960
GAGAAGAAAGACGACGAGAACATTCCGATGGAGACAGAGGAGACACACCTTGAAGAAACA 1020
ACAGAGAGCCAACAGAATGGTGAAGAAGGCACGTCTACTCCTGAGGACAAGGAGAGTGGG 1080
CAGGAGGGGGTCGACAGTATGGCAGAGGAAGGAACCAGTGATAGTAACACTGGCTCGGAG 1140
AGCAACAGTGCAACAGTGGAGGAGCCACCAACAGATCCCATACCAGAAGATGAGAAAAAA 1200
GAATAAGTGTTGCCTTGTTTTGTGTGTTCTAAATACTTTTTTTAATGAAAAAATGTTTTT 1260
TGG                                                            1263
```

APPENDIX 13

Nucleotide sequence of clone λ17.1

```
TCCGATGGAGACAGAGGAGACACACCTTGAAGAAACAACAGAGAGCCAACAGAATGGTGA   60
AGAAGGCACGTCTACTCCTGAGGACAAGGAGAGTGGGCAGGAGGGGGTCGACAGTATGGC  120
AGAGGAAGGAACCAGTGATAGTAACACTGGCTCGGAGAGCAACAGTGCAACAGTGGAGGA  180
GCCACCAACAGATCCCATACCAGAAGATGAGAAAAAAGAATAAGTGTTGCCTTGTTTTGT  240
GTGTTCTAAATACTTTTTTTAATGAAAAAATGTTTTTTGGTTTTAATGGTGTTACGTGGT  300
TTGTGTATTAATTTTTTTTCTTGTCCATATCATACCACCAAAGGCTTTTGGACCATTTAG  360
CATCATGAGCCTAATGGCTCAGTCAGTCACCTTTCTTAAGTGTTGTGAAGATGGCTCTTT  420
TCTTTGGATCTTGTTTCTAGCCCTCAACTGCTGAAAGCCTCAGAATTTAGATTAATTGAG  480
AAAACACCCACCTCTTTTAGAGAATTATCCTTTGATGCTGCAGAATCTACTCTTACAATG  540
CCTTCCTACAGCTCACTGGGGTGCCCAAAGCCATAGCTTTAAACCTTCCCAGTCCCCATC  600
AGAGCTTCCTGAAAGTCTCCTCTCTTGTTTACTTCTGCAAAGGGTAGCTTCTTAAAAACG  660
TGATCATGTATGAGTATGTATTTGTTCACTTACCCTTTTTTTACTTTTAATCAATGTCAG  720
ATACCAAGAGTTGTGTTAAGCTGAGTGTAGTATGTAACTAACTACACTTGGATCTTACTG  780
ATCCAGAAATAGTCCCCATAGTTAGAGTAGTTACTTATGAAGTGGTTATTAAAGTGAACA  840
ACAGCACATATACATTATCTATACTGCTTTTTGTTATGATTAATACGGGTGTTCTTAATA  900
TTCTTGTAATGGACTAGATAATTCAAACTGATTAGCCCATTCCAGAAGAAAAACAGCTGG  960
GAATTAAGTTAATCCACTTGAAATTGTTTTACAATAATCAGACATCCAAACCTCAAGGCT 1020
CAGGATCCCATAGACCAGAGTTTTCCTTTTTGATAAACTTAGTAAAGTCTTGGAGACTAG 1080
AAGCAAGATAGTTTGTGACACATAAGCTTCCTCGAAAAACTAGAATAGATTTTTACTGAT 1140
AGTGGTATATCTGATGGTATATGTTTCTTAAGGTCCGGAATTCCGGA               1187
```

FILE ON TOP

NO PUBLICATION NUMBER

XEROX

DOCTORAL DISSERTATIONS
PUBLISH ABSTRACT ONLY

Issue
APPENDIX 6

INSTRUCTIONS

1. Please read the Agreement on the reverse side of this form carefully. It authorizes the publication by Xerox University Microfilms of a 600-word abstract in *Dissertation Abstracts International*. Indicate your approval of the Agreement by signing item 11 below.

2. After reading the section on Reprints on the reverse side of this form, fill in the remainder of this side of the form. Use typewriter or ink. Print plainly. Be accurate.

DO NOT WRITE IN THIS SPACE

Abstract Length .....

Supervisor .....

Personal Data

1. Full Legal Name HAMEED (last name)
KHWAJA (first name) SHAHID (middle name)

2. Year of Birth 53 3. Country of Citizenship PAKISTAN

4. Present mailing address C/O FACULTY OF MEDICINE
MEMORIAL UNIVERSITY OF NEWFOUNDLAND

Future mailing address AS ABOVE

Effective date

Doctoral Degree Data

5. Abbreviation for Degree Authorized by University Ph.D

6. Degree date authorized by University

7. Full Name of University Conferring Degree
MEMORIAL UNIVERSITY OF NEWFOUNDLAND

Title Data

8. Title of Dissertation *exactly* as it appears on Title Page and Abstract. *(Please attach a copy of your title page.)*
MOLECULAR CLONING OF cDNA FOR A NUCLEAR MATRIX PROTEIN

9. Enter Subject Category selected for your dissertation
(Choose from list on overleaf)
Main Field MOLECULAR CELL BIOLOGY
Sub Field

Reprints

10. Please furnish _____ reprints of my abstract. Enclosed is my check or money order in the amount of $_____ (U.S.) for this service. A single tear slip copy is available at no charge upon request.

11. I agree to the provisions of the *Agreement* on the other side of this form.

Signature

Jan 20/93
Date

IMPORTANT! Please select subject matter on overleaf.

Xerox University Microfilms