ISAS - INTERNATIONAL SCHOOL FOR ADVANCED STUDIES

HUMAN PROTEINS THAT INTERACT SPECIFICALLY WITH A DNA REPLICATION ORIGIN

Thesis Submitted for the Degree of

Doctor Philosophiae

Candidate:
Lidija Marušić

Supervisor:
Prof. Arturo Falaschi

Academic Year 1991/92

TRIESTE

SISSA - SCUOLA INTERNAZIONALE SUPERIORE DI STUDI AVANZATI
TRIESTE
Strada Costiera 11

HUMAN PROTEINS THAT INTERACT SPECIFICALLY WITH A DNA REPLICATION ORIGIN

Thesis Submitted for the Degree of

Doctor Philosophiae

Candidate:

Lidija Marusic

Supervisor:

Prof. Arturo Falaschi

Academic Year 1991/92

ACKNOWLEDGMENTS

This work has been done at the International Centre for Genetic Engineering and Biotechnology. I am very grateful to Prof. Arturo Falaschi for his guidance during the preparation of the thesis. Special thanks are also due Dr. Eva Csordas-Toth and Alexander Ochem for their help and advices in performing the experiments and Dr. Georgine Faulkner for useful suggestions.

CONTENTS

1. INTRODUCTION

1.1 PROTEIN-DNA INTERACTION: STRUCTURAL MOTIFS USED IN DNA RECOGNITION

Protein recognition of specific DNA-binding sites is critical for the biological function of the cell. Fundamental processes such as transcription, replication and restriction are under the control of DNA binding proteins. In order to function as specific regulators, regulatory proteins must be able to recognize and to bind preferentially to certain DNA sequences against the background of a large number of competing sites of similar sequences that are present in the genome. The selection of the sequence is based on the physicochemical interaction between the protein and DNA. In a protein binding site the base pairs are arranged in specified order and are able to interact with the amino acid residues forming hydrogen bonds. Interactions other than hydrogen bonds can also contribute to binding such as hydrophobic interactions (especially with the $-CH_3$ group of thymine), conformation and stability of the DNA helix and ionic interactions between basic amino acids residues and

negatively charged phosphates of the double helix, however the hydrogen bond matrix remains the most important interaction in a specific sequence recognition by the protein. The regions responsible for DNA-binding activities are small domains of less than 100 amino acids. Four different structural patterns have been characterized: helix-turn-helix and two types of zinc fingers for DNA binding, leucine zipper and helix-loop-helix (both adjacent to a basic domain) for protein dimerization (Ohlendorf & Matthews, 1983; Pabo & Sauer,1984; Berg & von Hippel,1988; Struhl, 1989; Brennan & Matthews, 1989a; Latchman, 1990; Freemont et al., 1991a).

1.1A HELIX-TURN-HELIX

The helix-turn-helix was the first structural motif to be characterized for a DNA-binding domain. It has been found in several prokaryotic activator and repressor proteins (Brennan & Matthews, 1989b; Harrison & Aggarwal, 1990; Freemont et al., 1991a). It consists of two α helices separated by a β turn, with a second helix protruding from the surface of the protein that matches in shape the major groove of the DNA (**Figure 1.1**).

The Cro protein from bacteriophage λ is a prototypical example of these class of proteins. It has a simple structure consisting of three α helices and three antiparallel β sheets and binds as a dimer to a two fold symmetrical, 17 bp operator site (consensus sequence 5'- T(1)-A(2)-T(3)-C(4)-

A(5)-C(6)-C(7)-G(8)-C(9)-G(8')-G(7')-G(6')-T(5')-G(4')-A(3')-T(2')-A(1') -3' where the numbers in parentheses indicate the base-pair position). The structure of a Cro-operator complex has been recently solved by X-ray crystallography (Brennan et al., 1990). The α3 ("recognition") helix of the protein fits into the major groove of the DNA and recognizes a sequence specific site. It appears that Lys-32 contacts base-pairs G·C(4), T·A(5) and G·C(6), while Ser-28 contacts base-pairs A·T(3) and G·C(4). The side chains of the α-2 helix are important in the orientation and binding of the helix-turn-helix motif. Also, Try-26 in the α-2 helix contacts base-pair 1 in the operator site.

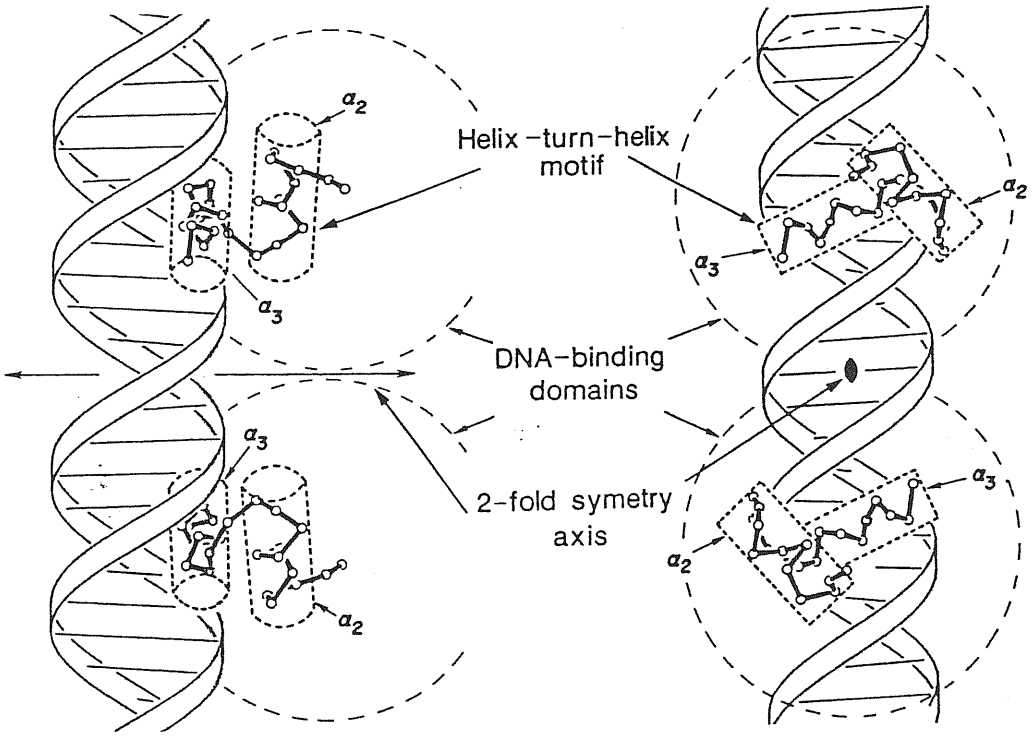


Figure 1.1 Binding of the helix-turn-helix motif to DNA. The recognition helix lies in the major groove of the DNA (from Matthews, 1988).

C-terminal residues Ser-60, Lys-62 and possibly Thr-64 interact with the phosphate backbone providing the major part of the binding strength. Upon binding of Cro, the DNA maintains B-form, although it becomes bent about 40^{o} (Brennan et al., 1990). The results of this study mainly agree with models previously proposed for the interaction of Cro with DNA (Ohlendorf & Matthews, 1983; Pabo & Sauer,1984) Recent structure determination of different repressor-operator complexes has shown that there is similarity in the structure of the domains responsible for the binding to the DNA. All of them contain the helix-turn-helix motif, although there are differences in interaction with the DNA in each case (Anderson et al., 1987; Aggarwal et al., 1988; Jordan & Pabo, 1988; Wolberger et al., 1988). Moreover, in the Trp repressor complex there are no direct contacts between amino acids and bases, most of the direct contacts are made with the DNA phosphate backbone (Otwinowski et al., 1988). This unusual structure could be a consequence of the crystallization condition which does not allow specific complex formation (Matthews, 1988). There is also a possibility that Trp was crystallized with a nonspecific binding site (Staacke et al., 1990).

In eukaryotes the helix-turn-helix motif was first predicted for the family of proteins involved in the development of the fruit fly *Drosophila Melanogaster* (Laughon & Scott, 1984). They contain a highly conserved amino acid region of about 60 amino acids called the homeodomain which is required for binding DNA (Muller et al., 1988; Mihara & Kaiser, 1988).

Similar binding motif occurs in many other transcription regulators in higher organisms (Shepherd et al., 1984; Akam, 1989; Scott et al., 1989). Nucleic magnetic resonance spectroscopic study of the *antennapedia* homeodomain protein confirmed the prediction that the homeodomain is a type of a helix-turn-helix (Quian et al., 1989). Further data has been obtained from the crystallographic study of the complex of the homeodomain of the *engrailed* protein of *Drosophilla* and its DNA binding site (Kissinger et al., 1990).

This study reveals some specificity of homeodomain-DNA interaction. The *engrailed* homeodomain contains three α helices and an extended N-terminal arm. The helices are much longer than the corresponding helices of prokaryotic repressors and there are extensive contacts with the sugar-phosphate backbone (**Figure 1.2A**). The homeodomain makes contacts in both the major and minor grooves. Differently from prokaryotic helix-turn-helix motifs, the residues in contact with the bases are situated near the center of the recognition helix (helix 3) and not towards its N-terminal. The TAAT subsite seems to be the most important part of the homeodomain binding site (Kissinger et al., 1990; **Figure 1.2B**).

Recently, a new motif, known as a POU domain, has been described in some mammalian regulatory proteins such as the pituitary specific protein Pit-1, the octamer binding proteins Oct-1 and Oct-2, and the nematode unc-86 protein (Herr et al.,1988). It consists of a homeodomain and of a 75-82 amino acid POU specific domain (Ruvkun & Finney 1991). The entire

region is involved in DNA binding and although the isolated homeodomain of the Pit-1 and Oct-1 are able to bind the DNA, the affinity and binding specificity are much higher in the presence of the POU domain (Kristie & Sharp, 1990, Ingraham et al. 1990). Selected helix-turn-helix DNA-binding proteins are given in **Table 1.1**.

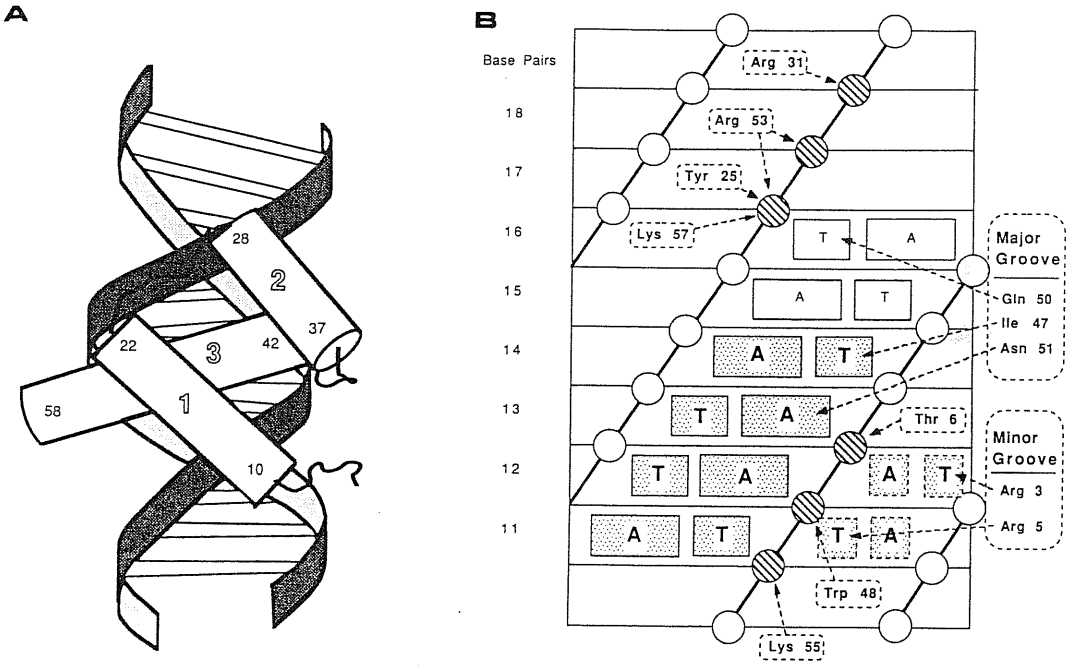


Figure 1.2 The homeodomain-DNA complex. (A) Sketch summarizing the relationship of α helices and the N-terminal arm with respect to the double-helical DNA. (B) Key contacts with the base pairs at the TAAT subsite (from Kissinger et al., 1990).

Table 1.1 DNA-binding motifs shared between prokaryotes and eukaryotes.

Motif	Structure	Role	Factors containing domain	DNA target
Helix-turn-helix	Two α helices separated by a β turn	DNA binding	λcI λCro Trp	TATCACCGCCAGTGGTA TATCACCGCCAGTGGTA CGAACTAGTTAACTAGTACG
Homeodomain	Helix-turn-helix	DNA binding	*antennapedia* *engrailed*	ATTA TAAT
POU	Helix-turn-helix and adjacent helical region	DNA binding	Oct-1 Oct-2 Pit-1	ATGCAAAT ATGCAAAT $T/_AT/_ATATNCAT$

1.1B ZINC FINGERS

The helix-turn-helix is not the only motif that has convenient geometric properties for DNA recognition. So-called zinc finger motifs (**Figure 1.3**) were identified in a number of proteins involved in the regulation of development and in the control of transcription including the yeast transcription factor TFIIIA (Miller et al.,1985), mammalian transcription factor Sp1 (Kadonaga et al., 1987), the yeast ADRI protein (Parraga et al., 1988), and the *Drosophila kruppel* protein (Redemann et al., 1988). All of these proteins require zinc for sequence specific binding to DNA. The model for a

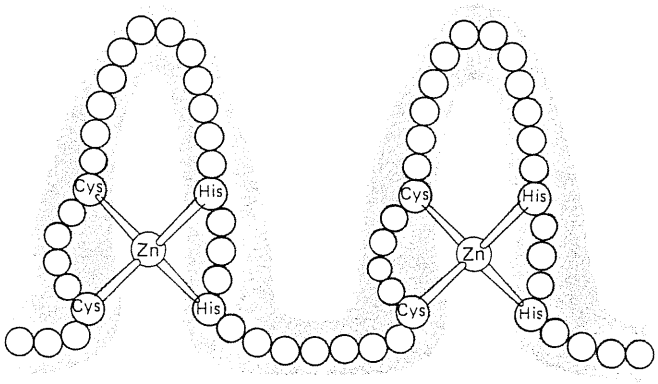


Figure 1.3 A highly schematic model for the general conformation of the "zinc finger" DNA-binding domain (adapted from Klug & Rhodes, 1987).

single finger was first proposed by Berg (1988) and has recently been elucidated using n.m.r. and crystallographic analyses. In the zinc finger motif, at both extremities of the domain, a pair of cysteine and a pair of histidine residues (C_2H_2) tetrahedrally coordinate an atom of zinc. The domain is composed of 25 to 30 amino acid residues with conserved phenylalanine, leucine and several basic residues protruding from the surface of the protein in a finger-like structure (Klug & Rhodes, 1987; Evans & Hollenberg, 1988). The number of fingers ranges from 2 to 37. As shown for the yeast ADR1 protein, a single, isolated finger has low affinity for DNA, indicating that there must be at least two fingers for specific recognition to occur (Parraga et al., 1988).

Concerning the structure of the zinc finger, it has been proposed based on the structure of other metalloproteins (Berg, 1988), that the interacting amino acids do not simply form a loop. The finger region consists of two anti-parallel β sheets and an adjacent α helix, held together by a zinc atom (**Figure 1.4A**). The results of recent n.m.r. and X-ray crystallography studies support this model (Nardelli et al., 1991; Pavletich & Pabo, 1991). The amino acids at the N-terminus of the α helix make contact with bases exposed in the major groove. Each finger recognizes a three base pair subsite. The periodicity of the protein structure matches that of a binding site sequence (**Figure 1.4B**).

There is another class of zinc finger proteins in which each zinc atom is coordinated by four cysteine residues. Multi-cysteine motifs have been identified in several proteins such

as the yeast transcription factor GAL4, adenovirus transcription factor E1A and in steroid-thyroid hormone receptors (Evans & Hollenberg, 1988; Evans, 1988). Differently from C_2H_2 class proteins, only two zinc fingers are found in steroid receptors. It is not possible to substitute

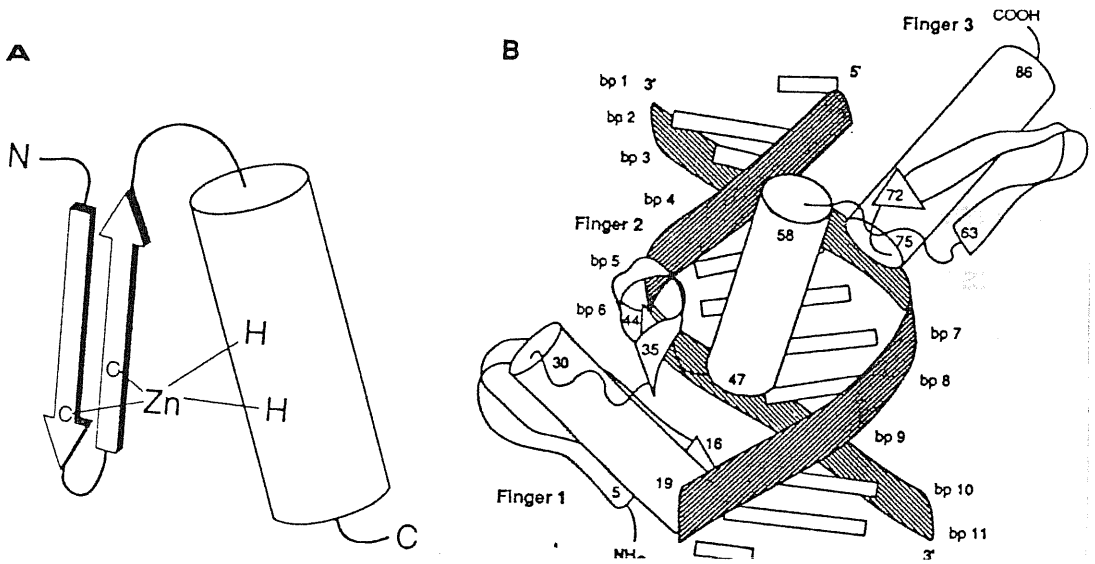


Figure 1.4 Zinc finger-DNA recognition. (A) Model of a single zinc-finger motif (adapted from Berg (1988). (B) Three zinc fingers of the Zif268 protein. Sketch showing the relationship of zinc fingers with respect to each other and with respect to the DNA (from Pavletich & Pabo, 1991).

cysteine with histidine residues, indicating that the two types of fingers are functionally distinct (Green & Chambon, 1987). Structural studies of cysteine fingers revealed the motif consisting of two α helices perpendicular to one another. Residues contacting the DNA are located at the N-terminus of the recognition helix. There is also a region responsible for dimerization of the receptor molecules (Hard et al., 1990; Schwabe et al., 1990). In the dimer molecule, the DNA binding helices are separated by 34 Å, as proposed for the oestrogen receptor dimer (Schwabe et al., 1990). This is an optimal spacing for the specific interaction with a palindromic sequence exposed in adjacent major grooves of the DNA molecule. Proteins having two distinct types of a zinc finger structure are given in **Table1.2.**

1.1C DOMAINS INVOLVING LEUCINE ZIPPER AND HELIX-LOOP-HELIX STRUCTURES

A structure called "leucine zipper" was proposed for several eukaryotic transcription factors such as the CAAT-box binding protein C/EPB, the yeast regulatory protein GCN and the proto-oncogene products Myc, fos and jun (Landschultz et al.,1988b). A common feature of these proteins is a 30 amino acid segment with a periodic repetition of leucine residues at every seventh position. This segment can adopt an α–helix conformation in which adjacent leucine residues are located at every other turn of the helix.

Table 1.2 Two types of a zinc finger motif.

Motif	Structure	Factors	Number of fingers	DNA target
Cysteine-histidine zinc finger (C2H2)	Multiple fingers, each coordinating a zinc atom	Sp1	3	GGGGCGGGG
		Zif268	3	GCGTGGGCG
		Krox-20	3	GCGGGGGCG
		TFIIIA	9	5S gene
		Xfin	27	not known
Cysteine-cysteine zinc finger (C4)	Single pair of fingers, each coordinating a zinc atom	glucocorticoid receptor	2	AGAACANNNTGTTCT
		estrogen receptor	2	AGGTCANNNTGACCT

Hydrophobic surfaces of two such helices adhere tightly to one another in parallel forming a coiled coil (O'Shea et al., 1989). Adjacent to the zipper region, at the N-terminus, there is a stretch of about 30 basic residues. This region is required for the specific binding to the DNA (Gentz et al., 1989; Kouzarides & Ziff, 1989). There are no X-ray or n.m.r. spectroscopic data for basic-zipper (b-Zip) proteins bound to DNA. Recently, a scissors-grip model was proposed for DNA recognition of these proteins (Vilson et al., 1989). Upon dimerization through a leucine zipper region, the basic regions generate a structure well suited for interaction with dyad-symmetric DNA sequences (**Figure 1.5**).

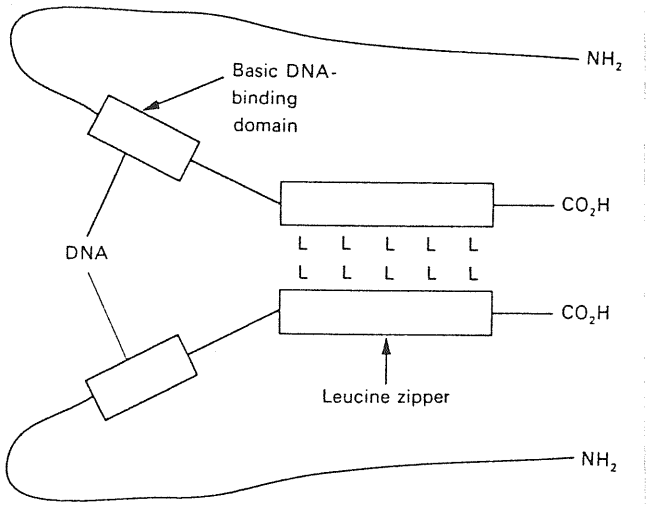


Figure 1.5 Leucine zipper and DNA binding domain. Schematic diagram showing how dimerization via the leucine zipper motif aligns the adjacent DNA binding domain in the correct orientation to bind to DNA (from Latchman, 1990).

The basic DNA binding domain has been identified in regulatory proteins that do not contain leucine zippers, including the muscle regulatory protein MyoD1, the *Drosophila daughterless* protein and the immunoglobulin enhancer binding proteins E12 and E47 (Murre et al.,1989a). In this case, adjacent to the basic domain there is a structure composed of two amphipatic helices which contain hydrophobic residues at every third or fourth position, separated by a non-helical loop (Murre et al., 1989a; Murre et al., 1989b).

This so-called helix-loop-helix motif was first thought to be a DNA binding motif, but later it became clear that it plays a role in mediating protein dimerization similarly to the leucine zipper (Pendergast & Ziff, 1989).

Dimerization plays an essential role in the specific DNA recognition of these proteins. Moreover, a factor can be activated or repressed depending upon its interaction with other proteins, illustrating a potential mechanism of regulating the control of gene expression (Jones, 1990).

There are at least 60 DNA binding proteins that recognize dyad-symmetric sequences via a basic-helix-loop-helix-zipper (bHLH-Zip) domain, including the oncogene Myc and its partner Max (Blackwood & Eisenman, 1991; **Figure 1.6; Table 1.3**). A model has been proposed for the bHLH-Zip protein motif binding to the sequence-specific DNA (Vinson & Garcia, 1992). The motif consists of two separate structural parts, the basic region-helix1 and the helix2-zipper region.

Helix1 and helix2 are separated by a loop of variable length that cannot adopt a helical structure. The helix2-zipper

region forms a coiled coil and is responsible for protein dimerization, the helix1 of each monomer interacts with the sides of the coiled coil while the basic region is involved in specific interactions with the DNA, adopting an α helix secondary structure (**Figure 1.7**). Proteins that do not contain Zip regions such as MyoD1, E12 and the yeast CBF-1 are considered to possess a shortened version of the same structure (Vinson & Garcia, 1992).

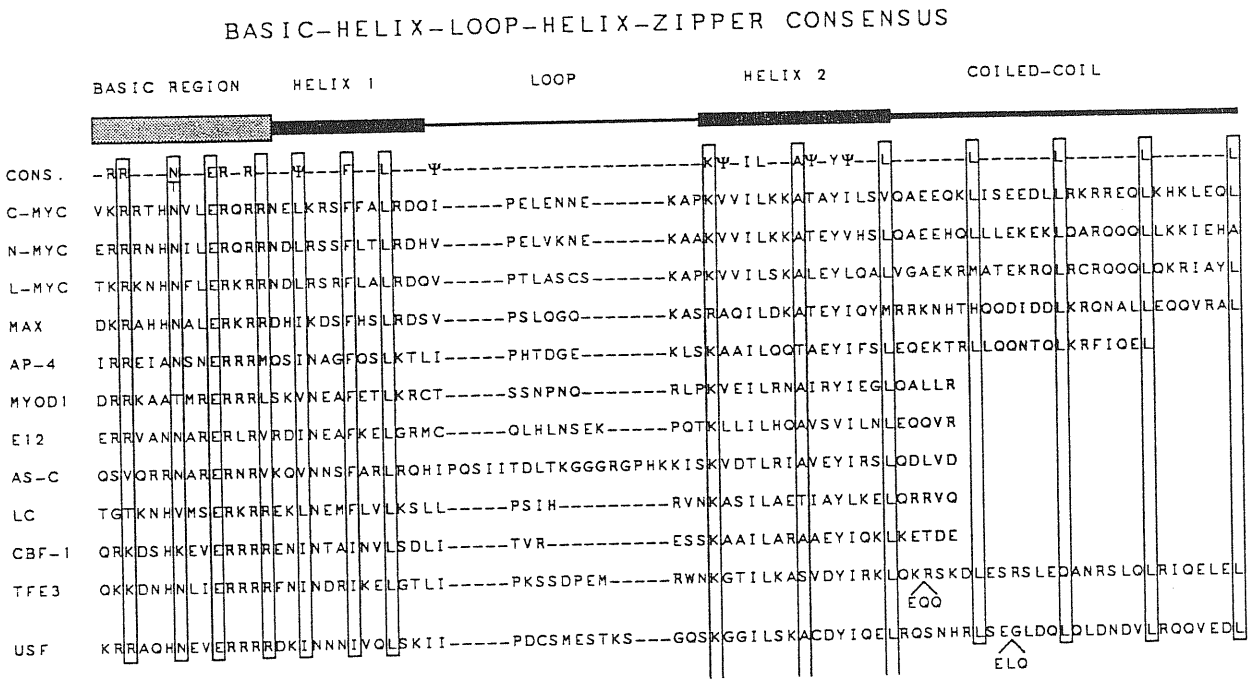


Figure 1.6 Amino acid sequences corresponding to the DNA binding and dimerization domains of seven bHLH-Zip and five bHLH proteins (from Vinson & Garcia, 1992).

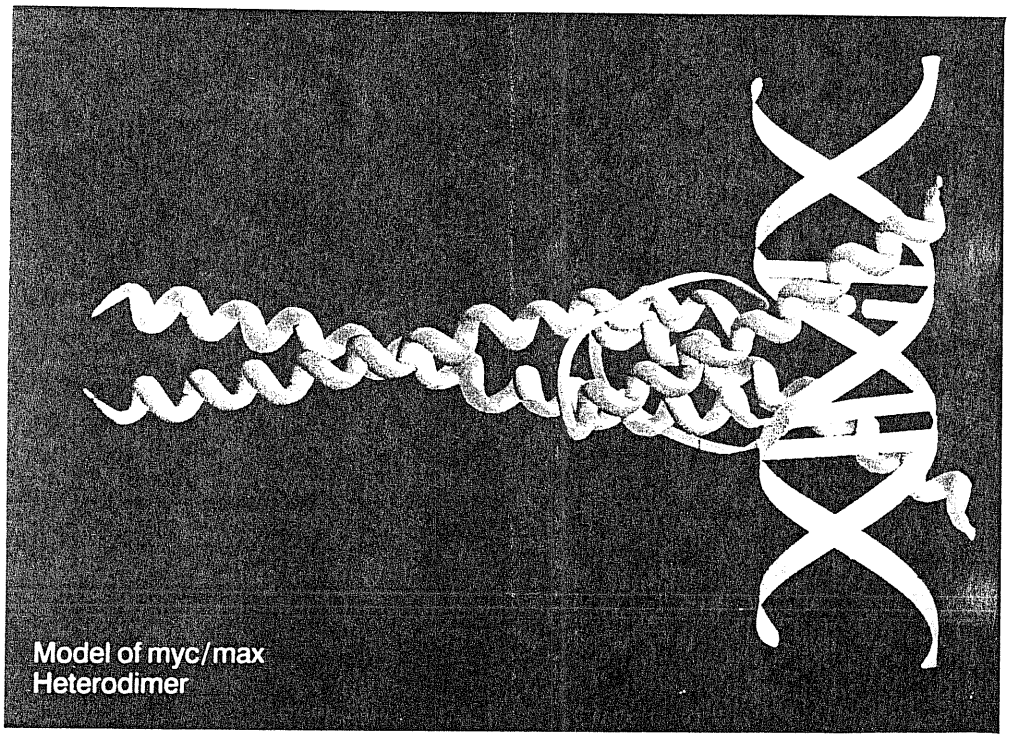


Figure 1.7 A hypothetical model of a Myc-Max dimmeric bHLH-Zip protein interacting with target DNA. The C-terminus is to the left. DNA is 20 bp long and represented as a ribbon. The CA--TG box is represented by cylinders. Helical regions of the model are presented as helices. The loop is presented as a thin ribbon connecting the helices (from Vinson & Garcia, 1992).

Table 1.3 Some proteins contain domains for both DNA binding and protein dimerization.

Factors	Protein dimerization domain	DNA binding domain	DNA target
C/EBP	Leucine coiled-coil	Basic region	GTGG$^{T}/_{A}{}^{T}/_{A}{}^{T}/_{A}$G
fos	Leucine coiled-coil	Basic region	GTGACTCAG
jun	Leucine coiled-coil	Basic region	GTGACTCAG
GCN4	Leucine coiled-coil	Basic region	ATGACTCTT
Myc	Leucine coiled-coil and helix-loop-helix	Basic region	CACGTG
Max	Leucine coiled-coil and helix-loop-helix	Basic region	CACGTG
USF/MLTF	Leucine coiled-coil and helix-loop-helix	Basic region	CACGTG
MyoD	Helix-loop-helix	Basic region	CAACTGAC

1.1D OTHER BINDING MOTIFS

Many regulatory proteins such as the CAAT box-binding factor CTF/NF1 (Mermod et al., 1989) and the transcription factor AP-2 (Williams et al., 1988) do not contain any of the sequence/structure motifs already described. Such proteins and their DNA-binding motifs are given in **Table 1.4.** The AP-2 DNA binding domain consists of a region rich in basic residues and a dimerization domain. Although the overall organization resembles that of bHLH-Zipp proteins, the AP-2 dimerization domain does not share any homology with helix-loop-helix or leucine zipper structures. This novel dimerization domain has been designated a helix-span-helix (HSH) motif (Williams & Tjian, 1991).

Sequence alignment analyses give an opportunity to identify similar motifs in different protein species. Comparison of the sequences of the hepatocyte-enriched transcription factor, HNF-3A and the *Dropsophila fork head* (Fkh) transcriptional regulator has revealed the sequence homology of functionally important regions of two proteins. In the region B, which is necessary for the DNA binding activity of HNF-3A, the amino acid identity with the Fkh region B is 86% (Weigel & Jackle, 1990). This new DNA binding motif, named the fork-head domain is thought to be conserved between Drosophila and mammals (Weigel & Jackle, 1990). There is another, cysteine

Table 1.4 Other binding motifs.

Motif	Structure	Role	Factors containing domain	DNA target
Fork head	Unknown	DNA binding?	Fork-head HNF3A	TATTGA$^{C}/_{T}$TT$^{A}/_{T}$G
ETS	Unknown	DNA binding?	Ets-1 PU.1	$^{G}/_{C}{}^{A}/_{C}$GGA$^{A}/_{T}$G$^{T}/_{C}$ AGAGGAACT
Cold-shock domain (CSD)	b-ribbon?	DNA binding	YB-1	ATTTTTCTGATTGGCCAAAG
Helix-span-helix (HSH)	α-helical?	DNA binding/ dimerization	AP-2	CCC$^{A}/_{C}$N$^{G}/_{C}{}^{G}/_{C}{}^{G}/_{C}$

rich, sequence motif shared by a number of diverse DNA binding proteins (Freemont et al., 1991b). It is similar to the zinc finger motif but it has not yet been proved to be required for DNA binding. It may also play a role in specific protein-protein interactions.

The family of eukaryotic DNA binding proteins including the proto-oncogene Ets-1, the *Drosophila* E74 protein and the mouse PU-1 protein contain a conserved domain of about 85 amino acids, with a highly basic region at the C-terminus and a region rich in leucine at the N-terminus. This new domain, designated ETS domain is involved in DNA binding (Karim et al., 1990).

The cold-shock domain is a putative DNA-binding motif found in prokaryotic cold-shock proteins and eukaryotic Y-box DNA-binding proteins (Wistow,1990; Wolffe et al., 1992). It consists of a highly conserved, 80 amino acid region in the N-terminus which is predicted to adopt a β sheet structure. Differently from other eukaryotic double-stranded DNA-binding proteins, these proteins seem to use one or more β ribbons to contact base pairs in the major or minor grooves of DNA (Doniger et al., 1992).

1.2 DNA SEQUENCE ELEMENTS (BINDING MODULES)

At the 5'-nontranscribed region of genes there are short, conserved sequences which are involved in regulation of transcription of various genes (**Table 1.5**). They are located mostly upstream of the transcription start point, occupying the region of about 100 bp and are modular in design (Lewin, 1990).

These so-called recognition elements or *cis*-regulatory regions of genes interact with sequence-specific DNA-binding proteins. Different genes contain different sets of regulatory sequences and the order in which they are located, as well as the distance between them, seem to be important. Initiation of transcription by RNA polymerase II occurs upon assembly of a complex of transcription factors which bind to sequence elements and interact with each other and with RNA polymerase II. DNA sequence elements are composed of enhancer and upstream promoter elements and a core promoter element. Enhancers are DNA sequences that influence transcription from a long distance. They are composed of independent cis-regulatory sequences and can be located thousands of base pairs away from the transcription start site. Specific trans-acting factors bind to those motifs and probably stimulate transcription by communicating with

Table 1.5 Main eukaryotic transcription elements.

Module	Concensus sequence	Factor binding	Size (kDa)	Tissue specificity
TATA box	TATAAAA	TFIID	~140	ubiquitous
INR	YAYTCYYY	TFII-I	120	
CAAT box	GGCCAATCT	CTF/NF1	52-66	ubiquitous
		C/EBP	42	liver, brain, fat
GC box	GGGCGG	Sp1	105	ubiquitous
octamer	ATGCAAAT	Oct-1/OTF-1	90	ubiquitous
		Oct-2	61	lymphoid cells
CA--TG box	CACPuTG	USF/MLTF	43/44	ubiquitous
		Myc	64/67	
		Max	21	
		TFE3	59	
		MyoD	45	proliferating myoblasts, differentiated myotubes

the promoter-associated transcription complex. The core promoter element contains an initiation site and a TATA element (Lewin, 1991; Latchman, 1991).

1.2A TATA BOX AND INR ELEMENT

The TATA element is an AT-rich sequence (consensus TATAA/TAA/T) which is usually located about 25 bp upstream of the transcription start site (Lewin, 1990). These two elements are necessary and sufficient for the basal level of transcription (Breathnach & Chambon, 1981). The general transcription factors are TFIIA,TFIIB, TFIID, TFIIE AND TFIIF. Only TFIID has a specific binding activity for the TATA element (Nakajima et al., 1988). The genes for yeast and human TFIID have been cloned recently (Hahn et al., 1989; Hoffmann et al.,1990; Kao et al.,1990; Peterson et al., 1990). Yeast TFIID is a polypeptide of 23-27 kD and is unable to support transcriptional activation by an upstream regulator Sp1 (Buratowski et al., 1989; Pugh & Tjian, 1990). The human cloned TFIID was found to migrate at 40 kD in SDS-PAGE (Peterson et al., 1990), while the endogenous protein migrates as a much larger protein of about 140 kD (Reinberg et al., 1987). It now appears that TFIID is a multisubunit complex composed of a TATA-binding protein (TBP) and TATA-binding protein-associated factors (TAFs) (Greenblatt, 1991; Dynlacht et al., 1991; Pugh & Tjian, 1991; Timmers & Sharp, 1991). The cloned 40 kDa protein corresponds to TBP and was

found to be involved in transcription by all three RNA polymerases, pol I, pol II and pol III. This is an interesting and intriguing finding because the different polymerases have their templates organized very differently and therefore were tought to utilize distinct sets of initiation factors (Sharp, 1992; White & Jackson, 1992). Moreover, the TBP protein is required for transcription of genes that do not contain TATA boxes (TATA-less class II and class III genes, and class I genes) (Pugh & Tjian, 1991; White et al., 1992; Comai et al., 1992). The possible explanation could be that in these cases TAFs participate in sequence recognition for promoter specificity and that different combination of TATA-binding protein and TAFs, assembling at different promoters, select an appropriate RNA polymerase (Sharp, 1992; White & Jackson, 1992).

TFIIA stabilizes the TFIID-DNA complex, but its function is not yet completely understood (Sawadogo & Sentenac, 1990). It has been purified by yeast TFIID affinity chromatography and was found to be unable to bind to DNA, although it stimulated transcription *in vitro* from EII and ML promoters (Usuda et al., 1991).

TFIIB binds to this complex and promotes the binding of RNA polymerase II (Buratowski et al., 1989). TFIIF (RAP 30/70) associates with polymerase II and modulates its activity (Greenblatt, 1991). A RNA polymerase II containing affinity column has been used to purify TFIIF factor. It has been shown to possess an ATP-dependent DNA helicase activity. After binding of the TFIIE factor, initiation of RNA synthesis can

occur (Ohkuma et al., 1991; Inostroza et al., 1991; Buratowski et al., 1991).

Recently, two novel activities have been identified, referred to as 700 kD and 40 kD factors, that are required for the basal transcription from adenovirus major late gene core promoter (MLP) (Parvin et al., 1992). It has also been shown that TFIIE is not required for the transcription of the IgH promoter (Parvin et al., 1992). This example indicates that eukaryotic promoters differ in their response to basal transcription factors, suggesting a mechanism for gene regulation.

Not all eukaryotic genes possess the TATA box. It is absent in housekeeping genes and in some tissue-specific genes (Smale & Baltimore, 1989) (**Figure 1.8**). In such cases pyrimidine-rich initiator elements (INR, consensus YAYTCYYY), located near or overlapping the transcription start site, determine the initiation point and constitute the minimal promoter required for basal transcription. A factor, designated TFII-I, that binds specifically to the INR elements in adenovirus ML, HIV-1 and TdT promoters has been identified (Roy et al., 1991). It is a 120 kD protein that also binds to the E-box (CACGTG) sequence. Moreover, it has been found to interact with USF at both, the INR and USF site. It is probable that TFII-I binds to these sites via distinct domains. Its ability to bind to different promoter elements suggests a novel promoter-activation pathway and a way of communication between upstream regulatory factors and the basic transcriptional machinery (Roeder, 1991).

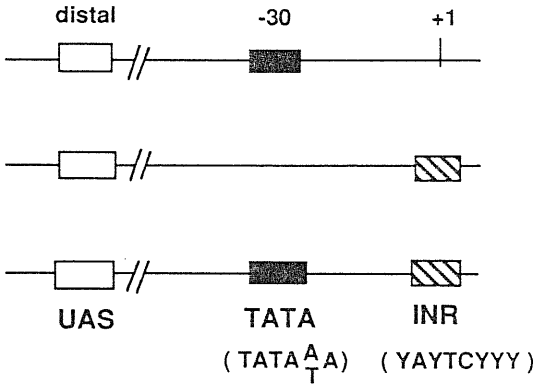


Figure 1.8 Eukaryotic promoters comprised of upstream activation sequence (UAS) elements in conjunction with TATA box and/or initiator (INR) core promoter elements (from Roeder, 1991).

1.2B CAAT BOX

One of the first modules to be identified was the CAAT box. It is a eukaryotic promoter element, located at about -80 and is required for transcription of a variety of cellular and viral genes (Lewin, 1990). In some promoters the CCAAT element interacts with another promoter sequences such as the CCACACCCG sequence in the mouse β-globin promoter (Myers et al., 1986). It is found together with the Sp1 binding sites in the herpes virus thymidine kinase (tk) gene promoter (McKnight et al., 1984; Jones et al., 1985) and

with a heat-shock element in the *Xenopus hsp* (70) gene (Bienz & Pelham, 1986). Transcription factor CTF was found to be responsible for selective recognition of eukaryotic promoters that contain a sequence CCAAT (Jones et al., 1987). It has been purified by sequence specific DNA affinity chromatography with CAAT-box elements from the human α-globin and Ha-ras genes (Jones et al., 1987). Purified CTF consists of a heterologous population of polypeptides of molecular weight of 52-66 kDa. It was found to be able to stimulate transcription of genes containing a CCAAT box and to stimulate initiation of replication of adenovirus DNA (Jones et al., 1987). The nucleotide sequence recognized by CTF matches a portion of the sequence (consensus TGG(A/C)N_5GCCAA) previously described as a binding site for the adenovirus replication protein, nuclear factor 1 (NF1) (Nagata et al., 1982). NF-1 binds a specific nucleotide sequence within the origin of replication of adenovirus and enhances its replication. No difference was detected between CTF and NF1 in their polypeptide composition, functional and immunological properties. These data indicated that CTF was identical to nuclear factor 1 (NF-1). The multiple polypeptide composition of CTF/NF1 could be generated by post-translational modification or by proteolytic degradation of a larger CTF/NF1 protein or could be the product of different but related genes (Jones et al., 1987). A synthetic oligomer corresponding to the expected coding sequence of the CTF peptide1 was selected as a specific DNA probe for screening a HeLa cDNA library (Santoro et al., 1988). Structural analyses

of ten independent cDNA clones suggested that a single gene gives rise to at least three distinct mRNA species which are most likely generated by alternative RNA splicing. The regions of the CTF/NF1 protein responsible for its DNA binding, dimerization, transcriptional and replication activities have been identified (Mermod et al., 1989). CTF/NF1 does not contain any of the previously characterized DNA binding motifs nor does it possess a leucine zipper dimerization motif. CTF/NF1 probably recognize their binding site by using a different type of a DNA binding structure (Mermod et al., 1989). A C-terminal 100 amino acid portion of the protein is needed for transcriptional activation. It contains approximately 25% of proline residues. It probably interacts directly with some components of the general transcription apparatus. A 185 amino acids N-terminal portion is required for DNA recognition, protein dimerization and adenovirus replication. These results show that interaction and mechanisms governing transcriptional activation by CTF are distinct from those mediating DNA replication (Mermod et al., 1989).

The CAAT/enhancer binding protein (C/EBP) has been purified from rat liver nuclei (Johnson et al., 1987). It binds to two regulatory elements that share no sequence similarity: the CAAT box element and the sequence TGTGGA/TA/TA/TG present in some eukaryotic viral enhancers (Johnson et al., 1987). The binding affinity for both promoter and enhancer cis-regulatory elements suggests that C/EBP could mediate interaction between those two elements (Landschulz et al.,

1988a). It has been found to be completely different from CTF/NF1 in its size (42 kDa), heat stability, and sequence specificity (Johnson et al., 1987). Also, the genes coding for CTF/NF-1 and C/EBP are non-homologous to one another (Santoro et al., 1988; Landschulz et al., 1988a). The binding affinity of C/EBP for the CAAT box element suggested that it might be a general ubiquitous regulator of transcription. However, its expression was found to be limited to fully differentiated cells in tissues that metabolize lipids and cholesterol (Birkenmeier et al., 1989).

A variety of regulatory proteins, isolated from different tissues and cell types can bind the CAAT box (Chodos et al., 1988; Raymondjean, 1988). Some of these factors have preferences for particular CAAT box promoters indicating that the same binding module is not always recognized by the same protein. There must be other factors or sequences that influence the preference of the CAAT box binding protein for the CAAT box at the particular promoter.

1.2C GC BOX

GC box sequence is a promoter element present in multiple copies and recognized by a single factor, a ubiquitous transcriptional regulator Sp1. Initially, it was identified as a protein that bound to multiple GGGCGG sequences upstream of a TATA box in the SV40 promoter (Dynan & Tjian, 1983; Gidoni et al., 1985). A purified Sp1 is

composed of two polypeptides of 105 and 95 kDa (Kadonaga & Tjian, 1986). Sp1 is encoded by a single gene. The DNA binding activity region is located at the C-terminus and contains three zinc fingers motifs (Kadonaga et al., 1987). Sp1 stimulates RNA synthesis by RNA polymerase II in all cell types, activating various promoters that contain GC boxes and in that way playing an essential role in constitutive transcription. In the thymidine kinase promoter the CCAAT binding site is located between two Sp1 binding sites (Jones et al., 1985). The presence of different transcription factors binding sites within the same promoter suggest the mechanism of modulation of transcription by interaction of these factors (Jones et al., 1985).

1.2D OCTAMER

A regulatory octamer DNA element ATGCAAAT has been found both in promoters and enhancers of several different cellular genes that are served by either RNA polymerase II or RNA polymerase III. Deletion mutation analyses have shown that this motif is absolutely required for the constitutive expression of the small nuclear RNAs, for the cell cycle specificity of histone H2B expression and the B cell specificity of immunoglobulin gene expression (Sive & Roeder, 1986). In the promoters of immunoglobulin heavy and light chain genes the octamer motif is found approximately 70 bp upstream of the transcription initiation site and

determines B-cell specific activity. It is also present, together with binding sites for several other factors, in the enhancer element of the immunoglobulin heavy-chain gene (Sen & Baltimore, 1986).

Two mammalian proteins, Oct-1 and Oct-2, have been found to recognize the octamer motif. Transcription factor Oct-1 (also called OTF-1, NF-A1, NFIII and OBP100) is a protein of 90 kDa that has been purified from HeLa nuclear extracts and found to stimulate transcription of a histone H2b gene (Fletcher et al., 1987). Oct-2 (or OTF-2 and NF-A2) has been found in extracts from B-cell lines (Staudt et al., 1986). It is a 61 kDa protein that activates transcription of the Ig promoter (Scheiderait et al., 1987). The Oct-2 protein and its mRNA are only synthesized in B-cells where Oct-2 controls the B-cell specificity of the immunoglobulin promoter (Clerc et al., 1988). Both proteins exhibit indistinguishable DNA binding characteristics and both possess nearly identical DNA binding POU domains. Nevertheless, they activate distinct octamer-containing promoters: Oct-1 is involved in activation of snRNA and histone H2B mRNA promoters while Oct-2 activates the octamer motif of B-cell specific mRNA promoters (Schaffner, 1989). Selectve transcriptional activation by Oct-1 and Oct-2 is conferred by activation domains of these two factors and specific protein-protein interactions with different types of RNApolII initiation complexes. This is an example where not only DNA binding specificity, but also activation domains determine the specificity of promoter activation (Tanaka et al., 1992).

1.2E CA--TG BOX

The consensus binding site CA--TG has been identified for the family of eukaryotic regulatory factors that possess helix-loop-helix and/or leucine zipper dimerization domains required for homo or heterodimer formation and efficient binding to DNA (Murre et al., 1989a; Murre et al., 1989b; Beckman et al., 1989; Blackwell et al., 1990; Blackwell & Weintraub, 1990; Carr & Sharp, 1990; Gregor et al., 1990; Prendergast & Ziff, 1991; Blackwood & Eisenman, 1991).

A sequence GGCCACGTGACC has been identified within the upstream activating element of the major late promoter (MLP) of adenovirus. A nuclear protein, upstream stimulatory factor (USF, also MLTF and UEF) has been found to bind to this sequence and stimulate transcription both *in vivo* and *in vitro* of the AdML promoter (Sawadogo & Roeder, 1985; Carthew et al., 1985, Miyamoto et al., 1985). A factor called MLTF has been purified from HeLa cells using DNA affinity chromatography (Chodosh et al., 1986). Proteins in the molecular weight range of 44 -48 kDa, which were eluted and renatured from an SDS-polyacrylamide gel, exhibited MLTF binding and transcription stimulatory activities. Purified USF has two forms, 43 and 44 kDa, which display identical affinity for the MLP upstream sequence (Sawadogo et al., 1988). USF probably corresponds to MLTF and seems to be an

ubiquitous transcription factor. It is present in diverse cell types (Carthew et al., 1987) and is required for the cell type specific transcription of the human growth hormone gene (Peritz et al., 1988) and the rat γ fibrinogen gene (Chodosh et al., 1987). The factor designated NF-μE3, that binds to a slightly different core sequence (CATGTG) within an immunoglobulin heavy-chain enhancer (called μE3 motif), has been identified (Sen & Baltimore, 1986). The purified factor consists of three polypeptides of similar size (between 42.5-45 kDa) (Peterson & Calame, 1989). It is probably highly related or identical to USF/MLTF.

The cDNA has been isolated for the novel μE3 binding factor, designated TFE3 (Beckman et al., 1990). This protein of about 59 kDa is also able to bind to the USF/MLTF binding site. It is highly related to another protein, termed TFEB, isolated by screening a phage λ expression library with a probe containing the binding site for USF/MLTF (Carr & Sharp, 1990). Both proteins contain a basic region adjacent to a helix-loop-helix and leucine zipper (bHLH-Zip) motifs. Genes for the 43 and 44 kDa components of USF have been cloned (Gregor et al., 1990; Sirito et al., 1992). Sequence analysis showed that these two factors also possess a bHLH-Zip domain and bind to the DNA as homo- and heterodimers (Gregor et al., 1990; Sirito et al., 1992). The bHLH-Zip motif is absolutely required for both DNA binding and specific protein-protein interactions of these proteins.

There is a striking structural similarity with c-Myc, a nuclear protooncogene that might be involved in cell

proliferation, mitogenesis and differentiation (Lusher & Eisenman, 1990). Although c-Myc exhibits nonspecific DNA binding activity, the presence of the bHLH-Zipp motif suggested that it might recognize similar DNA sequence elements (Beckmann et al., 1990). Dimers of a chimeric protein constructed of E12 and c-Myc as well as containing a c-Myc basic motif, bound specifically to the upstream stimulatory core sequence CACGTG. Moreover, it was specific for the unmethylated binding site, since methylation of the core CpG inhibited its binding activity (Prendergast & Ziff, 1991). The same result has been obtained by using the technique of selected and amplified binding sequence (SAAB) imprinting, where a carboxyl-terminal fragment of c-Myc bound *in vitro* to a sequence CACGTG in a sequence specific manner (Blackwell et al.,1990).

It has been proposed that c-Myc may form heteroligomeric complexes in vivo in order to act as a sequence specific DNA binding protein (Blackwood & Eisenman, 1991). A 21 kDa protein, termed Max, has been found to form complexes with Myc proteins.

The specific binding activity of the c-Myc-Max complex for the CACGTG core sequence is increased significantly, compared to that of either protein alone (Blackwood & Eisenman, 1991). Max is also a helix-loop-helix-zipper protein. Upon dimerization with c-Myc each member of the heterodimer probably recognizes one half site of the sequence.

SAAB imprinting has been used to select preferred DNA sequences from a pool of partially randomized oligonucleotides for the MyoD and E2A proteins (Blackwell & Weintraub, 1990). It was found that homo and heterodimers of these helix-loop-helix proteins recognize a common consensus sequence CA--TG, although nucleotide differences at internal and surrounding positions are also very important. These results indicate that helix-loop-helix proteins show similarities in specific sequence recognition. Homo- and heterodimer formation enables them to modulate binding sequence preferences. Also, different versions of the CA--TG motif may have different binding properties and be recognized by different protein species (Blackwell & Weintraub, 1990).

1.3 INITIATION OF DNA REPLICATION IN EUKARYOTIC CHROMOSOMES

Replication of double-stranded DNA is a complicated process involving a great number of specific protein-protein and protein-DNA interactions. DNA replicates in a semiconservative mode with each parental DNA strand serving as a template for one new (daughter) template. Thus, the sequence of bases in each new strand is complementary to the base sequence in the old strand being copied. The chromosome of a prokaryotic cell acts as a single unit of replication, which means that two copies of the chromosome are produced by an initiation of replication at a single site on the circular chromosome, followed by propagation of the replication forks and separation of the daughters. A new initiation can take place on the chromosome before the previous cycle is completed, so that multiple replication forks progress around one molecule simultaneously (Kornberg & Baker, 1992). Replicon is defined as a unit of DNA replication consisting of a specific origin, initiating protein(s), elongating proteins and termination mechanisms. Studies in *Escherichia coli* including isolation of various temperature sensitive mutants for DNA replication and

cloning of the replication origin (*ori*C) resulted in a model for initiation at *ori*C (**Figure 1.9**). Initiation of replication consists of several stepwise processes starting with recognition of the origin by a sequence specific origin binding

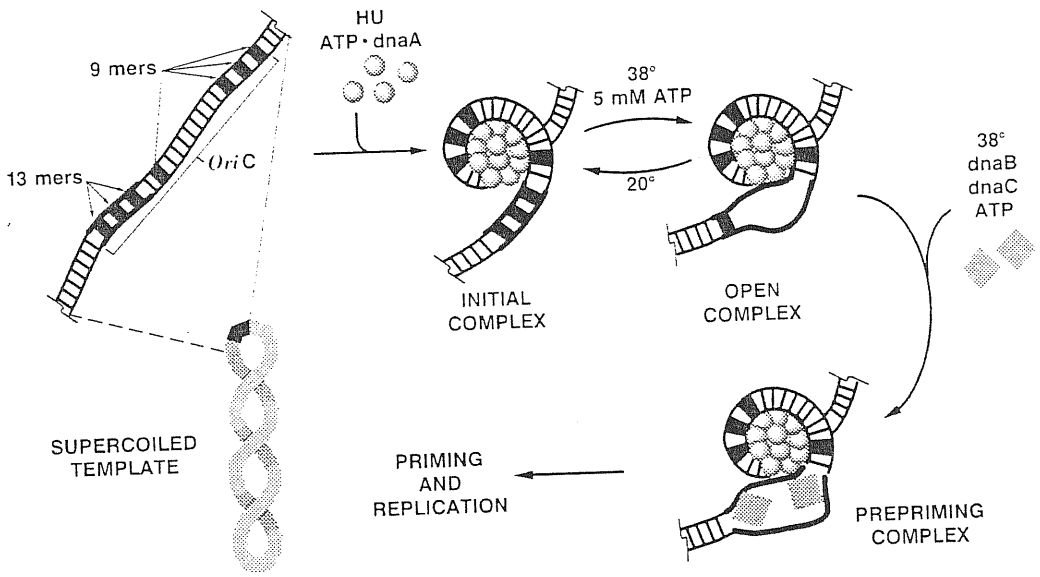


Figure 1.9 A scheme for initiation at *ori*C The *dna*A protein binds the four 9-mers organizing *ori*C around the protein core to form the initial complex. The three 13-mers are then melted serially by *dna*A protein to create the open duplex. The *dna*B-*dna*C complex can now be directed to the 13-mer region to extend the duplex opening and generate a pripriming complex (from Bramhill and Kornberg,1988).

protein, *dna*A. It binds tightly to four 9-mer repeats to form an initiation complex and successively melts the 13-mer, AT-rich repeats. Binding of the initiator protein causes the opening of the DNA duplex which allows the assembly of the multienzyme complex required for the formation of the replication fork. The *dna*B helicase binds the single-stranded 13-mer structure, unwinds the duplex and directs priming of DNA replication (Bramhill & Kornberg, 1988). Similar events occur in a variety of bacteriophages and eukaryotic viruses. In eukaryotic cells, the need to duplicate precisely is complicated by the size of eukaryotic genomes. In order to reproduce their genomes within a realistic time scale, eukaryotic chromosomes initiate replication at multiple sites along their length. Each unit of DNA synthesis (replicon) initiates once in every cell cycle. The rate of elongation varies little in different cell types and therefore the frequency of initiation of replication is the primarily determinant of the length of S-phase (Harland, 1981). The S-phase in somatic cells lasts about 10 hours with infrequent initiation points and ordered gene expression and replicon activation whereas in embryonic cells the replicating regions are very frequent, allowing the DNA to be replicated in an interphase of 3-4 minutes. The possible explanation could be that in embryonic cells different classes of sequences can be used as origins because the DNA is not transcriptionally active. The activity of sequences which specify the initiation of replication could be controlled by binding of specific

inhibitory factors or by DNA modification (Harland, 1981; Jackson, 1990).

1.2A ORIGIN MAPPING

The existence of origins in eukaryotic chromosomes has not been formally demonstrated, mostly because there are no suitable methods for mapping in large and complex mammalian genomes. Numerous experiments have been performed often giving contrasting results. The evidence that initiation does not need to be sequence specific comes from experiments in which circular DNA molecules were injected into *Xenopus laevis* eggs. No specific DNA sequences were required for replication initiation, similarly to embryonic DNA (Harland & Laskey, 1980).

Sequences that might serve as origins of replication have been identified in the yeast *Saccharomyces cerevisiae*, where numerous chromosomal DNA elements have been found to serve as autonomously replicating sequences (ARSs) (Stuhl et al., 1979). Two dimensional gel electrophoretic analysis in the yeast 2 μm plasmid and in ARS1 containing plasmid has demonstrated that a single ARS element is required for replication (Brewer & Fangman, 1987; Huberman et al., 1987). Nevertheless, some experiments revealed that not all ARS elements can be used as chromosomal origins *in vivo* (Umek et al., 1989). The studies of the yeast ribosomal DNA revealed the presence of multiple ARSs that are not used as origins. The same has been observed for the left end of the

chromosome III (Umek et al., 1989). These results show that ARS elements and chromosomal origins do not always coincide. Some ARSs are used as chromosomal origins whereas some are blocked. A possible explanation could be that in the case of two ARS elements close to each other, replication that has started from the first one somehow blocks the ARS activity of the other. Also, transcription as well as neighboring cis-acting suppressing sequences can exert a negative influence on ARS function (Umek et al., 1989). Nevertheless there is still enough evidence that ARSs are potent origins of replication. There are approximately 400 such sequences per haploid genome and they replicate only during S-phase and only once per cell cycle. All the ARSs have similar organization. There is a 11 bp core sequence (element A, consensus $^{A}/_{T}TTTAT^{A}/_{G}TTT^{T}/_{A}$ (Broach et al., 1982). Mutations within this core consensus sequence significantly abolish the ARS function. Also, a ~100 bp region located 3' to the core sequence (element B) is essential for replication. This sequence is rich in A+T and shows little primary sequence homology among ARS elements. It is thought to be required for the melting of the duplex DNA (Umek & Kowalski, 1988).

After the discovery of ARS elements in yeast numerous attempts have been made to detect similar elements in higher eukaryotic cells. In one attempt, random human DNA sequences were inserted into selectable plasmids (Biamonti et al., 1985). The plasmids were introduced into mouse and human cells, but no evidence of autonomous replication of

recombinant plasmids could be detected, in spite of the fact that one origin of replication should be present every 50-100 kb (Biamonti et al., 1985).

Different approaches have been tried to isolate mammalian replication origins by using physical means. In one experiment African green monkey kidney cells were arrested at the G1/S border using the DNA replication inhibitor aphidicolin (Kaufmann et al., 1985). Nascent DNA from the onset of S-phase was isolated and cloned. The cloned fragments were examined for their time of replication by hybridization to cellular DNA fractions synthesized at various intervals of S-phase. Four out of five studied sequences hybridized with early replicating fractions. The cloned fragments should be proximal to the origins and perhaps contain replication origins at or close to their centres. It has been found that six "ors" sequences out of nine contain a 21 bp consensus sequence. There is an abundance of AT rich regions and a high frequency of potential "stem-loop" structures, showing similarities between mammalian replication origins and prokaryotic origins (Kaufmann et al., 1985).

In the studies of an amplified and nonamplified dihydrofolate reductase gene (DHFR) region in Chinese hamster ovary (CHO) cells an initiation zone has been located approximately 17 kb downstream of the gene (Heitz & Hamlin, 1982; Burhans et al., 1986a; Burhans et al., 1986b; Leu and Hamlin, 1989; Anachkova & Hamlin, 1989; Handeli et al., 1989; Vassilev et al., 1990; Burhans et al., 1990) (**Figure 1.10**).

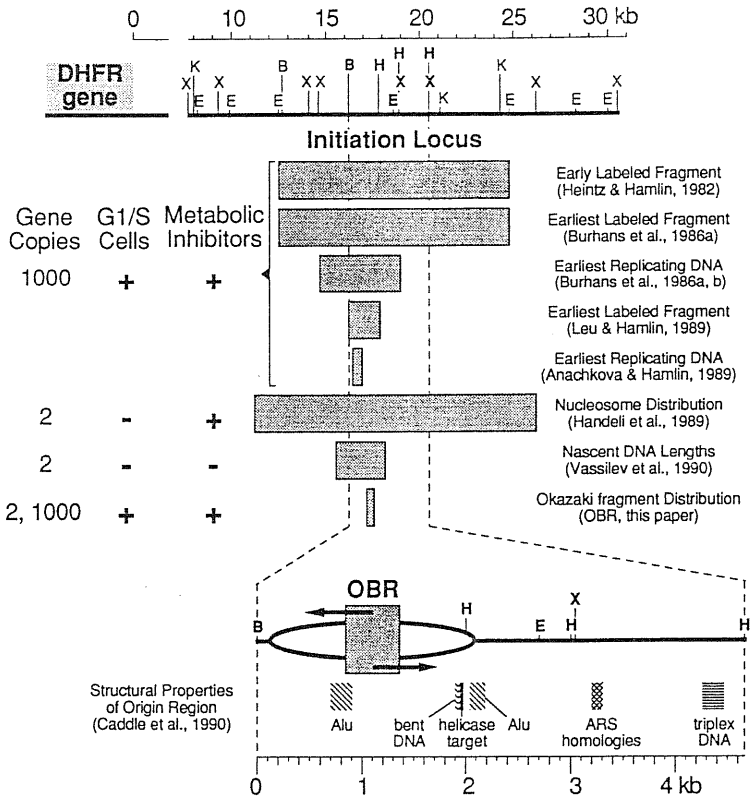


Figure 1.10 Mapping of the initiation zone in the DHFR gene region of CHO and CHOC 400 cells (from Burhans et al., 1990).

In one approach an extremely sensitive polymerase chain reaction (PCR) was used to map initiation points in non synchronized CHO cells containing a single copy of the DHFR gene (Vassilev et al., 1990). Nascent DNA chains were labelled *in vivo,* fractionated according to size, purified by immunoprecipitation with antibromodeoxyuridine antibodies, amplified by PCR using pairs of primers from different regions and then hybridized to the probes homologous to the amplified segments (Vassilev and Johnson, 1989). Specific probes for the segments nearest to the origin hybridize with DNA chains of all sizes, whereas those further away will only hybridize with longer chains. A chromosomal origin of DNA replication has been identified ~17 kb downstream of the DHFR gene with bidirectional DNA replication starting within 2.5 kb long initiation zone (Vassilev et al., 1990). Using the same method an origin of chromosomal replication has been localized ~ 1.5 kb upstream of the first exon of the c-Myc proto-oncogene in HeLa cells (Vassilev & Johnson, 1990). Replication from this origin has been found to proceed bidirectionally within an initiation zone of 2 kb. The initiation zone contains the segments previously reported to exhibit ARS activity (Iguchi-Ariga et al., 1988; McWhinney & Leffak,1990), and a binding site for the nuclear factor CTF/NF1 (Rosenfeld et al., 1989).

To search for the transition site from discontinuous to continuous DNA synthesis in the same DHFR gene region previously reported to contain an "ori", Okazaki fragments

were labeled at the onset of S-phase, isolated and hybridized to cosmid clones containing either 5'-3' or 3'-5' complementary strands of segments spanning the genomic region containing the DHFR gene (Burhans et al., 1990). An origin of bidirectional replication has been located within 450 nucleotides large initiation zone, ~ 17 kb downstream from the 3' end of the DHFR gene in CHO cells (Burhans et al., 1990). Apparently contradictory results have been obtained analyzing replication intermediates in the amplified DHFR gene region using two different and complementary two-dimensional gel electrophoretic techniques (Vaughn et al., 1990). The data suggest that initiation of replication can occur at many random sites within a 28 kb zone and argue against the presence of a small cis-regulatory element (referred to as *ori*) required for initiation. It is also in contrast with known initiation mechanisms in which a protein (or a complex) binds to a cis-element and destabilizes the duplex DNA (Bramhill & Kornberg, 1988).

It has been widely assumed that initiation of DNA replication in higher eukaryotes uses the replication fork mechanism, although only two of its characteristics have ever been demonstrated: the presence of Okazaki fragments and the presence of replication bubbles (Umek et al., 1989). In an attempt to summarize the data about mammalian replication origins, a model of replication initiation has been proposed (Linskens & Huberman, 1990). According to the model, a replication origin should contain an origin of bidirectional replication (a point in which the polarity of Okazaki

fragments switches) that might colocalize with "ori" (a sequence recognized by an initiation protein(s)) and a large initiation zone (10-30 kb) that is required for the initial unwinding and formation of a macrobubble that expands bidirectionally (Linskens & Huberman, 1990).

1.3B ORIGIN BINDING PROTEINS

Although there are some differences among origins of replication in different organisms, a generalized model for the initiation of DNA synthesis can be made (**Figure 1.11**). Replication is a stepwise process, which starts by binding of an initiator protein to a particular origin sequence and then continues with the subsequent melting of an adjacent A+T-rich region (Bramhill & Kornberg, 1988). The A+T region, also called a DNA unwinding element, has special structural properties necessary for easy unwinding, which is an essential feature of an origin. DNA bending is another characteristic that may promote duplex opening and that may be enhanced by the binding of specific proteins (Ramstein & Lawry, 1988). An origin sequence of similar organization has been found in yeast replicators (Eckdahl & Annderson, 1987) and it is likely that specific proteins contribute to DNA unwinding in a similar manner as their counterparts in prokaryotes, bacteriophages and eukaryotic viruses. One of the best characterized yeast origin sequences is ARS1. It has

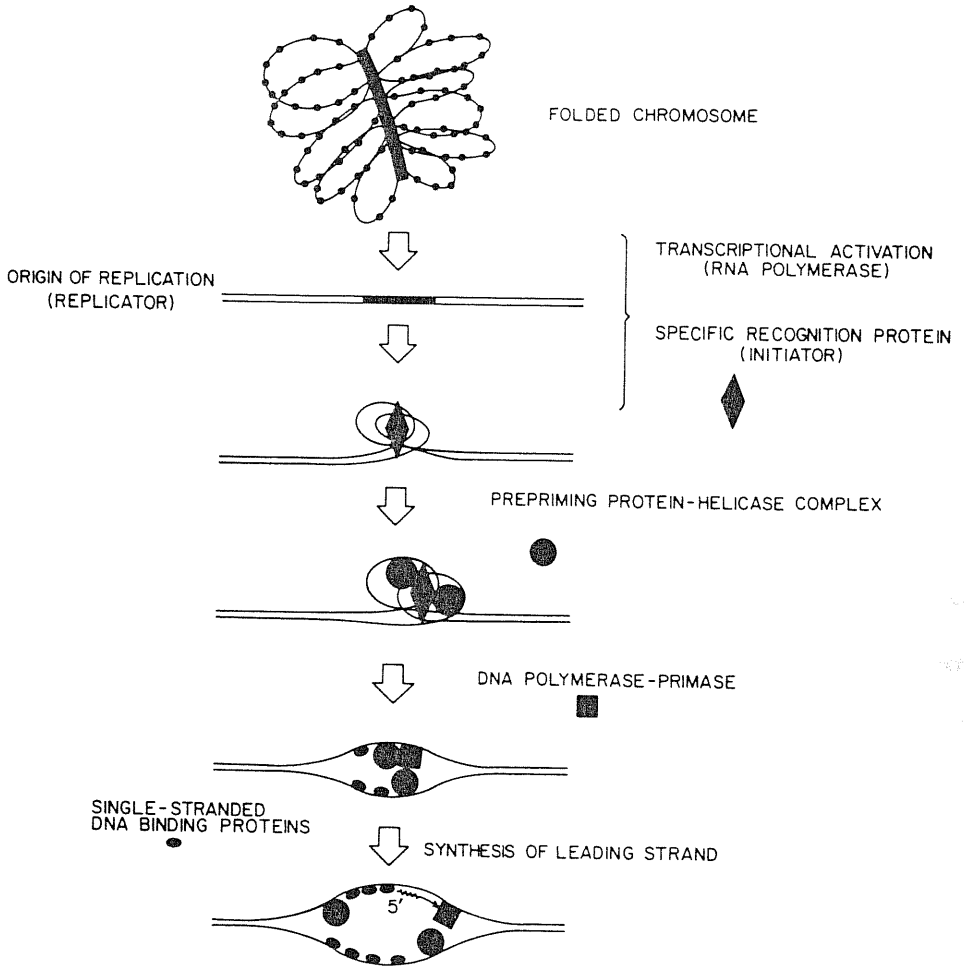


Figure 1.11 The replicon model (from Campbell, 1988).

been divided into two different elements: a 11 bp core consensus sequence (A element) and a B element consisting of three subdomains, B1, B2 and B3 (Marahrens & Stillman, 1992). Contrary to the A element, the B element is not indispensable for the ARS function, but improves its efficiency. A protein, ARS-binding factor1 (ABF1), has been found to bind specifically to the B3 subdomain (Diffley & Stillman, 1988). This protein is known to be a transcriptional activator and its transcriptional activation domain is required for the stimulatory influence of the B3 element on replication efficiency (Della Seta et al., 1990). Recently, the purification of a multiprotein complex that binds to ARS1 has been described and designated as the origin recognition complex (ORC) (Bell & Stillman, 1992)(**Figure 1.12**). The complex is composed of at least six different polypeptides and has a relative molecular mass of ~250 000 kDa. The purification was followed by DNase I footprinting, because the complex could not be detected by gel retardation assay. Using a UV cross-linking assay, three polypeptides (72, 62 and 53 kDa) were found to bind specifically to the origin. An interesting observation was the requirement of ATP for the ORC binding activity, which is not the case for most DNA binding proteins. However, an *E. coli* initiatior protein *dna*A which binds as a multimeric complex to the E. coli origin sequence, also requires ATP to exert its function (Sikemitzu et al., 1987). Association of proteins with ARS1 sequence has also been tested by genomic footprinting (Diffley & Cocker, 1992). Two protected regions have been detected: one over the

B3 element, probably bound by an ABF1 and another over the A and slightly extending to the B1 element, similar to that of the ORC. Also, five regions of hypersensitivity correspond to those generated by ORC (Diffley & Cocker, 1992).

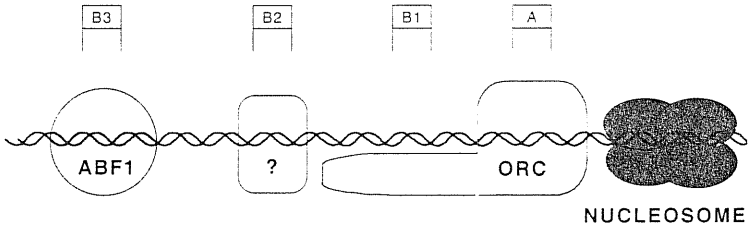


Figure 1.12 Model of possible protein-DNA interactions at ARS1 *in vivo* (from Blow, 1992).

The identification of an initiator protein should explain how replication starts in complex eukaryotic genomes. It should also answer the questions why and how some origins are activated early in S-phase and some late, as has already been postulated for ARS501 which is located at the right end of the S. cerevisiae chromosome V (Ferguson & Fangman,1992). It has been proposed that the late activation of ARS501 is caused by its position at the end of the chromosome, near telomeric sequences (Ferguson & Fangman,1992).

Identification of proteins that are involved in initiation of DNA replication in mammalian chromosomes has been elusive so far. It seems likely that mammalian origins have the same structure and organization as those of lower organisms. Some of the most extensively studied regions containing putative origins are the DHFR gene domain in CHO cells and and a region near the c-Myc gene. Both initiation zones contain the A and T repeats and bent DNA structures (Caddle et al., 1990a). The portion of the DHFR gene domain, which has repeated ATT motifs adjacent to the bent DNA has been found to bind a protein, designated RIP60 (Caddle et al., 1990b). There is a possibility that RIP60 represents an initiator protein in mammalian cells, since in variety of already characterized replication origins an initiator protein binds to a similar sequence adjacent to the bent or A+T rich regions. Moreover, RIP60 forms a complex and copurifies with another protein, RIP100 which has been found to possess an ATP-dependant helicase activity, required for the extensive unwinding of origin DNA during the elongation phase of DNA synthesis (Dailey et al., 1990). Although both RIP60 and RIP100 proteins possess interesting features, their role in DNA replication still remains to be resolved.

Recently, a protein has been purified that binds to the PUR element within a replication initiation zone located upstream of the c-Myc gene (Bergemann & Johnson, 1992). The PUR element (consensus GGNNGAGGGAGARRRR) has been found to be conserved among eukaryotes, although it is not present in all chromosome sequences reported to contain an origin. The

protein, referred to as Pur, is a single polypeptide of ~28 kDa that has a low affinity for double-stranded DNA, but specifically recognizes the purine-rich single strand of the PUR element (Bergemann & Johnson, 1992). The role of the Pur protein in DNA replication remains unclear. The position of its binding site near the initiation zone and the c-Myc transcription regulatory zone suggests that it has a role in replication (as a helix-destabilizing protein) as well as in control of transcription (Bergemann & Johnson,1992).

1.3C TRANSCRIPTION AND REPLICATION

Many questions remain to be resolved in order to explain the initiation of mammalian replication. An interesting issue is the relationship between transcription and replication. According to the data available there is a relationship between replication and gene expression. In one study the temporal order of replication of tissue-specific genes in nine types of differentiated cells has been examined (Hatton et al., 1988). The results obtained showed that actively transcribed genes usually replicate during the first quarter of the S-phase, whereas genes whose transcripts were not detected early replicated later or during the second half of the S-phase. Unlike tissue-specific genes, house-keeping genes can replicate during any interval of the S-phase. The timing and site of replication could be controlled by chromatin organization and by the use of a different subset

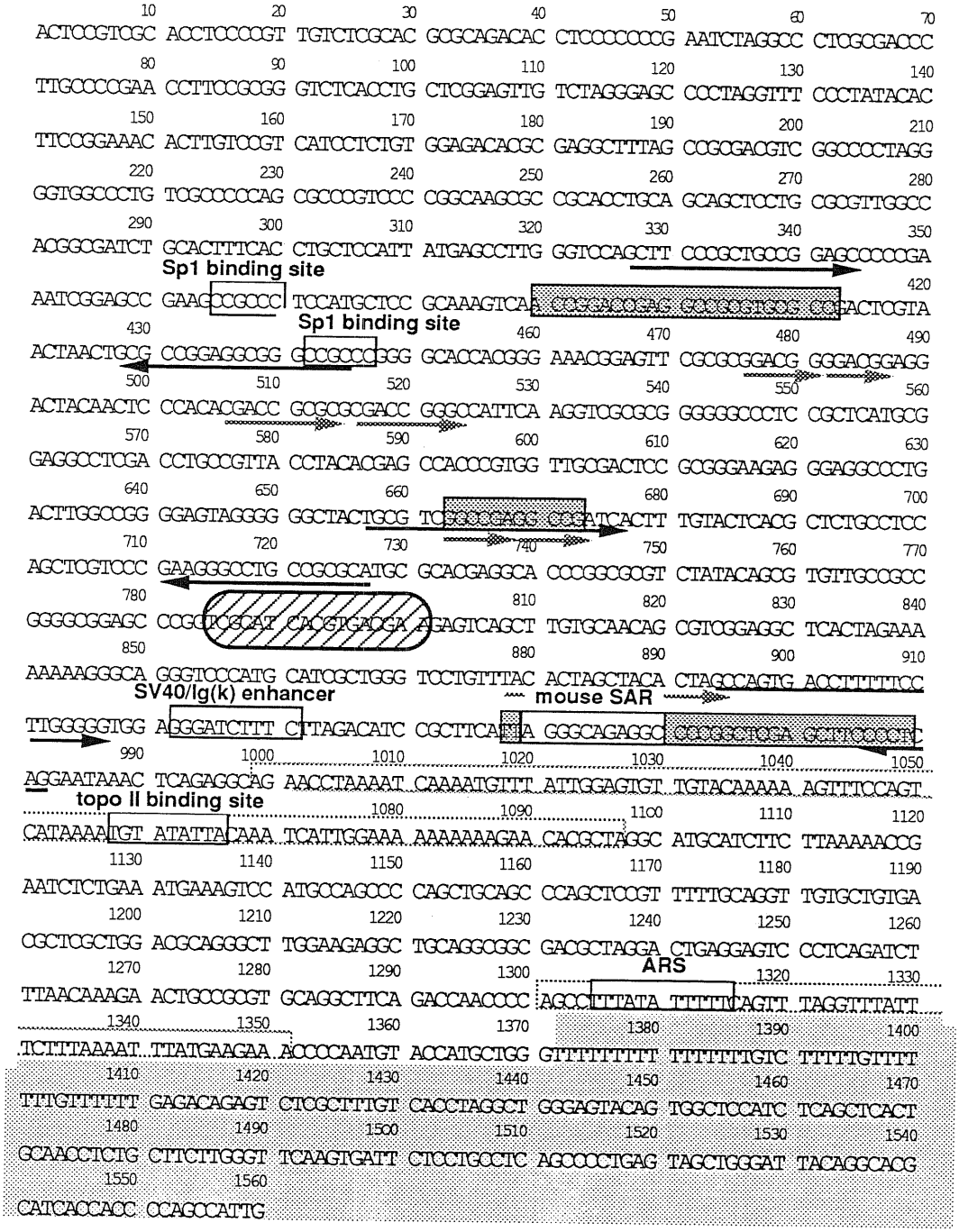
of many potential origins in each differentiated cell type (Hatton et al., 1988). There is still little information about the role of cis-acting transcriptional elements in the regulation of cellular DNA replication. In eukaryotic viruses the situation is more clear (DePamphilis, 1988). In the case of SV 40, polyoma virus (PyV) and adenovirus (Ad) there is a core component where replication begins and an auxiliary component with promoter or enhancer elements that may be involved in transcription as well as replication. Enhancer and promoter elements function in either orientation, but they must be close to the AT-rich end. *Ori* cores from these viruses contain all cis-acting information necessary to initiate replication. But, as in the case of Ad 2 where the cellular protein called nuclear factor 1 (NF 1) facilitates up to 10 fold the interaction of the initiation complex with *ori* core, proteins recognizing transcriptional elements may stabilize binding of initiation proteins to *ori*-core by promoting localized strand separation or modifying the chromatin structure of *ori*-core and therefore making it more accessible for the initiation protein complex. Bovine papilloma virus (BPV) and Epstein-Barr virus (EBV) enhancer and *ori*-core are separated by 300-1000 bp and function independently of their relative orientation and distance. In EBV two *cis*-acting components separated by about 1000 bp are required for *ori*P activity. In mitochondrial DNA, it is a transcription from an upstream promoter that provides RNA primers for initiation of DNA synthesis (DePamphilis, 1988). The requirement of *cis*-acting element for the function of

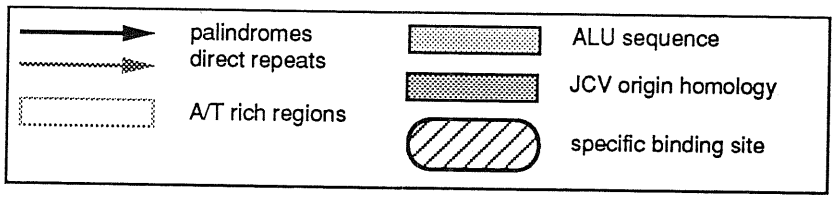
virus and plasmid *ori* regions indicate that there may be a role for cis-acting transcriptional elements also in regulation of eukaryotic DNA replication. In yeast, repression of the yeast silent mating type loci requires *cis*-acting sequences located over 1 kb from regulated promoters. One of the silencers also behaves like an origin of DNA replication having ARS activity. It might be that DNA replication initiated at the silencer is also important for transcriptional repression (DePamphilis, 1988). Promoter and other transcription regulatory elements often overlap with most known eukaryotic replication origins, and therefore transcription and replication should be coordinated. A presence of binding sites for transcription factors AP1 and Oct-1 (OTF1/NFIII) within the DHFR gene origin region suggests an interaction between transcription factors and initiator proteins during the initiation of DNA synthesis (Dailey et al., 1990).

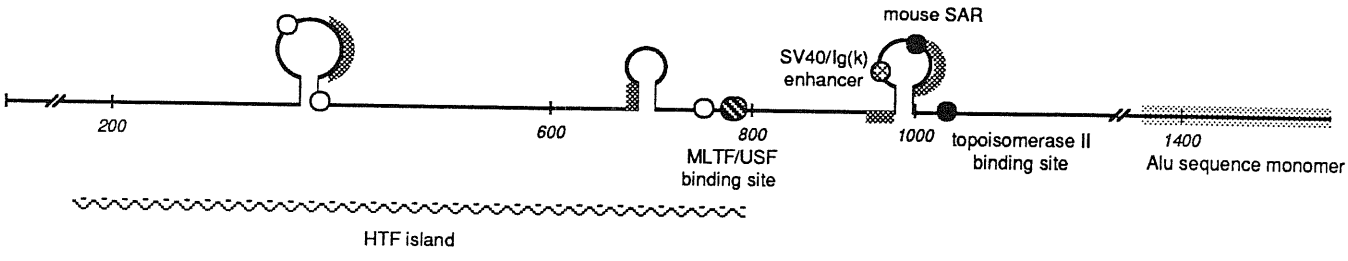
1.4 ABOUT THE PROJECT

In an attempt to identify origins in eukaryotic chromosomes human promyelocitic HL60 cells were synchronized with aphidicholin at the G1/S border and sequences that replicated immediately after the onset of S-phase were isolated (Tribioli et al., 1987). The longest early

replicated DNA fragment, the insert of plasmid pB48, was found to contain a number of characteristic features showing its possible role in the regulation of transcription and initiation of DNA replication such as: a 600 bp long CpG-rich region with the properties of an HTF island, three possible thermodynamically stable stem and loop structures, three sequences homologous (70-100%) to the central (GC-rich) palindrome in the origin of human Papova virus JVC, three Sp1 binding sites located close together, in the stem-and-loop area, a 9 bp sequence homologous to the SV40 and human Ig (κ chain) enhancer and a sequence nearly identical to the MLTF/USF biding site (Tribioli et al., 1987; Falaschi et al., 1988)(**Figure 1.13**). CAT assay experiments were performed to show the presence of an active promoter in pB48 (Falaschi et al., 1988). Transcription was found to occur, but only in one orientation, towards the HTF island. In further studies a chromosomal region of 13.7 kb (designated L30E) was identified by screening a genomic library using pB48 as a probe (Biamonti et al., 1992a). The L30E region maps on chromosome 19, at the subtelomeric p13.3 G-negative band, replicates in the first minute of S-phase in synchronized HL-60 cells and contains two closely spaced, nonoverlapping genes. A replication initiation site is thought to be located between the two transcription units. One of the genes was found to correspond to a B-type nuclear lamin, a member of a family of proteins synthesized only in S-phase, providing


Sp1 binding site
Sp1 binding site
SV40/Ig(k) enhancer
mouse SAR
topo II binding site
ARS


palindromes
direct repeats
A/T rich regions
ALU sequence
JCV origin homology
specific binding site


mouse SAR
SV40/Ig(k) enhancer
200
600
800
1000
1400
MLTF/USF binding site
topoisomerase II binding site
Alu sequence monomer
HTF island

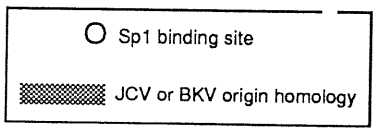

Sp1 binding site
JCV or BKV origin homology

Figure 1.13

evidence that actively transcribed genes are replicated early (Biamonti et al., 1992).

In the search for possible binding sites for specific nuclear factors by band shift assay, retarded bands were observed using the central 521 bp AvaI-AvaI fragment as a probe (Tribioli et al., 1987). In order to analyze the binding site, the 195 bp fragment was submitted to DNase footprinting analysis. A protected region was found to be palindromic in the central 12 nucleotides except for one mismatch and it was homologous to the upstream element in the major late promoter (MLP) of adenovirus 2. Similar sequences are also found in the long terminal repeat (LTR) of HIV-1, in the promoter of the mouse metallothionein I gene, in yeast centromeres, in the TRP1ARS1 element of yeast and in the regulatory region of chorion genes of several insects. The results of competition experiments show that the MLP sequence and pB48 binding site compete for the same factor. Lower, but still significant competition was found with the sequence present in the LTR of HIV-1 (Falaschi et al., 1988). A protein that binds an upstream element in the MLP of adenovirus has been identified and purified by different investigators (Sawadogo & Roeder, 1985; Carthew et al., 1985; Chodosh et al., 1986; Sawadogo et al., 1988). It is able to activate transcription both *in vitro* and *in vivo* (Carthew et al., 1985; Sawadogo, 1988). Since it is present in uninfected cell extracts, it is probably involved in the regulation of transcription of cellular genes.

These preliminary data suggested a role for the pB48 binding site in the transcription of an early replicated chromosomal region, possibly involved in control of DNA replication. The aim of the work was to purify and characterize a nuclear protein(s) that specifically recognize(s) the pB48 binding site.

2. MATERIALS AND METHODS

2.1 DNA PROBES

For DNase I footprinting assay a 165 bp long end-labelled DNA fragment was prepared by HindIII/EcoRI digestion of pL15 plasmid containing a 106 bp AluI-AluI (from nucleotides 703-808) fragment from plasmid pB48 (Tribioli et al., 1987). Labelling was performed with either with [γ-^{32}P]ATP and T4 polynucleotide kinase or with [α-^{32}P]dATP and DNA polymerase I Klenow fragment, according to the published protocols (Maniatis et al., 1982).

Oligonucleotides were synthesized using an Applied Biosystems 380B DNA synthesizer and were further purified on denaturing polyacrylamide gels. A 25 mer blunt-end oligonucleotide 5'-GATCTCGCATCACGTGACGAAGATC-3' (corresponding to the pB48 binding site) was synthesized for gel retardation assays. It was end-labelled and annealed to the complementary oligonucleotide prior to use.

For Southwestern experiments an oligonucleotide containing 4 copies of the pB48 binding site was synthesized together with its complementary strand. The complementary oligonucleotides were annealed, phosphorylated, ligated to

form random sized, tandem, head to tail concatamers ("superprobe") and labelled by nick-translation (Maniatis et al., 1982).

2.2 DNA-BINDING ASSAYS

Gel retardation experiments were performed in 20 µl volumes as described by Tribioli et al., 1987. An aliquot of protein fractions was incubated with 0.2-0.3 ng of the end-labelled oligonucleotide (containing pB 48 binding) site and 1-3 µg of nonspecific DNA competitor (poly[d(I-C)]) in 20 mM HEPES, pH 7.9, 50 mM NaCl, 5 mM $MgCl_2$, 2 mM DTT, O.2 mM EDTA and 5% glycerol. After 30 min of incubation at room temperature the samples were loaded directly onto a low-ionic strength 5% polyacrylamide gel in 0.5 x TBE (25 mM Tris base, 25 mM boric acid, 1mM EDTA, pH 8.3). After electrophoresis the gel was dried and subjected to autoradiography.

For *competition experiments* the conditions were the same, except that unlabeled specific and nonspecific competitor oligonucleotides were incubated with the protein for 15 min at room temperature prior to addition of the labelled probe.

DNase I footprinting assay was performed using a 165 bp fragment containing the pB48 binding site. About 7 ng of the probe was incubated with different protein fractions in 20 mM HEPES, pH 7.9, 50 mM KCl, o.1 mM EDTA, 5 mM DTT, 10

% glycerol for 45 min on ice. After the incubation the reaction mixtures were treated with the appropriate amount of DNase I for 1 min at 20°C. The reaction was stopped by adding 200 µl of stop solution (20 mM Tris, pH 7.5, 0.1 M NaCl, 1% SDS, 5 mM EDTA, 50 µg/ml proteinase K and 25 µg/ml yeast RNA). DNA was precipitated with ethanol, the pellet was washed with 70% ethanol, dried and resuspended in 4 µl formamide dyes. The samples were heated 3 min at 95°C, cooled on ice and loaded on a 6% sequencing gel in 1 x TBE.

Southwestern analyses were carried out essentially as described by Silva et al., 1987, with some modifications. Proteins were separated by SDS gel electrophoresis. The gel was incubated in renaturation buffer (10 mM Tris, pH 7.2, 50 mM NaCl, 20 mM EDTA and 4 M urea) for 1 hour to remove SDS. Proteins were electroblotted to a nitrocellulose filter paper (0.2 or o.4 µm thick) in Tris-glycine buffer (25 mM Tris, 190 mM glycine, pH 8). The filter was treated in binding buffer 10 mM Tris, pH 7.2, 50mM NaCl, 5 mM $MgCl_2$, 0.1 mM EDTA and 1 mM EDTA) containing 5% powdered lean milk for 1 hour, then incubated with the labelled oligonucleotide (10^6 cpm/ml) in the presence of 50 µg/ml of single- and double-stranded sonicated salmon sperm DNA (to prevent the unspecific binding) for 1 hour. The filter was washed three times with binding buffer, dried briefly, covered with a thin foil and exposed for autoradiography.

2.3 PREPARATION OF NUCLEAR EXTRACT AND PROTEIN PURIFICATION

Nuclear extracts were prepared from 300g frozen HeLa cells (obtained from suspension culture) according to the procedure of Dignam et al., 1983. and then precipitated with ammonium sulfate (0-35%).

Protein purification was performed at 4°C. Buffers and solutions were made using sterile water and were sterilized by filtration through 0.2 μm cellulose acetate filters (Nalgene). Protease inhibitors 1mM PMSF, 1mM sodium metabisulfite, 1μM pepstatin and 1μM leupeptin and 1 mM DTT were added to all the buffers immediately before use.

The precipitated proteins were dissolved in buffer D (20 mM HEPES, pH 7.9, 20% glycerol, 0.2 mM EDTA, 12.5 mM $MgCl_2$ and 0.1 M NaCl), dialysed against the same buffer and loaded onto a BIO-REX 70 column at a flow rate 13ml/hr. The column was washed with buffer D and bound proteins were eluted with a linear gradient of 0.1-0.6 M NaCl in buffer D. Protein concentration was monitored by continuous UV absorption at 280 nm. The active fractions were pooled and dialysed against buffer D and loaded onto a HiLoad S Sepharose column (Pharmacia) at a flow rate 130 ml/hr. The proteins were eluted with a linear gradient of 0.1-1.0 M NaCl. The pooled fractions were dialysed against buffer A (20 mM Tris, pH 8, 20% glycerol, 12.5 mM $MgCl_2$, 0.2 mM EDTA and 0.1 M NaCl), filtered through a 0.45 μm syringe filter (Milipore) and

applied to a 8 ml MonoQ column (Pharmacia) at 1 ml/min. The column was then washed with the same buffer and the binding activity was eluted using a linear gradient of 0.1-1.0 M NaCl. After the dialysis aginst buffer S (20 mM HEPES, pH 7.9, 20% glycerol, 12.5 mM $MgCl_2$, 0.2 mM EDTA, 0.1% NP-40 and 0.1 M NaCl) the sample was loaded onto a specific DNA-Sepharose affinity column. To prepare the specific DNA affinity matrix a 21 mer oligonucleotide containing the pB48 binding site was synthesized together with its complementary strand, annealed, phosphorylated with T4 polynucleotide kinase, ligated with T4 ligase and the oligomers were coupled to CNBr-activated Sepharose as described by Kadonaga & Tjian, 1986. After loading of the MonoQ active fraction, the column was washed extensively with buffer S. The bound proteins were eluted with 0.3, 1.0 and 2.0 M NaCl step elutions.
Protein content of different column fractions was analyzed on 12% SDS polyacrylamide gels stained with Coomassie Brilliant Blue or Bio-Rad silver reagent. Protein concentrations were determined by Bradford assay using Bio-Rad protein assay reagent.

2.4 PROTEIN SEQUENCE ANALYSIS

The protein fractions obtained after the 2nd specific affinity column (1M pool in **Figure 1**, lane 6 in **Figure 7a**) were further purified by reverse phase HPLC on an AQUAPORE 300 (ABI, Foster City, USA) 2.1x200 mm column

using a 30 min gradient from 0.1% TFA/water to 90% acetonitrile/10% water, 0.09 TFA. the resulting material was subjected subjected to microsequencing on an ABI 471A pulsed liquid phase sequencer. The sequence obtained (XXXGXKAAVVLXMXVGFXM) was found to be identical with that of the N-terminus of the 86 kDa large subunit or the systemic lupus Ku-antigen (Yaneva et al., 1989; Mimori et al, 1990). In order to determine if the 70 kDa component is identical to the small subunit of the Ku-antigen we first attempted N-terminal sequencing of the electroblotted 70 kDa band. This attempt was unsuccessful which is in accordance with the fact that this subunit is supposed to be acetylated on the N-terminus (Reeves and Sthoeger, 1989). As the amino acid sequence of the Ku-antigen has only one acid-labile Asp-Pro bond situated at residue 341 of the small subunit, the electroblotted 70 kDa band was hydrolyzed with dilute HCl, pH=2.0 at 108 °C for 2 hours. N-terminal sequencing of the cleaved material gave the sequence XXLMLMGFKP, that in fact corresponds to residue positions 342-351 vicinal to the acid-scissile bond of the 70 kDa small subunit of the Ku-antigen.

3. RESULTS

The purification scheme is shown in **Figure 3.1**. The DNA-binding activity was monitored by gel retardation assay, in which a synthetic duplex oligonucleotide containing 17 bp pB48 binding site (**Figure 3.2**) was used as a probe. Using crude HeLa nuclear extract two distinct DNA-protein complexes were observed by virtue of their differential mobilities on nondenaturing polyacrylamide gel. After heat-treatment (3 min at 65^{o}C) the faster migrating complex disappears indicating that the two DNA-protein complexes represent two different protein species with different heat stability (**Figure 3.3**).

3.1 CHROMATOGRAPHIC STEPS

Details of the entire purification are given in **Table 3.1**. Nuclear extracts were prepared from HeLa cells according to the procedure described by Dignam et al., (1983). As a first step, the nuclear extract was concentrated by 35% ammonium sulfate to precipitate nuclear proteins, as well as

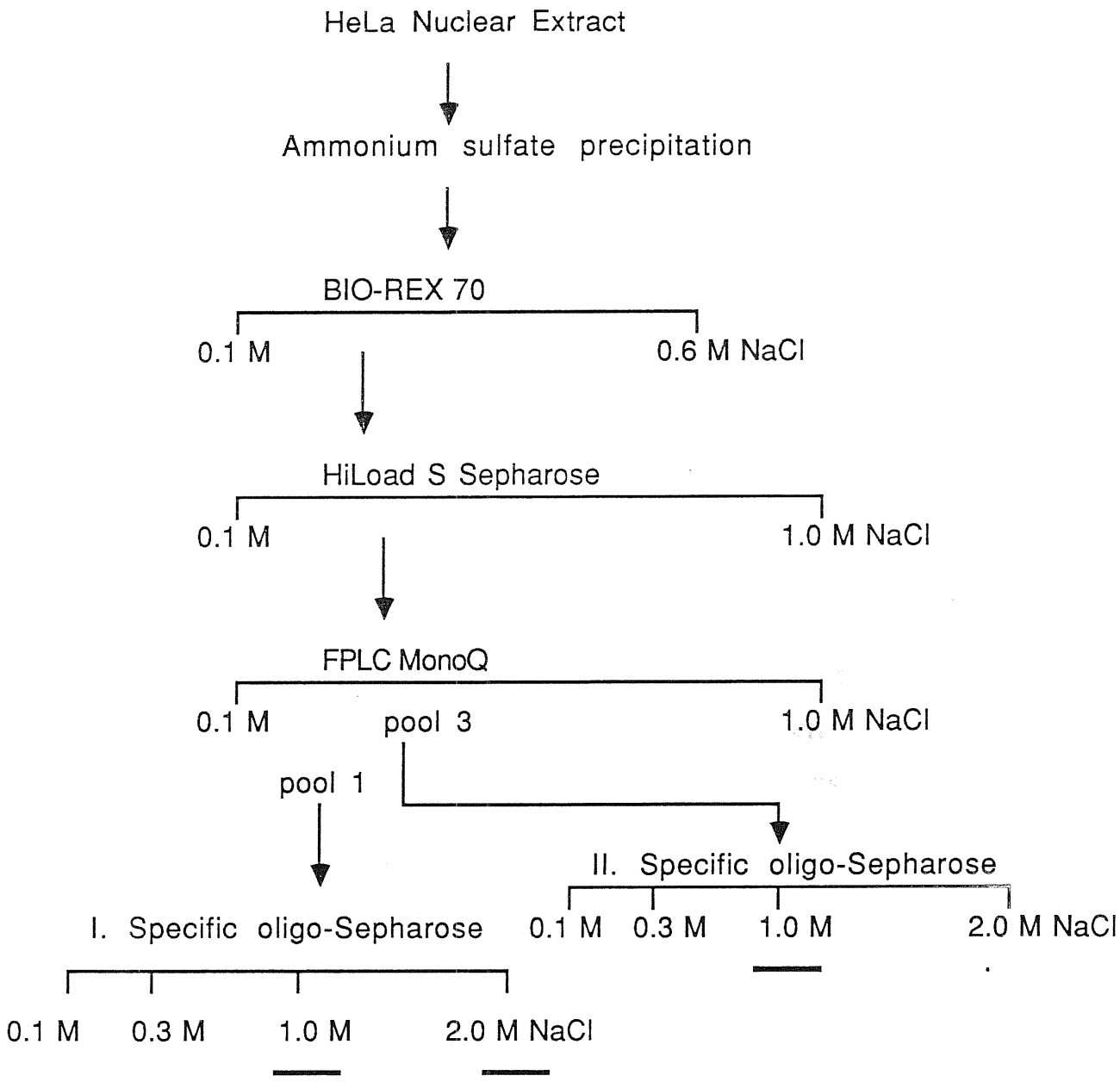


Figure 3.1 Purification scheme. The molarity of NaCl used to elute each fraction is indicated. Binding activity from the specific affinity column was eluted with step elutions.

5' - GATCTCGCATCACGTGACGAAGATC - 3' pB 48 binding
3' - CTAGAGCGTAGTGCACTGCTTCTAG - 5' site

5' - GATCTCGCATCATATGACGAAGATC - 3' mutated
3' - CTAGAGCGTAGTATACTGCTTCTAG - 5'

Figure 3.2 Sequences of the oligonucleotides used are shown. pB48 binding site is derived from an early replicated DNA sequence that contains the conserved CACGTG dyad symmetry binding motif (Tribioli et al., 1987). Mutated nucleotides are indicated with horizontal bars.

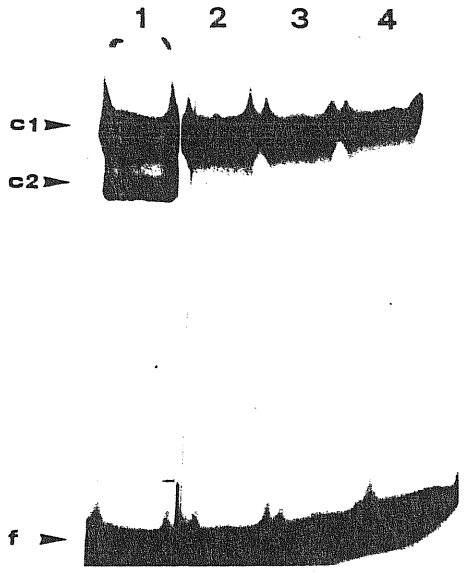


Figure 3.3 Heat stability of the pB48 site binding proteins. 8 µg of nuclear extract, heat treated for 0, 3, 5 and 7 min (lanes 1, 2, 3 and 4, respectively) was incubated with 0.3 ng of pB48 binding site (pB48 BS) oligoprobe and 3 µg of poly [d(I-C)] for 30 min at room temperature. c1-pB48-1 specific band-shift; c2- pB48-2 specific band-shift; f- free DNA probe.

Table 3.1 Purification of pB48 site-binding proteins.

Fraction	Total protein (mg)	Total activity	% Yield	Specific activity (pmol bound/mg of protein)	Purification (fold)
pB48-1					
HeLa nuclear extact	2166	1300	100	0.6	
BIO-REX chromat.	288	262	20.15	0.91	1.52
Hi-Load S-Sepharose	175	245	18.84	1.4	2.3
Mono Q	16.63	116	8.92	7.0	11.67
Specific oligo Sepharose	<0.001	<50	<3.80	50,000	83,000
pB48-2					
HeLa nuclear extract	2166	541	100	0.25	
BIO-REX chromat.	288	180	33.2	0.62	2.48
Hi-Load S-SEpharose	175	165	30.5	0.94	3.76
Mono Q	8.32	158	29.2	18.9	75.60
Specific oligo Sepharose	0.35	150	27.8	429	1716

to decrease the total volume to be handled during the chromatography. Proteins pelleted in this way were dissolved and dialysed against the buffer and then loaded onto a weak cationic exchange resin, BIO-REX 70, equilibrated with buffer containing 0.1 M NaCl. After washing the column with several volumes of buffer, the bound proteins were eluted with a linear NaCl gradient (0.1-0.6 M). This step removed 90% of the loaded proteins and significantly reduced the volume of the columns required for subsequent purification steps. The specific DNA-binding activity was eluted between 0.19 and 0.28 M NaCl concentration (**Figure 3.4a**). The BIO-REX 70 gradient active pool (BRAP) was precipitated with ammonium sulfate (40%), redissolved in buffer without salt, dialyzed and loaded onto a HiLoad S Sepharose column which was selected because of its high loading capacity and fast flow rate. The bound proteins were eluted with a linear 0.1 to 1.0 M NaCl gradient (**Figure 3.4b**). Active fractions, eluted between 0.18-0.29 M NaCl, were dialyzed, applied to a strong anionic exchanger, FPLC MonoQ column and subsequently eluted with a gradient from 0.1 to 1.0 M NaCl (**Figure 3.4c**). This fractionation step resulted in the separation of two different protein species, as it was revealed in gel retardation assay and Southwestern hybridization analysis (**Figure 3.5**). The protein that causes the higher shift eluted between 0.16 to 0.20 M NaCl concentration (pool 1), while the other one, causing the lower shift, was eluted between 0.23 to 0.26 M NaCl (pool 3). Proteins that eluted between 0.20 and 0.23 M NaCl (pool 2) represented a fair mixture of pool 1 and

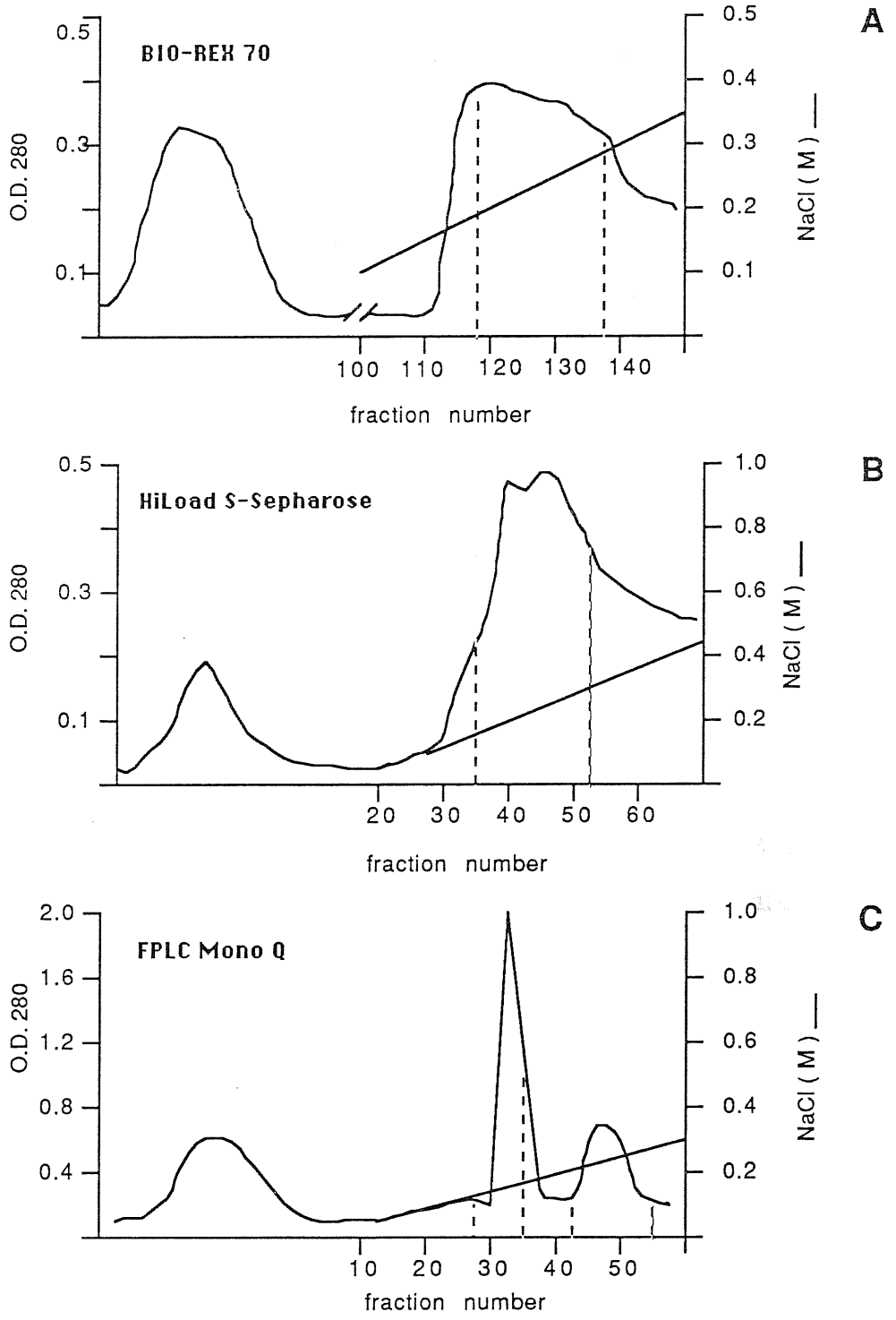


Figure 3.4 Column fractionation of pB48-site binding proteins. Elution profile (O.D. 280 nm) for each column is shown. The dashed lines indicate the pooled fractions containing binding activity. (A) BIO-REX 70, (B) HiLoad S Sepharose, (C) FPLC MonoQ.

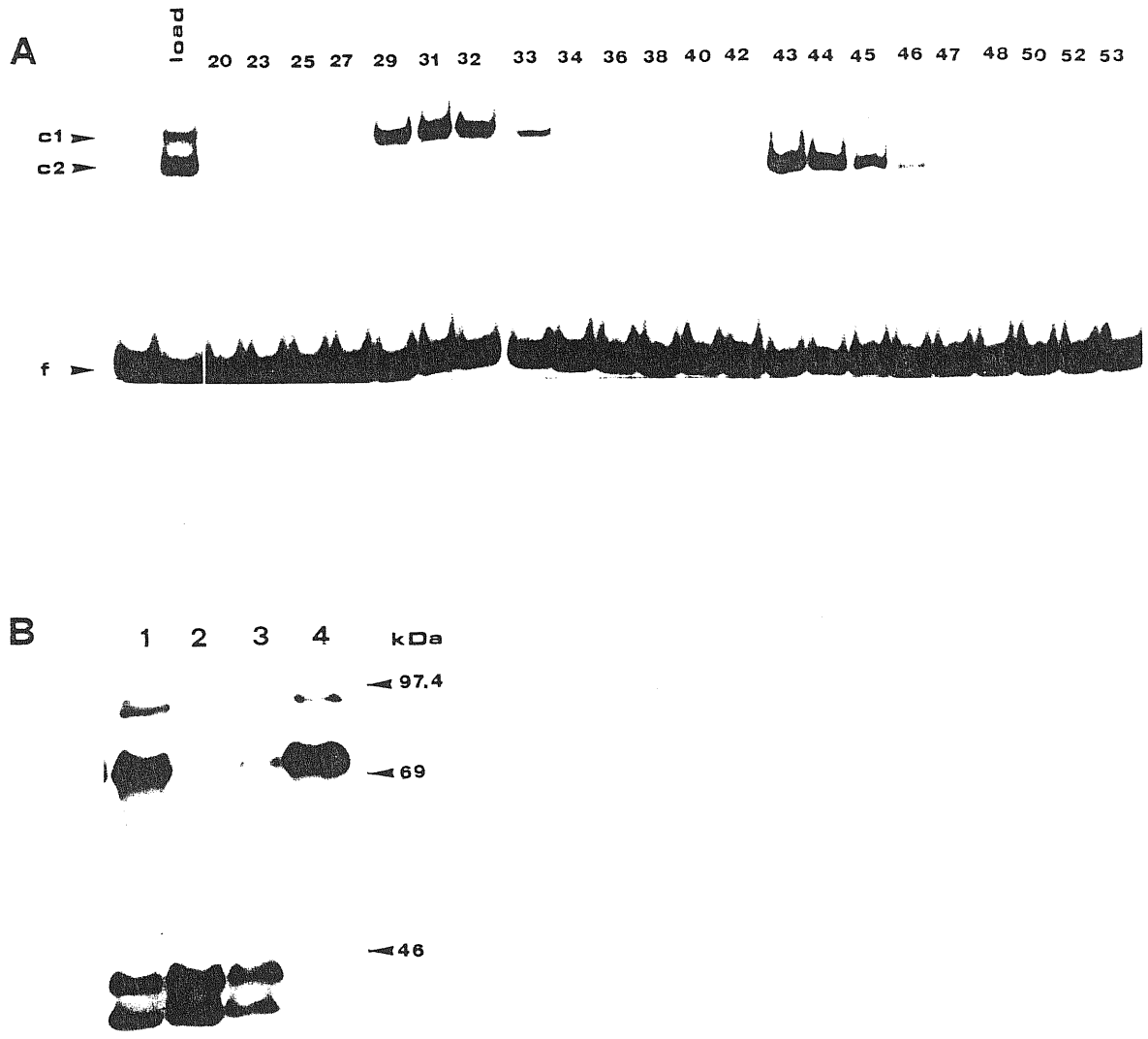


Figure 3.5 (A) Gel retardation assay with MonoQ fractions. The NaCl gradient (0.1-1.0 M) was started at fraction 20. 2 µl aliquotes of indicated fractions were incubated with 0.3 ng of pB48 binding site oligoprobe in the presence of 3µg of poly [d(I-C)]. c1 - pB48-1 specific shift; c2 - pB48-2 specific shift; f - free DNA (B) Southwestern analysis with MonoQ gradient pools. Proteins were separated on a 12% gel, renatured, transferred to a nitrocellulose filter and incubated with a pB48 BS superoligoprobe. Lane 1, HiLoad S Sepharose active pool (60 µg); lane 2, MonoQ pool1 (25 µg); lane 3 MonoQ pool 2 (25 µg); lane 4 MonoQ pool 3 (27 µg). Position of molecular weight markers is indicated.

pool 3 proteins. As a final purification step, the dialysed pool1 and pool 3 were loaded separately on a synthetic oligoaffinity column containing the pB48 binding site. The column was extensively washed with buffer containing 0.1 M NaCl and the proteins were eluted with 0.3, 1.0 and 2.0 M NaCl steps. In both cases the binding activity was eluted at high salt concentration (**Figure 3.6**). These results indicate that proteins from both pool 1 and pool 3 were able to bind with high affinity to the oligo-DNA-Sepharose matrix. Samples from each of the MonoQ pools and affinity column fractions were run on SDS polyacrylamide gels (**Figure 3.7a**). As can be seen from the MonoQ pools, majority of proteins were not able to bind to the affinity matrix. Affinity purified fractions from pool 3 contained two dominant polypeptides of 72 and 87 kDa (**Figure 3.7a**, lane 6). These proteins were relatively abundant and were showing a faster migrating DNA-protein complex in gel retardation assay. In Southwestern experiments, 72 kDa protein bound more strongly (or was present in a larger amount), while the signal of 87 kDa protein was considerably weaker (**Figure 3.7b**, lanes 2 and 5). It is possible that the 87 kDa component contributes less in the DNA binding, explaining the results of SDS-PAGE where both the 72 and 87 kDa proteins are present in approximately equal amounts. Pool 1 affinity purified proteins were much less abundant. After silver staining two bands were observed, one of about 40 kDa and the other very faint around 44 kDa (**Figure 3.7a**, lane 2). At the same time Southwestern analysis revealed the presence of two proteins bands

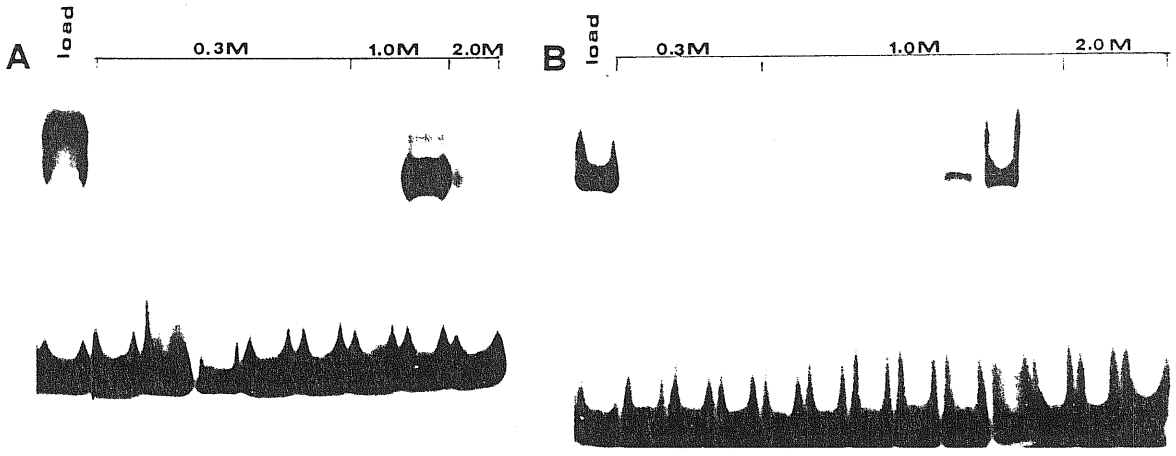


Figure 3.6 Affinity purification of pB48-site binding proteins. Gel retardation assays. 0.3 ng of pB48 BS oligoprobe was incubated with 4 µl of indicated protein fractions and 1 µg of poly[d(I-C)]. (A) 1st specific affinity chromatography (loaded pool 1). (B) 2nd specific affinity chromatography (loaded pool 3).

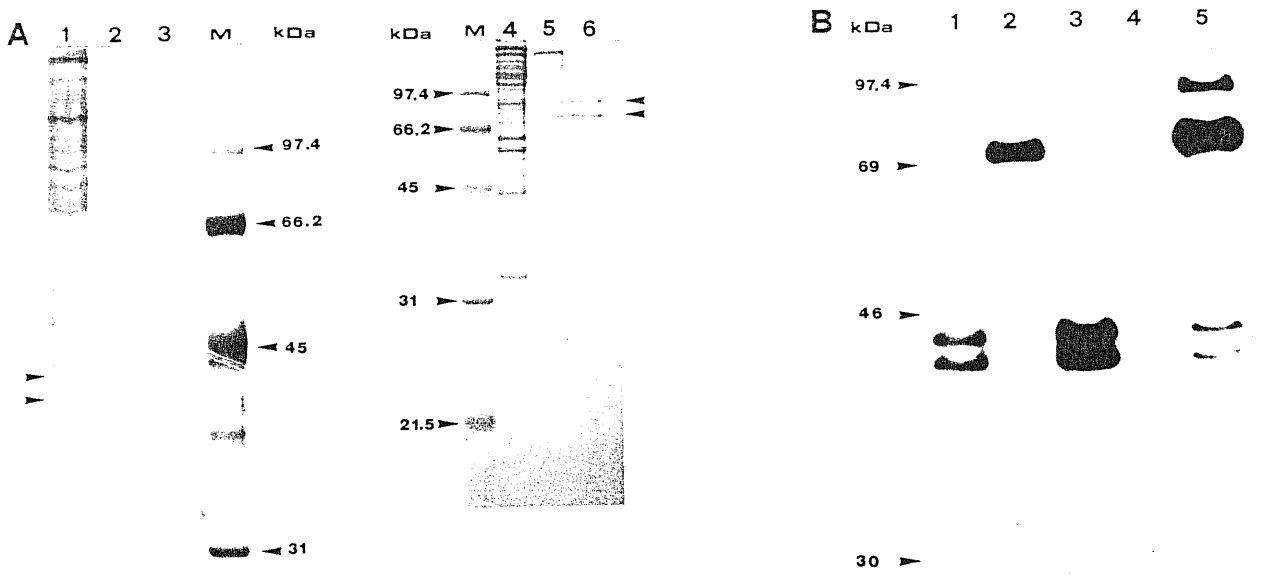


Figure 3.7 Affinity purification of pB48-site binding proteins. (A) SDS-PAGE. Samples were concentrated by ultrafiltration, electrophoresed on 12% gel and silver or Coomassie Blue stained. Lane 1, pool 1 (25 µg); lane 2, 1st specific 1M pool (< 1 µg); lane 3, 1st specific 2M pool (< 1 µg); lane 4, pool 2 (12 µg); lane 5, pool 3 (15 µg); lane 6, 2nd specific 1M pool (3.5 µg); M, molecular weight markers. (B) Southwestern analysis. Concentrated samples were electrophoresed, renatured and blotted to nitrocellulose filter. The filter containing transferred proteins was incubated with pB48 BS superoligoprobe. Lane 1, pool 1 (25 µg); lane 2, pool 3 (27 µg); lane 3, 1st specific 1M pool (< 1 µg); lane 4, 1st specific 2M pool (< 1 µg), lane 5, 2nd specific 1M pool (3.5 µg). Position of molecular weight markers is indicated.

(41.5 and 44 kDa) that bound to the probe with equal intensity (**Figure 3.7b**, lanes 1, 3 and 4). These data suggest that 41.5/44 kDa (pB48-1) and 72/87 kDa (pB48-2) polypeptides might bind to the DNA as heterodimeric complexes.

3.2 CHARACTERIZATION OF THE PURIFIED FRACTIONS

In the course of purification the protein fractions were tested in different DNA-binding assays in order to obtain more data about the characteristics of the pB48-site binding proteins. It was observed previously that there are more proteins from crude nuclear extract that bind to the pB48 binding site (Giacca et al., 1989). Southwestern experiments were performed using a mutated oligonucleotide as a probe. The sequence of the mutated oligonucleotide is shown in **Figure 3.2**. Both protein species seem to bind specifically to the pB48 probe, since the experiment with a mutated probe gave negative results (**Figure 3.8**).

The ability of pB48-1 and pB48-2 proteins to bind specifically to the pB48 binding site was also tested in gel retardation competition assays using unlabeled pB48 binding site and mutated oligonucleotides. Binding of pB48-1 was decreased in proportion to the concentration of unlabeled specific oligonucleotide while the mutated oligonucleotide was not able to compete significantly (**Figure 3.9**). In the case of pB48-2 and under usual experimental conditions, band shift activity was observed even in the presence of high

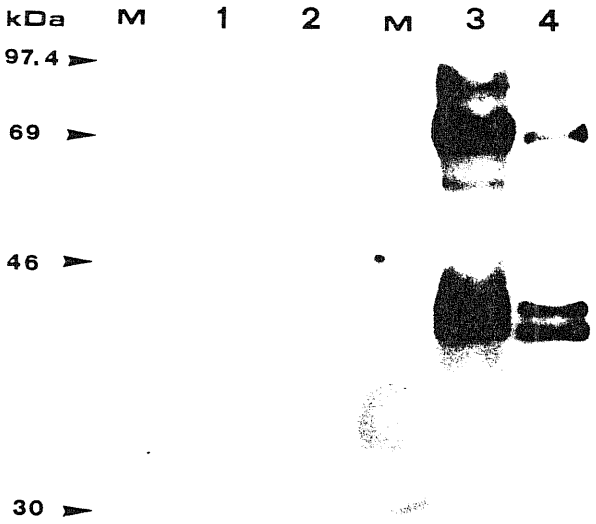


Figure 3.8 Specificity of the proteins for the pB48 BS. In Southwestern analysis 80 µg of HeLa nuclear extract (lane 1 and 3) and 25 µg of pool 1 + pool 3 (lanes 2 and 4) were electrophoresed on 12% gel, renatured, transferred to nitrocellulose filter and incubated with mutated (lanes 1 and 2) or pB48 BS (lanes 3 and 4) superoligoprobe. M - radioactive rainbow molecular weight markers.

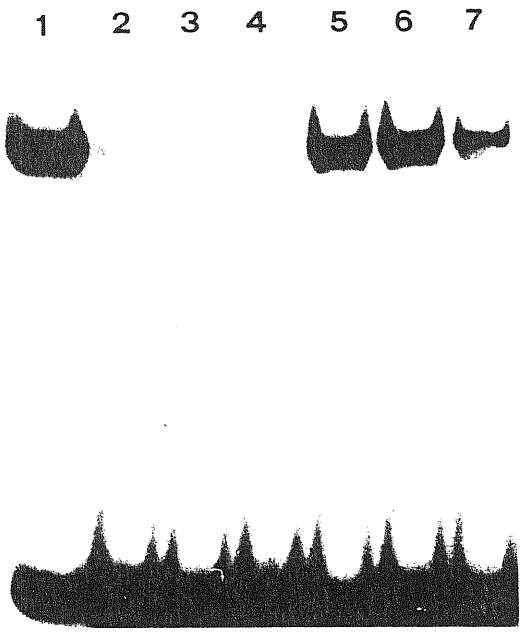


Figure 3.9 Gel retardation competition assay. 1.2 µg of pool 1 was preincubated with 0 ng (lane 1) and 20, 100 and 200 ng of unlabeled pB48 BS oligonucleotide (lanes 2,3 and 4, respectively) and with 20, 100 and 200 ng of mutated oligonucleotide (lanes 5,6 and 7, respectively) in the presence of 3 µg of poly[d(I-C)]. After 15 min, 0.2 ng of labelled pB48 BS oligo was added and the reaction mixture was incubated for a further 30 min at room temperature.

concentration of the unlabeled specific oligonucleotide sequence. Similar results were obtained using different oligo-sequences as unlabeled competitors (data not shown). However, when the pB48-2 protein was preincubated with increasing concentration of poly [d(I-C)] before the adding of the specific probe, the retarded complex was not formed (**Figure 3.10**). These competition assays showed the apparent lack of sequence specificity of the 72/87 polypeptides. This finding is in contradiction with the sequence specificity indicated in the Southwestern experiments and with the fact that the 72/87 polypeptides were strongly bound to the specific DNA-affinity matrix and were eluted at high salt concentration.

3.3 COMPARISON BETWEEN pB48 SITE-BINDING PROTEINS AND OTHER PROTEINS

The molecular weight of pool 1 purified proteins and their high specificity for the CACGTG box within the binding site suggested the possible relationship to the USF/MLTF transcription factor that bind to an upstream element of Adenovirus 2. Furthermore, previous experiments showed the ability of USF/MLTF binding site to compete for the formation of the retarded complex in gel retardation assays (Falaschi et al., 1988). Polyclonal serum against 43 kDa USF component (kindly supplied by R.G. Roeder) was used

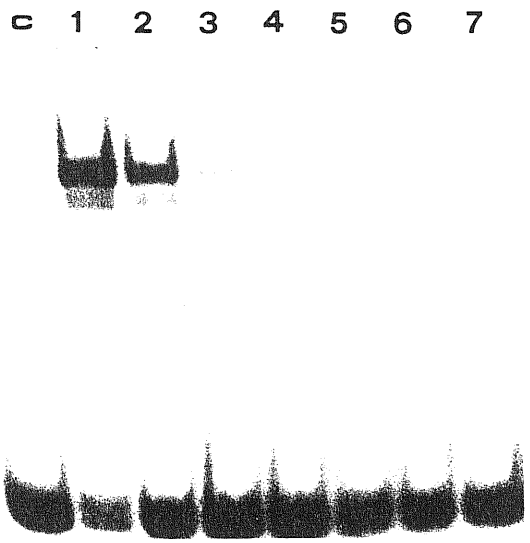


Figure 3.10 Gel retardation competition assay with purified pB48-2 protein. Approximately 40 ng of purified 72/87 kDa protein was preincubated with 0, 1, 5, 10, 50, 100 and 500 ng of poly [d(I-C)] (lanes 1-7, respectively). After 15 min, 0.2 ng of labelled pB48 BS oligo was added and incubation continued for a further 30 min. c- control, no protein.

in gel supershift assays. The pB48-1 specific gel retardation complex was supershifted, while no effect was observed in the case of the pB48-2 protein-DNA complex. Also, the form of a retarded complex formed by pB48-1 showed the striking similarity with that formed by the recombinant USF protein (**Figure 3.11**). These results indicate that pB48-1 is immunologically related or maybe even identical to the 43 kDa component of USF. Also, footprinting analysis of purified pB48-1 gave a protection pattern very similar to that of rUSF 43 kDa (**Figure 3.12a**, lanes 4 and 5 and **figure 3.12b** lane 7).

The N-terminal sequence of the pB48-2 protein 87 kDa component was found to be identical to that of the larger subunit of a DNA-binding protein, called the systemic lupus Ku antigen, recognized by autoantibodies from patients with systemic lupus erythomatosus (Mimori et al., 1986, Mimori et al., 1990). At the same time, the smaller subunit did not give any N-terminal signal, which is in accordance with the finding that the 70kDa subunit of Ku is acetylated (Reeves & Sthoeger et al., 1989). However, an internal sequence of the smaller subunit of was determined and was found to perfectly match with the respective Ku-sequence. On the basis of these partial sequence data we can conclude that the pB48-2 polypeptide is identical with or closely resembling to the Ku-antigen.

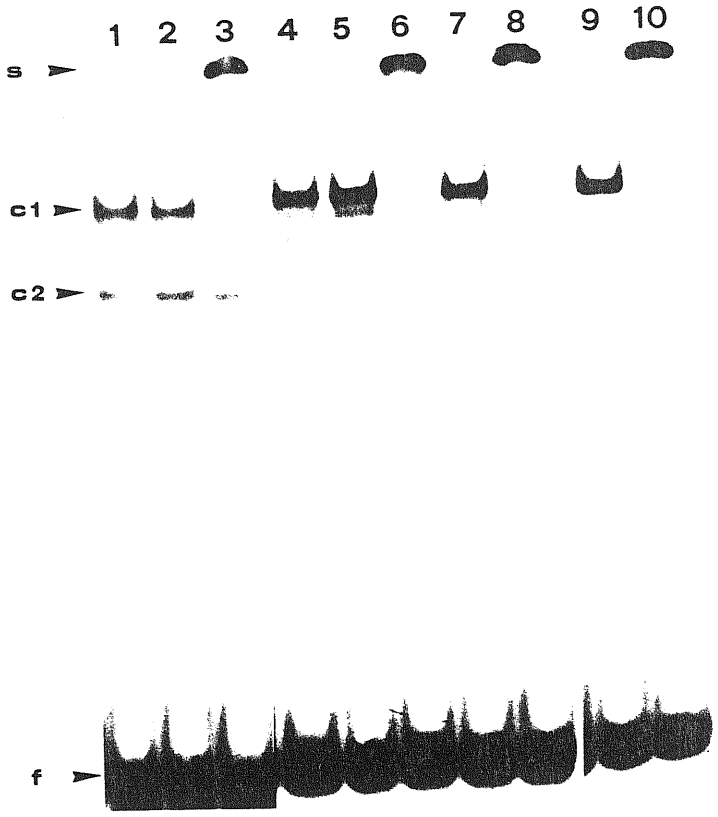


Figure 3.11 Antigenic cross-reactivity of pB48-1 with the rUSF-43kDa component. 2 μg of BIO-REX active pool (lanes 1-3), 0.5 μg of pool 1 (lanes 4-8) and 0.5 ng of rUSF were incubated with 0.2 ng of labelled pB48 BS oligo, 2 μg (lanes 1-8) or 0.5 μg (lanes 9 and 10) of poly [d(I-C)] and 0.1 μl of either naive (lanes 2, 5, 7, 9) or immune serum (lanes 3, 6, 8, 10). Antisera used were obtained against full-size USF-43kDa (residues 1-310) (lanes 3, 8, 10) or a portion of it (residues 18-105) (lane 6). s- supershift complex; c1- pB48-1 specific band-shift; c2- pB48-2 specific band-shift; f- free DNA probe.

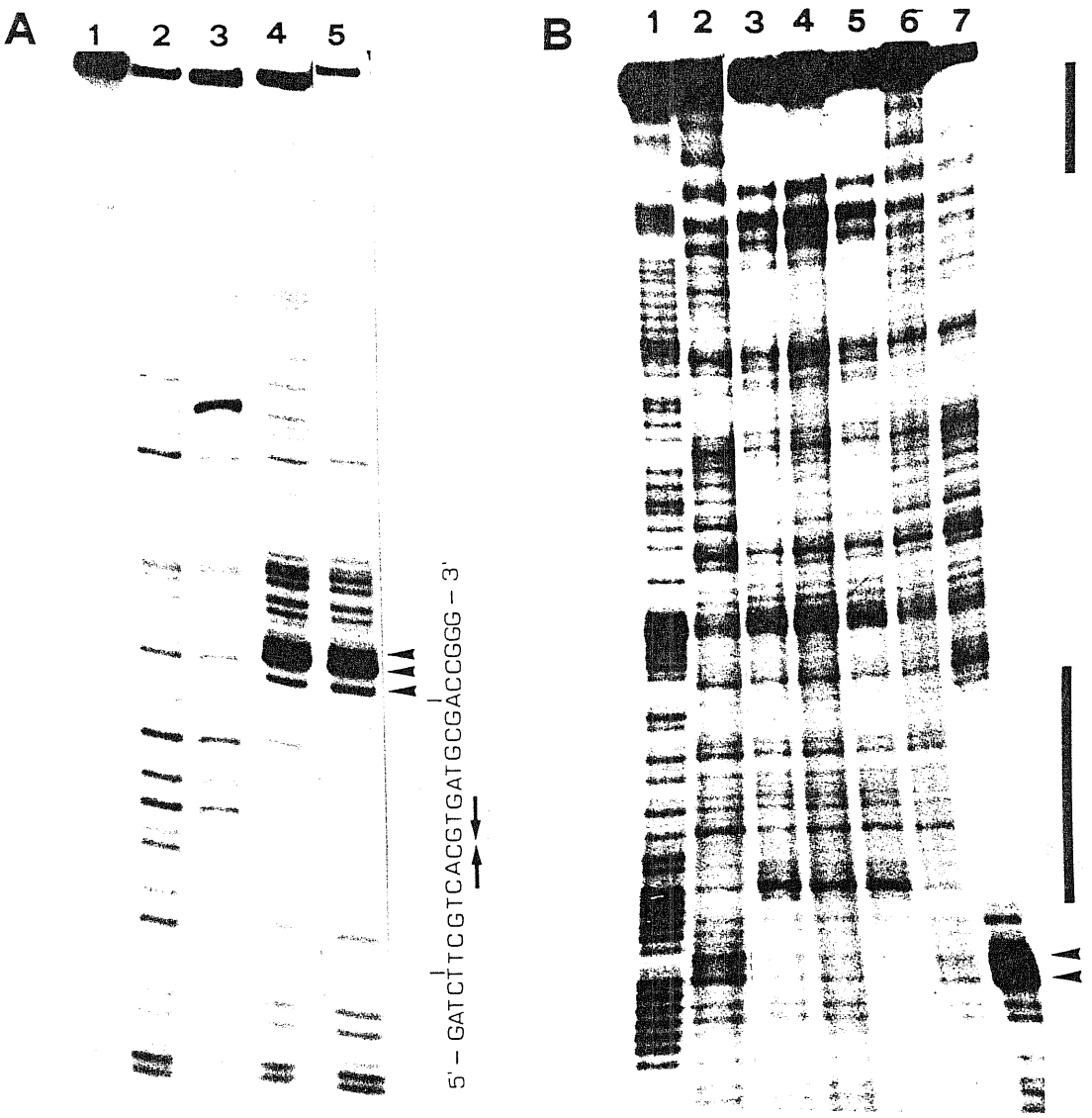


Figure 3.12 DNase I footprinting assay with affinity purified pool 1 and pool 3 proteins. The DNA probe, end-labelled at Hind-III site, was incubated with protein fractions and digested with DNase I for 1 min. The arrows indicate nuclease hypersensitive sites. The protected nucleotide sequence is indicated. (A) Lane 1, G+A reaction; lane 2, control (no protein); lane 3, 1st specific 0.3 M pool (5 µl); lanes 4 and 5, 1st specific 1 M pool (5 and 8 µl, respectively). (B) Lane 1, G+A reaction; lanes 2 and 6, control (no protein); lane 3 and 4, 2nd specific 1M pool (5 and 8 ml, respectively); lane 5, pool 3 (5 µl) and lane 7, rUSF (2 µl).

One of the characteristics of the Ku protein is that it selectively recognizes ends of double stranded DNA molecules *in vitro* (Mimori & Hardin, 1986). To study the DNA binding properties of pB48-2 DNase I footprinting and gel retardation experiments were performed. Using purified pB48-2 only the end region of the DNA fragment was found to be protected against DNase I, as it is shown in **Figure 3.12b**.

In a gel retardation experiment the pB48-2 protein was incubated with increasing amounts of the circular and linearized plasmid containing four copies of pB48 binding site (pUF4). Linearized pUF4 competed for the binding of pB48-2, while there was no competition with the circular plasmid (**Figure 3.13**).

A random sequence, a HindIII/HaeII 165 bp fragment of plasmid pUC18 was used as a probe in a gel retardation assay. The pB48-2 protein bound to this fragment and, similarly to the Ku antigen, formed multimer DNA-protein complexes (**Figure 3.14**).

These data suggest that pB48-2 and Ku are identical proteins or related members of the same family.

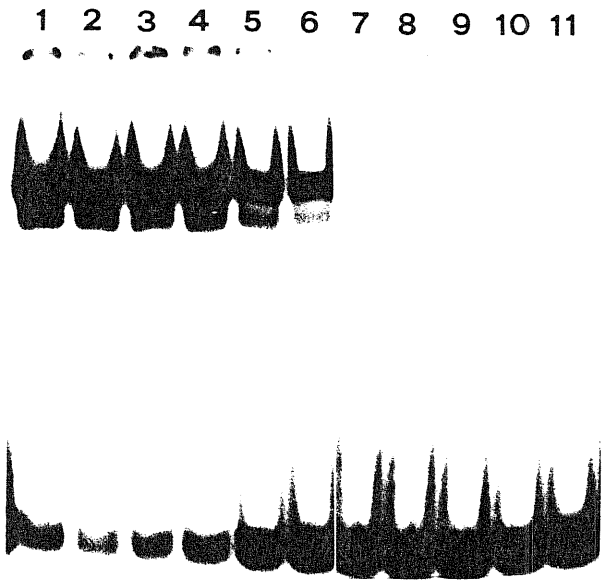


Figure 3.13 pB48-2 needs ends to bind to DNA. Approximately 40 ng of purified pB48-2 protein was incubated with increasing amounts of circular (75, 150, 750, 1500, 4500 fold weight excess, lanes 2-6, respectively) and linearized (75, 150, 750, 1500, 4500 fold weight excess, lanes 7-11, respectively) pUF4 plasmid. After 15 min 0.2 ng of labelled pB48 oligo was added in the reaction mixture and incubated for a further 30 min at room temperature.

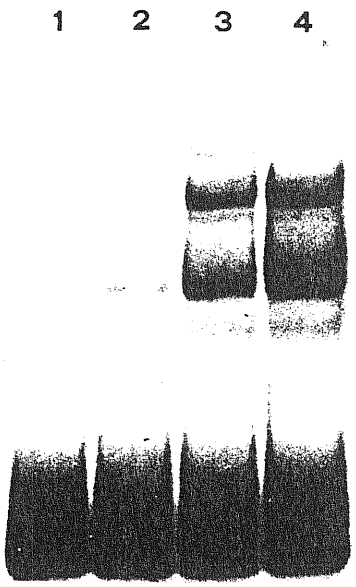


Figure 3.14 pB48-2 binds to pUC18/HindIII-HaeII fragment. Approximately 10 ng of end-labelled probe was incubated with 0, 20, 80 and 120 ng of pB48-2 (lanes 1, 2, 3 and 4, respectively) in the presence of 3 μg of poly[d(I-C)] for 30 min at room temperature.

4. DISCUSSION

4.1 TWO DISTINCT PROTEIN SPECIES BIND TO THE pB48 BINDING SITE

The evidence obtained by a newly developed quantitative PCR technique indicates that a chromosomal origin of DNA replication is located within approximately 3000 bp in a 13.7 kb region (L30E) of the human chromosome 19 (Biamonti et al., 1992b). The L30E region is replicated within the first minute of S-phase and contains two nonoverlapping units of transcription separated by a promoter. One of the genes encodes for lamin B2, a protein expressed mainly in S-phase. The origin region encompasses several elements characteristic of actively transcribed region including a binding site for the transcription factor USF/MLTF and an active promoter overlapping a putative CpG island (Falaschi et al., 1988; Giacca et al., 1989). The search for nuclear proteins that display the activities expected of factors involved in initiation of DNA replication resulted in the purification of two proteins that bind to a 17 bp sequence

(pB48 binding site) present in this early replicating DNA. Two proteins, designated pB48-1 and pB48-2 have been purified from HeLa cells to greater than 95% homogeneity with 4% yield after 83 000 fold purification (pB48-1) and with 30% yield after 1700 fold purification (pB48-2). The purification approach combined conventional chromatography on ion-exchange columns together with fast flow protein liquid chromatography which gave a high resolution. The final purification step, specific affinity chromatography resulted in preparations of high purity. Before starting the purification two distinct DNA-binding activities were detected, distinguishable by their different gel mobility and heat stability. The approximate molecular weight of these DNA-binding proteins was determined using Southwestern analysis. Following purification it was observed that the DNA-protein complex showing slower mobility in gel retardation assay corresponds to two polypeptides of 41.5 and 44 kDa (pB48-1), while two other polypeptides, of 72 and 87 kDa (pB48-2) form the faster migrating DNA-binding activity.

4.2 pB48-1 IS A MEMBER OF BASIC-HELIX-LOOP-HELIX FAMILY OF PROTEINS ?

Two polypeptides of 41.5 and 44 kDa consistently copurified with pB48 site-binding activity. Copurification of these proteins strongly suggests that they are both required

for binding to the DNA. Indeed, the Southwestern experiments show that both polypeptides interact directly with the binding site. The specificity of binding is highly dependant on the integrity of the CACGTG core sequence within the binding site because the mutation of two central nucleotides (CG=TA) hinders the binding activity to a large degree. It was previously reported that a pB48 binding site sequence is very similar to that of the upstream element in the major late promoter of adenovirus 2 (Tribioli et al., 1987). The center of the 17 bp protected region, the dyad symmetry element CACGTG has also been found in many other gene regulatory elements of different species. The binding activities recognizing similar sequences (consensus CACPuTG) have been identified and in many cases purified, like proteins USF (Sawadogo & Roeder, 1985, Sawadogo et al., 1988) and MLTF (Carthew et al., 1985, Chodosh et al., 1986), TFE3 (Beckman et al., 1990), TFEB (Carr & Sharp, 1990), Myc (Blackwell et al., 1990; Prendergast & Ziff, 1991) and Max (Blackwood & Eisenman, 1991). The CACPuTG binding proteins are members of a basic-helix-loop-helix (bHLH) family, mediating their specific interaction with DNA through a basic region adjacent to a helix-loop-helix domain (Murre et al., 1989). A recent mutational analysis of MyoD showed that the basic region is required for DNA-binding and muscle-specific gene activation, but not protein dimerization, which required only the helix-loop-helix motif (Davis et al., 1990) The leucine zipper has been found in several transcription factors and has been shown to mediate protein-protein interactions (O'Shea et

al., 1989). It is often found next to the basic region, but in most of the members of the bHLH family of proteins the leucine zipper is adjacent to the helix-loop-helix domain. The DNA-binding activity is dependent of both the dimerization domains and the basic region. Homo- and hetero-dimeric proteins have emerged as a common motif in the eukaryotic transcription field suggesting the possibility that different binding specificities can be generated by combining heterologous subunits. The function of these proteins maybe repressed or enhanced by specific interactions with other proteins via the leucine repeat or helix-loop-helix domain as has already been shown for the Myc and Max proteins (Blackwood & Eisenman, 1991).

The molecular weight range of pB48-1 and its heat stability characteristics were reminiscent of data on USF (Sawadogo et al., 1988). Furthermore, the specific band-shift activity of both proteins seems to depend on the length of the probe. Using a labelled tandem repeat of four pB48 binding sites, two equal pB48-1 specific shifts appear (results not shown), whereas using a 25 bp oligo containing only one binding site, the upper pB48-1 specific shift is of much stronger intensity. These facts, together with an immunological relationship which was established using an antibodies raised against rUSF 43 kDa component in gel retardation assay, suggest that Pb48-1 is highly related or maybe even identical to USF. The existence of multiple family members seems to be a common feature of many eukaryotic transcription factors. Further characterization, possibly amino acid sequence analysis, is

necessary to determine whether pB48-1 is simply another member of the family of USF-like proteins.

4.3 pB48-2 BEARS A CLOSE RESEMBLANCE TO THE AUTOANTIGEN Ku

The purified pB48-2 activity is composed of two subunits of 72 and 87 kDa which copurify with the pB48-site binding activity and are present in equal amounts during the final stage of purification, as visualized on SDS-PAGE gels. The various forms of chromatography failed to separate the two subunits while retaining activity. The results of the Southwestern hybridization experiments suggest that the 72 kDa component contributes more for the binding to DNA. The effect of competitor DNA, however, showed that the complex was nonspecific, as it was not possible to detect the preference of pB48-2 for the pB48 binding site. Only circular DNA molecule was not able to compete for the binding of pB48-2 suggesting that the protein requires the ends to bind. Nevertheless, in DNase I footprinting assays the protected region was located at the ends of the fragment with no apparent sequence specificity. When a larger, random DNA fragment was used as a probe, multimeric protein-DNA complexes were formed. It seems that the protein, upon binding to a DNA end, moves to the internal portions of the fragment and covers it completely at sufficiently high concentrations. The properties of the pB48-2 protein

resemble those of the autoantigen Ku, a protein that was originally identified as an autoantigen recognized by sera from patients with rheumatic diseases (Mimori et al., 1986; Yaneva & Busch 1986). The Ku antigen contains equal amounts of 70 and 80 kDa polypeptides which are products of different genes (Reeves and Sthoeger, 1989; Mimori et al., 1990). It forms heterodimers or tetramers in solution and is associated with active chromatin in a cell-cycle dependent manner. The Ku protein possibly constitutes an integral component of the chromatin DNA-protein complex that anchors DNA strands to the nuclear membrane region (Higashiura et al., 1992). It has been shown to bind specifically to end region of double-stranded DNA (Mimori & Hardin 1986). No particular nucleotide sequence appeared to be recognized preferentially. It also binds weakly to single-stranded DNA and not at all to RNA or RNA-DNA hybrids (Mimori and Hardin, 1986). However, some Ku related proteins have recently been shown to bind DNA in a sequence specific manner such as Ku-2, an octamer binding protein (May et al., 1991), PSE-1, which binds the proximal sequence element of the human U1 promoter and promotes transcription from a minimal promoter containing this element (Knuth et al.,1990) and TREF which binds the human transferrin receptor promoter and is also involved in transcription (Roberts et al., 1989). NFIV was identified as a protein that binds to the origin of replication of type 2 adenovirus (De Vries et al., 1989). It was shown to be a hetrodimer of 72 and 84 kDa which requires the presence of molecular ends and upon

binding moves freely without energy input on any double-stranded DNA forming regular DNA-multimeric protein complexes (De Vries et al., 1989).

The close similarity between pB48-2 and Ku is confirmed by their sequence homology. Complete sequencing of pB48-2 is required to reveal the full extend of its homology with Ku. Considering the data obtained, pB48-2 is a protein which has relatively high nonspecific affinity for DNA and thus there is a difficulty of using competitor DNA to examine its possible sequence-specific DNA-binding activity. A similar problem has been encountered in the study of sequence-specific DNA-binding by *Drosophila* P-element transposase (Kaufman et al., 1989). The Southwestern hybridization experiment data suggest that the high sequence nonspecific binding activity (under the gel retardation assay conditions) might mask the possible specificity of binding. *In vitro* DNA-binding sequence specificity may be demonstrated if more refined DNA binding techniques are used such as "selected and amplified binding sites" (SAAB) imprinting, described for the c-Myc protein (Blackwell et al., 1990)

4.4 POSSIBLE FUNCTIONS OF pB48-1 AND pB48-2 *IN VIVO*

According to the data available, transcription signals are always found near or inside origins of replication, although there is little information about the role of

regulators of transcription in DNA replication. The picture is more clear in eukaryotic viruses where factors have been identified that interact both with transcription and replication, like Sp1 (Kadonaga et al.,1986), T antigen (DePamphilis & Wasserman, 1982), NFIII/Oct-1 (Prujin et al., 1987; O'Neill et al., 1988) and CTF/NF1 (Jones et al., 1987, Santoro et al., 1988, Mermod et al., 1989). Binding of CTF/NF1 perturbs the local distribution of nucleosomes, thus increasing the accessibility of the origin region to other regulatory factors. One may conclude that there could also be a role for cis-acting transcriptional elements in regulation of eukaryotic DNA replication. The data we obtained by now strongly indicate that the L30E region contains an origin of replication. Thus it is conceivable that proteins of the bHLH family (whose member is most probably pB48-1) could participate in origin activation, as well as in transcription.

There is a sequence similarity between both subunits of Ku and Myc gene family products. The corresponding regions contain a periodic repeat of leucines and serines, thus forming a leucine zipper (Reeves and Sthoeger, 1989; Mimori et al., 1990). The presence of an acidic domain within the 70 kDa component sequence indicates that Ku might be a transcriptional activator. Indeed, Ku-like proteins, PSE1 and TREF have been shown to be involved in the transcriptional regulation (Knuth et al.,1990; Roberts et al., 1989). The affinity of pB48-2 for DNA ends and the similarity with Ku indicate that it may be involved in the processes that require DNA termini, like DNA replication and repair. It may also

recognize a specific DNA structure (a wide distribution of Ku on the chromosomal DNA favors this possibility), like the nonhistone high mobility group 1 protein (HGM1), which was shown to bind four-way junction structure with high specificity, but independently of base sequence (Bianchi et al., 1989). Palindromic DNA sequences (inverted repeats) are common in both prokaryotic and eukaryotic genomes. They can, in theory, undergo intrastrand base pairing and form cruciforms or four-way junctions. Indirect evidence is accumulating that this may occur *in vivo* and may be of physiological significance. A palindrom is present in the pB48 region probably very close to the activated origin. The Ku antigen is also able to bind to hairpins (Paillard & Straus, 1991) that are probably formed transiently according to the modulation of expression of the specific portions of the genomes. Hence, the Ku and Ku-like proteins, interacting with other, possibly more sequence specific proteins, could play an important role in eukaryotic gene expression and origin activation.

5. BIBLIOGRAPHY

Aggarwal A.K, Rodgers D, Drottar M, Ptashne M & Harrison SC (1988): Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. Science 242:899-907

Akam M (1989): Hox and HOM: homologous gene clusters in insects and vertebrates. Cell 57:347-349

Anachkova B & Hamlin JL (1989): Replication in the amplified dihydrofolate reductase domain in CHO cells may initiate at two distinct sites, one of which is a repetitive sequence element. Mol Cell Biol 9:532-540

Anderson JE, Ptashne M & Harrison SC (1987): Structure of the repressor-operator complex of bacteriophage 434. Nature 326:846-852

Beckmann H, Su L-K & Kadesch T (1990): TFE3: a helix-loop-helix protein that activates transcription through the immunoglobulin enhancer μE3 motif. Genes Dev 4:167-179

Bell SP & Stillman B (1992): ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature 357:128-134

Berg OG & von Hippel PH (1988): Selection of DNA binding sites by regulatory proteins. Trends Biochem Sci 13:207-211

Berg JM (1988): Proposed structure for the zinc-binding domains from transcription factor III A and related proteins. Proc Natl Acad Sci USA 85:99-102

Bergemann AD & Johnson EM (1992): The HeLa Pur factor binds single-stranded DNA at a specific element conserved in gene flanking regions and origins of DNA replication. Mol Cell Biol 12:1257-1265

Biamonti G, Della Valle G, Talarico D, Cobianchi F, Riva S & Falaschi A (1985): Fate of exogenous recombinant plasmids introduced into mouse and human cells. Nucleic Acids Res 13:5545-5561

Biamonti G, Giacca M, Perini G, Contreas G, Zentilin L, Weighardt F, Della Valle G, Saccone S, Riva S & Falaschi A (1992a): The gene for novel human lamin maps at a highly transcribed locus of chromosome 19 which replicates at the onset of S-phase. Mol Cell Biol 12:3499-3506

Biamonti G, Perini G, Weighardt F, Riva S, Giacca M, Norio P, Zentilin L, Diviacco S, Dimitrova D & Falaschi A (1992b): A human DNA replication origin: localization and transcriptional characterization. Chromosoma 102:24-31

Bianchi ME, Beltrame M & Paonessa G (1989): Specific recognition of cruciform DNA by nuclear protein HGM1. Science 243:1056-1059

Bienz M & Pelham HRB (1986): Heat shock regulatory elements function as an inducible enhancer in the Xenopus hsp 70 gene and when linked to a heterologous promoter. Cell 45:753-760

Birkenmeier EH, Gwynn B, Howard S, Jerry J, Gordon JI, Landschulz WH & McKnight SL (1989): Tissue-specific expression, developmental regulation and genetic mapping of the gene encoding CCAAT/enhancer binding protein. Genes Dev 3:1146-1156

Blackwell TK, Kretzner L, Blackwood EM, Eisenman RN & Weintraub H (1990): Sequence-specific DNA binding by the c-Myc protein. Science 250:1149-1151

Blackwell TK & Weintraub H (1990): Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. Science 250:1104-1110

Blackwood EM & Eisenman RN (1991): Max: A helix-loop-helix zipper protein that forms a sequence-specific DNA binding complex with Myc. Science 251:1211-1217

Blow J (1992): A protein complex present at origins of DNA replication in yeast cells. BioEsseys 14:561-563

Bramhill D & Kornberg A (1988): A model for initiation at origins of DNA replication. Cell 54:915-918

Breathnach R & Chambon P (1981): Organization and expression of eukaryotic split genes coding for proteins. Annu Rev Biochem 50:349-383

Brennan RG & Matthews BW (1989a): The helix-turn-helix DNA binding motif. J Biol Chem 264:1903-1906

Brennan RG & Matthews B. (1989b): Structural basis of DNA-protein recognition. Trends Biochem Sci 14:286-290

Brennan RG, Roderick SL, Takeda Y & Matthews BW (1990): Protein-DNA conformational changes in the crystal structure of a l Cro-operator complex. Proc Natl Acad Sci USA 87:8165-8169

Brewer BJ & Fangman WL (1987): The localization of replication origins on ARS plasmids in S. cerevisiae. Cell 51:463-471

Broach JR, Li Y-Y, Feldman J, Jayaram M, Abraham J, Nasmyth KA & Hicks JB (1982): Localization and sequence analysis of yeast origins of DNA replication. Cold Spring Harbor Symp Quant Biol 47:1165-1173

Buratowski S, Hahn S, Guarente L & Sharp PA (1989): Five intermediate complexes in transcription initiation by RNA polymerase II. Cell 56:549-561

Buratowski S, Sopta M, Greenblatt J & Sharp PA (1991): RNA polymerase II associated proteins are required for a DNA conformation change in the transcription initiation complex. Proc Natl Acad Sci USA 88:7509-7513

Burhans WC, Selegue JE & Heintz NH (1986a): Isolation of the origin of replication associated with the amplified Chinese hamster dihydrofolate reductase domain. Proc Natl Acad Sci USA 83:7790-7794

Burhans WC, Selegue JE & Heintz NH (1986b): Replication intermediates formed during initiation of DNA synthesis in Methotrexate-resistant CHOC 400 cells are enriched for sequences derived from a specific, amplified restriction fragment. Biochemistry 25:441-449

Burhans WC, Vassilev LT, Caddle MS, Heintz NH & DePamphilis ML (1990): Identification of an origin of bidirectional DNA replication in mammalian chromosomes. Cell 62:955-965

Caddle MS, Lussier RH & Heintz NH (1990a): Intramolecular DNA triplexes, bent DNA and DNA unwinding elements in the initiation region of an amplified dihydrofolate reductase replicon. J Mol Biol 211:19-33

Caddle MS, Dailey L & Heintz NH (1990b): RIP60, a mammalian origin-binding protein, enhances DNA bending near the dihydrofolate reductase origin of replication. Mol Cell Biol 10:6236-6243

Campbell J (1988): Eukaryotic DNA replication: yeast bares its ARSs. Trends Biochem Sci 13:212-217

Carr CS & Sharp PA (1990): A helix-loop-helix protein related to the immunoglobulin E-box proteins. Mol Cell Biol 10:4384-4388

Carthew RW, Chodosh LA & Sharp (1985): An RNA polymerase II transcription factor binds to an upstream element in the adenovirus major late promoter. Cell 43:439-448

--------- (1987): The major late transcription factor binds to and activates the mouse metallothionein I promoter. Genes Dev 1:973-980

Chodosh LA, Carthew RW & Sharp PA (1986): A single polypeptide possesses the binding and transcription activities of the adenovirus major late transcription factor. Mol Cell Biol 6:4723-4733

Chodosh LA, Carthew RW, Morgan JG, Crabtree GR & Sharp PA (1987): The adenovirus major late transcription factor activates the rat g-fibrinogen promoter. Science 238:684-688

Chodosh LA, Oelsen J, Hahn S, Baldwin AS, Guarente L & Sharp PA (1988): A yeast and a human CCAAT-binding protein have heterologous subunits that are functionally interchangeable. Cell 53:25-35

Clerc RG, Corcoran LM, Le Bowitz JH, Baltimore D & Sharp PA (1988): The B-cell specific Oct-2 protein contains POU-box and homeobox-type domains. Genes Dev. 2:1570-1581

Comai L, Tanese N & Tjian R (1992): The TATA-binding protein and associated factors are integral components of the RNA polymerase I transcription factor, SL1. Cell 68:965-976

Cormack BP & Struhl K (1992): The TATA-binding protein is required for transcription by all three nuclear RNA polymerases in yeast cells. Cell 69:685-696

Dailey L, Caddle MS, Heintz N & Heintz NH (1990): Purification of RIP60 and RIP100, mammalian proteins with origin-specific DNA-binding and ATP-dependent DNA helicase activities. Mol Cell Biol 10:6225-6235

Davis RL, Cheng P-F, Lassar AB & Weintraub H (1990): The MyoD DNA binding protein contains a recognition code for muscle-specific gene activation. Cell 60:733-746

Della Seta F, Ciafre S-A, Marck C, Santoro B, Presutti C, Sentenac A & Bozzoni I (1990): The ABF1 factor is the transcriptional activator of the L2 ribosomal protein genes in Saccharomyces cerevisiae. Mol Cell Biol 10:2437-2441

DePamphilis ML & Wasserman PM (1982): Organization and replication of papova virus DNA. CRS Press, Boca Raton, FL, pp 37-114

DePamphilis ML (1988): Transcriptional elements as components of eukaryotic origins of DNA replication. Cell 52:635-638

De Vries E, Van Driel W, Bergsma WG, Arnberg AC & Van der Vliet P (1989): HeLa nuclear protein recognizing DNA termini and translocating on DNA forming a regular DNA-multimeric protein complex. J Mol Biol 208:65-78

Diffley JFX & Stillman B (1988): Purification of a yeast protein that binds to origins of DNA replication and a transcriptional silencer. Proc Natl Acad Sci USA 85:2120-2124

Diffley JFX & Cocker JH (1992): Protein-DNA interactions at a yeast replication origin. Nature 357:169-172

Dignam JD, Lebovitz RM & Roeder RG (1983): Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res 11:1475-1491

Doniger J, Landsman D, Gonda MA & Wistow G (1992): The product of unr, the highly conserved gene upstream of N-ras, contains multiple repeats similar to the cold-shock domain (CSD), a putative DNA-binding motif. New Biol 4:389-395

Dynan WS & Tjian R (1983): The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV 40 early promoter. Cell 35:79-87

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

Eckdahl TT & Anderson JN (1990): Conserved DNA structures in origins of replication. Nucleic Acids Res 18:1609-1612

Evans RM (1988): The steroid and thyroid hormone receptor gene superfamily. Science 240:889-895

Evans RM & Hollenberg (1988): Zinc fingers: guilt by association. Cell 52:1-3

Falaschi A, Biamonti G, Cobianchi F, Csordas-Toth E, Faulkner G, Giacca M, Pedacchia D, Perini G, Riva S & Tribioli C (1988): Presence of transcription signals in two putative DNA replication origins of human cells. Biochim Biophys Acta 951:430-442

Ferguson BM & Fangman WL (1992): A position effect on the time of replication origin activation in yeast. Cell 68:333-339

Fletcher C, Heintz N & Roeder RG (1987): Purification and characterization of OTF-1, a transcription factor regulating cell cycle expression of a human histone H2b gene. Cell 51:773-781

Freemont PS, Lane AN & Sanderson MR (1991a): Structural aspects of protein-DNA recognition. Biochem J 278:1-23

Freemont PS, Hanson IM & Trowsdale J (1991b): A novel cysteine-rich sequence motif. Cell 64:483-484

Gehring WJ (1987): Homeo boxes in the study of development. Science 236:1245-1250

Gentz R, Rausher FJ, Abate C & Curran T (1989): Parallel association of Fos and Jun leucine zippers juxtaposes DNA binding domains. Science 243:1695-1699

Giacca M, Gutierrez MI, Demarchi F, Diviacco S, Biamonti G, Riva S & Falaschi A (1989): A protein target site in an early replicated human DNA: a highly conserved sequence. Bichem Biophys Res Comm 165:956-965

Giacca M, Gutierrez MI, Menzo S, D'Adda di Fagagna F & Falaschi A (1992): A human binding site for transcription factor USF/MLTF mimics the negative regulatory element of human immunodeficiency virus type 1. Virology 186:133-147

Gidoni D, Kadonaga JT, Barrera-Saldana H, Takahashi K, Chambon P & Tjian R (1985): Bidirectional SV40 transcription mediated by tandem Sp1 binding interactions. Science 230:511-517

Green S & Chambon P (1987): Oestradiol induction of a glucocorticoid response gene by a chimaeric receptor. Nature 325:75-78

Greenblatt J (1991a): RNA polymerase-associated transcription factors. Trends Biochem Sci 16:408-411

Greenblatt J (1991b): Roles of TFIID in transcroptional initiation by RNA polymerase II. Cell 66:1067-1070

Gregor PD, Sawadogo M & Roeder RG (1990): The adenovirus major late transcription factor USF is a member of the helix-

loop-helix group of regulatory proteins and binds to DNA as a dimer. Genes Dev 4:1730-1740

Hahn s, Buratowski S, Sharp PA & Guarente L (1989): Isolation of the gene encoding the yeast TATA binding protein TFIID: a gene identical to the SPT15 suppressor of Ty element insertions. Cell 58:1173-1181

Handeli S, Klar A, Meuth M & Cedar H (1989): Mapping replication units in animal cells. Cell 57:909-920

Hard T, Kellenbach E, Boelens R, Maler BA, Dahlman K, Freedman LF, Carlstedt-Duke J, Yumamoto KR, Gustafsson JA & Kapstein R (1990): Structure of glucocorticoidreceptor domain. Science 249:157-160

Harland RM (1981): Initiation of DNA replication in eukaryotic chromosomes. Trends Biochem Sci 6:71-74

Harland RM & Laskey RA (1980): Regulated replication of DNA microinjected into eggs of X. laevis. Cell 21:761-771

Harrison SC & Aggarwal AK (1990): DNA recognition by proteins with the helix-turn-helix motif. Annu Rev Biochem 59:933-969

Hatton KS, Dhar V, Brown EH, Iqbal MA, Stuart S, Didamo VT & Schildkraut CL (1988): Replication program of active and inactive multigene families in mammalian cells. Mol Cell Biol 8:2149-2158

Heintz NH & Hamlin J (1982): An amplified chromosomal sequence that includes the gene for dihydrofolate reductase initiates replication within specific restriction fragments. Proc Natl Acad Sci USA 79:4083-4087

Herr W, Sturm RA, Clerc RG, Corcoran LM, Baltimore D, Sharp PA, Ingraham HA, Rosenfeld MG, FInney M, Ruvkin G & Horvitz HR (1988): The Pou domain: a large conserved region in the mammalian pit-1, oct-1, oct-2 and C. Elegans unc-86 gene products. Genes Dev 2:1513-1516

Higashiura M, Shimizu Y, Tanimoto M, Morite T & Yagura T (1992): Immunolocalization of Ku-proteins (p80/p70): localization of p70 to nucleoli and periphery of both

interphase nuclei and methaphase cheomosomes. Exp Cell Res 201:444-451

Hoffmann A, Horikoshi M, Wang CK, Schroeder S, Weil PA & Roeder RG (1990): Cloning of the Schizosaccharomyces pombe TFIID gene reveals a strong conservation of functional domains present in saccharomyces cerevisiae TFIID. Genes Dev 4:1141-1148

Huberman JA, Spotila LD, Nawotka KA, El-Assouli SM & Davis LR (1987): The in vivo replication origin of the yeast 2 µm plasmid. Cell 51:473-481

Iguchi-Ariga SMM, Okazaki T, Itani T, Ogata M, Sato Y & Ariga H (1988): An initiation site of DNA replication with transcriptional enhancer activity present upstream of the c-myc gene. EMBO J 7:3135-3142

Ingham PW (1988): The molecular genetics of embryonic pattern formation in Drosophila. Nature 335:25-34

Ingraham HA, Chen R, Mangalan HJ, Elsholtz HP, Flynn SC, Linn CR, Simmons DM, Swanson, L. & Rosenfeld, M.G. (1988): A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype. Cell 55:519-529

Ingraham HA, Flynn SE, Voss JW, Albert VR, Kapiloff MS, Wilson L & Rosenfeld MG (1990): The POU specific domain is essential for sequence-specific high affinity DNA binding and DNA-dependent Pit1-Pit1 interactions. Cell:1021-1033

Inostroza J, Flores O & Reinberg D (1991): Factors involved in specific transcription by mammalian RNA polymerase II: purification and functional analysis of general transcription factor IIE. J Biol Chem 266:9304-9308

Jackson DA (1990): The organization of replication centres in higher eukaryotes. BioEssays 12:87-89

Johnson PF, Landschulz WH, Graves BJ & McKnight SL (1987): Identification of a rat liver nuclear protein that binds to the enhancer core element of three animal viruses. Genes Dev 1:133146

Jones KA, Yamamoto KR & Tjian R (1985): Two distinct transcription factors bind to the HSV thymidine kinase promoter in vitro. Cell 42:559-572

Jones KA, Kadonaga JT, Rosenfeld PJ, Kelly TJ & Tjian R (1987): A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. Cell 48:79-89

Jones N (1990): Transcriptional regulation by dimerization: two sides to an incestuous relationship. Cell 61:9-11

Jordan SR & Pabo CO (1988): Structure of the lambda complex at 2.4 A resolution: Details of the repressor-operator interactions. Science 242:893-899

Kadonaga JT & Tjian R (1986): Affinity purification of sequence specific DNA binding proteins. Proc Natl Acad Sci USA 83:5889-5893

Kadonaga JT, Jones KA & Tjian R (1986): Promoter-specific activation of RNA polymerase II transcription by Sp1. Trends Biochem Sci 11:20-23

Kadonaga JT, Carner KR, Masiarz FR & Tjian R (1987): Isolation of cDNA encoding the transcription factor Sp1 and functional analysis of the DNA binding domain. Cell 51:1079-1090

Kao CC, Lieberman PM, Schmidt MC, Zhou Q, Pei R & Berk AJ (1990): Cloning of a transcriptionally active human TATA binding factor. Science 248:1646-1650

Karim FD, Urness LD, Thummel CS, Klemsz MJ, McKercher SR, Celeda A, Van Beveren C, Maki RA, Gunther CV, Nye JA & Graves BJ (1990): The ETS-domain: a new DNA binding motif that recognizes a purine rich core DNA sequence. Genes Dev 4:1451-1453

Kaufmann G, Zannis-Hadjopoulos M & Martin R (1985): Cloning of nascent monkey DNA synthesized early in the cell cycle. Mol Cell Biol 5:721-727

Kaufman PD, Doll RF & Rio DC (1989): Drosophila P element Transposase recognizes internal P element DNA sequences. Cell 59:359-371

Kissinger CR, Liu B, Martin-Blanco E, Kornberg TB & Pabo CO (1990): Crystal structure of an engrailed homeodomain-DNA complex at 2.8 A resolution: A framework for understanding homeodomain-DNA interaction. Cell 63:579-590

Klug A & Rhodes D (1987): Zinc fingers: a novel protein model for nucleic acids recognition. Trends Biochem Sci 12:464-469

Knuth MW, Gunderson SI, Thompson NE, Strasheim LA & Bergess RR (1990): Purification and characterization of proximal sequence-binding protein 1, a transcription activating protein related to Ku and TREF that binds the proximal sequence element of the human U1 promoter. J Biol Chem 265:17911-17920

Kornberg A & Baker TA (1992): DNA replication. WH Freeman & Company, New York

Koshland D, Kent JC & Hartwell LH (1985): Genetic analysis of the mitotic transmission of minichromosomes. Cell 40:393-403

Kouzarides T & Ziff E (1988): The role of leucine zipper in the Fos-Jun interaction. Nature 336:646-651

Kristie TM & Sharp PA (1990): Interaction of the Oct-1 POU subdomains with specific DNA sequences and with the HSV alpha transactivator protein. Genes Dev 4:2383-2396

Landschulz WH, Johnson PF, Adashi EY, Graves BJ & McKnight SL (1988a): Isolation of a recombinant copy of the gene encoding C/EBP. Genes Dev 2:786-800

Landschulz WH, Johnson PF & McKnight SL (1988b): The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science 240:1759-1764

Latchman D S (1990): Eukaryotic transcription factors. Biochem J. 270:281-289

Latchman DS (1991): Eukaryotic transcription factors. Academic press, London, pp 1-15

Laughon A & Scott MP (1984): Sequence of a Drosophila segmentation gene: protein structure homology with DNA-binding proteins. Nature 310:25-31

Leu T-H & Hamlin J (1989): High resolution mapping of replication fork movement through the amplified dihydrofolate reductase domain in CHO cells by in-gel renaturation analysis. Mol Cell Biol 9:523-531

Lewin B (1990): Genes IV. Oxford University Press, Oxford, pp 543-577

Linskens MHK & Huberman JA (1990): The two faces of higher eukaryotic DNA replication origins. Cell 62:845-847

Luscher B & Eisenman RN (1990): New light on Myc and Myb. Part I. Myc. Genes Dev 4: 2025-2035

Maniatis T, Fritsch EF & Sambrook J (1982): Molecular cloning: a laboratory manual. Cold Spring Harbour University Press, Cold Spring Harbour.

Marahrens Y & Stillman B (1992): A yeast chromosomal origin of DNA replication defined by multiple functional elements. Science 255:817-823

Matthews BW (1988): No code for recognition. Nature 335:294-295

May G, Sutton C & Gould H (1991): Purification and characterizationof Ku-2, an octamer binding protein related to autoantigen Ku. J Biol Chem 266:3052-3059

McKnight SL, Kingsbury RC, Spence A & Smith M (1984): The distal transcription signals of the herpes virus tk gene share a common hexanucleotide control sequence. Cell 37:253-262

McWhinney C & Leffak M (1990): Autonomous replication of a DNA fragment containing the chromosomal replication origin of the human c-myc gene. Nucl Acids Res 18:1233-1242

Mermod N, O'Neil EA, Kelly TJ & Tjian R (1989): The proline rich transcriptional activator CTF/NF-1 is distinct from the replication and DNA binding domain. Cell 58:741-753

Mihara H & Kaiser ET (1988): A chemically synthesized Antennapedia homeodomain binds to a specific DNA sequence. Science 242:925-927

Miller J, McLachlan AD & Klug A (1985): Repetitive zinc-binding domains in the protein transcription factor III A from Xenopus oocytes. EMBO J 4:1609-1614

Mimori T, Hardin JA & Steitz JA (1986): Characterization of the DNA-binding protein antigen Ku recognized by autoantibodies from patients with rheumatic disorders. J Biol Chem 261:2274-2278

Mimori T & Hardin JA (1986): Mechanism of interaction between Ku protein and DNA. J Biol Chem 261:10375-10379

Mimori T, Ohosone Y, Hama N, Suwa A, Akizuki M, Homma M, Griffith AJ & Hardin JA (1990): Isolation and characterization of cDNA encoding the 80-kDa subunit protein of the human autoantigen Ku (p70/p80) recognized by autoantibodies from patients with scleroderma-polymiositis overlap syndrome. Proc Natl Acad Sci USA 87:1777-1781

Miyamoto NG, Moncollin V, Egly JM & Chambon P. (1985): Specific interaction between a transcription factor and the upstream element of the adenovirus-2 major late promoter. EMBO J 4:3563-3570

Muller M, Afrolter M, Leupin W, Otting G, Wuthrich K & Gehring VJ (1988a): Isolation and sequence specific DNA binding of the Antennapedia homeodomain. EMBO J 7:4299-4304

Muller MM, Rupert S, Schaffner W & Matthias P (1988b): A cloned octamer transcription factor stimulates transcription from lymphoid-specific promoters in non-B cells. Nature 336:544-551

Murre C, McCaw PS & Baltimore D (1989a): A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD and myc proteins. Cell 56:777-783

Murre C, McCaw PS, Vaessin H, Caudy M, Jan LY, Jan YN, Cabera CV, Buskin JN, Hauschka SD, Lassar AB, Weintraub H & Baltimore D (1989b): Interactions between heterologous

helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell 58:537-544

Myers RM, Tilly K & Maniatis Y (1986): Fine structure genetic analysis of a β-globin promoter. Science 232:613-618

Nagata K, Guggenheimer RA, Enomoto T, Lichy JH & Hurwitz J (1982): Adenovirus DNA replication in vitro: identification of a host factor that stimulates synthesis in the preterminal protein-dCMP complex. Proc Natl Acad Sci USA 79:64386442

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

Nardelli J, Gibson TJ, Vesque C & Charnay P (1991): Base sequence discrimination by zinc finger DNA binding domains. Nature 349:175-178

Ohlendorf DH & Matthews BW (1983): Structural studies of protein- nucleic acid interaction. Ann Rev Biophys Bioeng 12:259-284

Ohkuma Y, Sumimoto H, Horikoshi M & Roeder RG (1990): Factors involved in specific transcription by mammalian polymerase II: purification and characterization of general transcription factor TFIIE. Proc Natl Acad Sci USA 87:9163-9167

O'Neill EA, Fletcher C, Burrow CR, Heintz N, Roeder RG & Kelly TJ (1988): Transcription factor otf-1 is functionally identical to the DNA replication factor NF-III. Science 241:1210-1213

O'Shea EK, Rutkowski R & Kim ps (1989): Evidence that the leucine zipper is a coiled coil. Science 243:538-542

Otwinowski Z, Schevitz RW, Zhan R-G, Lawson CL, Joachimiak A, Marmorstein RQ, Luisi BF & Sigler PB (1988): Crystal structure of trp repressor/operator complex at atomic resolution. Nature 335:321-329

Pabo CO & Sauer RT (1984): Protein-DNA recognition. Ann Rev Biochem 53:293-321

Paillard S & Strauss F (1991): Analysis of the mechanism of interaction of simian Ku protein with DNA. Nucleic Acids Res 19:5619-5624

Parraga G, Horvath SJ, Eisen A, Taylor WE, Hood L, Young ET & Klevitt RE (1988): Zinc dependent structure of a single finger domain of yeast ADRI. Science 241:1489-1492

Parvin JD, Timmers HTM & Sharp PA (1992): Promoter specificity of basal transcription factors. Cell 68:1135-1144

Pavletich N & Pabo CO (1991): Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science 252:809-816

Peritz LN, Fodor EJB, Silversides DW, Cattini PA, Baxter JD & Eberhardt NL (1988): The human growth hormone gene contains both positive and negative control elements. J Biol Chem 263:5005-5007

Peterson CL & Calame K (1989): Proteins binding to site C2 (μE3) in the immunoglobulin heavy-chain enhancer exist in multiple oligomeric forms. Mol Cell Biol 9:776-786

Peterson MG, Tanese N, Pugh BF & Tjian R (1990): Functional domains and upstream activation properties of cloned human TATA binding protein. Science 248:1625-1630

Prendergast GC & Ziff EB (1989): DNA-binding motif. Nature 341:342

Prendergast GC & Ziff EB (1991): Methylation-sensitive sequence-specific DNA binding by the c-Myc basic region. Science 251:186-188

Prujin GJM, Van Driel W, Van Miltenberg RT & Van der Vliet PC (1987): Promoter and enhancer elements containing a conserved sequence motif are recognized by nuclear factor III, a protein stimulating adenovirus DNA replication. EMBO J 6:3771-3778

Pugh BF & Tjian R (1990): Mechanism of transcriptional activation by SP1: evidence for coactivators. Cell 61:1187-1197

Pugh BF & Tjian R (1991): Transcription from a TATA-less promoter requires a multisubunit TFIID complex. Genes Dev 5:1935-1945

Quian YQ, Billeter M, Otting G, Muller M, Gehring WJ & Wuthrich K (1989): The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. Cell 59:573-580

Ramstein J & Lavery R (1988): Energetic coupling between DNA bending and base pair opening. Proc Natl Acad Sci USA 85:7231-7235

Raymondjean M, Cereghini S & Yaniv M (1988): Several distinct CCAAT box binding proteins coexist in eukaryotic cells. Proc Natl Acad Sci USA 85:757-761

Redemann N, Gaul U & Jackle H (1988): Disruption of a putative Cys-zinc interaction eliminates the biological activity of the Kruppel finger protein. Nature 332:90-92

Reeves WH & Sthoeger ZM (1989): Molecular cloning of cDNA encoding the p70 (Ku) lupus autoantigen. J Biol Chem 264:5047-5052

Reinberg D, Horikoshi M & Roeder RG (1987): Factors involved in specific transcription by mammalian RNA polymerase II. J biol Chem 262:3322-3330

Roberts Mr, Miskimins WK & Ruddle FH (1989): Cell Regulation 1:151-164

Roeder RG (1991): The complexities of eukaryotic transcription initiation: regulation of preinitiation complex assembly. Trends Biochem Sci 16:402-408

Rosenfeld PJ, O'Neill EA, Wides RJ & Kelly TJ (1989): Sequence-specific interactions between cellular DNA-binding proteins and the adenovirus origin of DNA replication. Mol Cell Biol 7:875-886

Roy AL, Meisterernst M, Pognonec P & Roeder RG (1991): Cooperative interaction of an initator-binding transcription initation factor and the helix-loop-helix activator USF. Nature 354:245-248

Ruvkun G & Finney M (1991): Regulation of transcription and cell identity by POU domain proteins. Cell 64:475-478

Santoro C, Mermod N, Andrews PC & Tjian R (1988): A family of human CCAAT-box-binding proteins active in transcription and DNA replication: cloning and expression of multiple cDNAs. Nature 334:218-224

Sawadogo M & Roeder RG (1985): Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. Cell 43:165-175

Sawadogo M, Van Dyke MW, Gregor PD & Roeder RG (1988): Multiple forms of the human gene-specific transcription factor USF. J Biol Chem 263:11985-11993

Sawadogo M & Sentenac A (1990): RNA polymerase B (II) and general transcription factors. Annu Rev Biochem 59:711-754

Schaffner W (1989): How do different transcription factors binding the same DNA sequence sort out their jobs. Trends Genet 5:37-39

Scheidereit C, Heguy A & Roeder RG (1987): Identification and purification of a human lymphoid-specific octamer-binding protein (OTF-2) that activates transcription of an immunoglobulin promoter in vitro. Cell 51:783-793

Schvabe JWR, Neuhaus D & Rhodes D (1990): Solution structure of the DNA binding domain of the oestrogen receptor. Nature 348:458-461

Scott MP & Carroll SB (1987): The segmentation and homeotic gene network in early Drosophila development. Cell 51:689-698

Scott MP, Tamkun JW & Hartzell GW (1989): The structure and function of the homeodomain. Biochim Biophys Acta 989:25-48

Sekimizu K, Bramhill D & Kornberg A (1987): ATP activates dnaA protein in initiating replication of plasmids bearing the origin of the E. coli chromosome. Cell 50:259-265

Sen R & Baltimore D (1986): Multiple nuclear factors interact with the immunoglobulin enhancer sequences. Cell 46:705-716

Sharp PA (1992): TATA-binding protein is a classless factor. Cell 68:819-821

Shepherd JCW, McGinnis W, Carrasco AE. De Robertis EM & Gehring WJ (1984): Fly and frog homeodomains show homologies with yeast mating type regulatory loci. Nature 310:70-71

Silva CM, Tully DB, Petch LA, Jewell CM & Cidlowski JA (1987): Application of a protein-blotting procedure to the study of human glucocorticoid receptor interactions with DNA. Proc Natl Acad Sci USA 84:1744-1748

Sirito M, Walker S, Lin Q, Kozlowski MT, Klein WH & Sawadogo M (1992): Members of the USF family of helix-loop-helix proteins bind DNA as homo- as well as heterodimers. Gene Expression 2:231-240

Sive HL & Roeder RG (1986): Interaction of a common factor with conserved promoter and enhancer sequences in histone H2B, immunoglobulin and U2 small nuclear (sn) RNA genes. Proc Natl Acad Sci USA 83:6382-6386

Smale ST & Baltimore D (1989): The initiator as a transcription control element. Cell 57:103-113

Staacke D, Walter B, Kisters-Woike B, Wicken-Bergman B & Mueller-Hill B (1990): How trp repressor binds to its operator. EMBO J 9:1963-1967

Staudt LH, Singh H, Sen R, Wirth T, Sharp PA & Baltimore D (1986): A lymphoid -specific protein binding to the octamer motif of immunoglobulin genes. Nature 323:640-643

Stillman B (1989): Initiation of eukaryotic DNA replication in vitro. Annu Rev Cell Biol 5:197-245

Struhl K, Stinchcomb DT, Scherer S & Davis RW (1979): High-frequency transformation of yeast: autonomous replication of hybrid DNA molecules. Proc Natl Acad Sci USA 76:1035-1039

Struhl K (1989): Helix-turn-helix,zinc-finger and leucine-zipper motifs for eukaryotic transcriptional regulatory proteins. Trends Biochem Sci 14:137-140

Sturm RA, Das G & Herr W (1988): The ubiquitous octamer binding protein Oct-1 contains a POU domain with a homeobox. Genes Dev. 2:1582-1599

Tafuri SR & Wolfe AP (1990): Xenopus Y-box transcription factors: Molecular cloning, functional analysis and developmental regulation. Proc Natl Acad Sci USA 87:9028-9032

Tafuri SR & Wolfe AP (1992): DNA-binding, multimerization and transcription stimulation by the Xenopus Y-box in vitro. New Biol 4:349-359

Tanaka M, Lai J-S & Herr W (1992): Promoter-selective activation domains in Oct-1 and Oct-2 direct differential activation of an snRNA and mRNA promoter. Cell 68:755-767

Timmers HTM & Sharp PA (1991): The mammalian TFIID protein is present in two functionally distinct complexes. Genes Dev 2:1946-1956

Tribioli C, Biamonti G, Giacca M, Colonna M, Riva S & Falaschi A (1987): Characterization of human DNA sequences synthesized at the onset of S-phase. Nucleic Acids Res 15:10211-10232

Umek RM & Kowalski D (1988): The ease of DNA unwinding as a deteminant of initiation at yeast replication origins. Cell 52:559-567

Umek RM, Linskens MHK, Kowalski D & Huberman JA (1989): New beginnings in studies of eukaryotic DNA replication origins. Biochim Biophys Acta 1007:1-14

Usuda Y, Kubota A, Berk AJ & Handa H (1991): Affinity purification of transcription factor IIA from HeLa cell nuclear extracts. EMBO J 10:2305-2310

Vassilev L & Johnson EM (1989): Mapping initiation sites of DNA replication in vivo using polymerase chain reaction

amplification of nascent strand segments. Nucl Acids Res 17:7693-7705

Vassilev LT, Burhans WC & DePamphilis ML (1990): Mapping of an origin of DNA replication at a single-copy locus in mammalian cells proliferating under physiological conditions. Mol Cell Biol 9:4685-4689

Vassilev L & Johnson EM (1990): An initiation zone of chromosomal DNA replication located upstream of the c-myc gene in proliferating HeLa cells. Mol Cell Biol 10:4899-4904

Vaughn JP, Dijkwel PA & Hamlin JL (1990): Replication initiates in a broad zone in the amplified CHO dihydrofolate reductase domain. Cell 61:1075-1087

Vinson CR, Sigler PB & McKnight SL (1989): A scissors-grip model for DNA recognition by a family of leucine zipper proteins. Science 246:911-916

Vinson CR & Garcia KC (1992): Molecular model for DNA recognition by the family of basic-helix-loop-helix-zipper proteins. New Biol 4:396-403

Weigel D & Jackle H (1990): The fork head domain: a novel DNA binding motif of eukaryotic transcription factors? Cell 63:455-456

Williams T, Admon A, Luscher B & Tjian R (1988): Cloning and expression of AP-2, a cell-type-specific transcription factor that activate inducible enhancer elements. Genes Dev 2:1557-1569

Williams T & Tjian R (1991): Analysis of the DNA-binding and activation properties of the human transcription factor AP-2. Genes Dev 5:670-682

Wistow G (1990): Cold shock and DNA binding. Nature 344:823-824

Wolberger C, Dong Y, Ptashne M & Harrison SC (1988): Structure of a phage 434 Cro/DNA complex. Nature 335:789-795

Wolffe AP, Tafuri S, Ranjan M & Familari M (1992): The Y-box factors: a family of nucleic acid binding proteins conserved from Escherichia coli to man. New Biol 4: 290-298

Yaneva M & Busch H (1986): A 10S particle released from deoxyribonucelase-sensitive regions of HeLa cell nuclei contains the 86-kilodalton-70-kilodalton protein complex. Biochemistry 25:5057-5063

Yaneva M, Wen J, Ayala A & Cook R (1989): cDNA-derived amino acid sequence of the 86-kDa subunit of the Ku antigen. J Biol Chem 264:13407-13411