ISAS - INTERNATIONAL SCHOOL FOR ADVANCED STUDIES

Protein-DNA interactions at a Human DNA Replication Origin

Thesis submitted for the Degree of
Doctor Philosophiae
(Academic Year 1996/97)

Candidate: Paolo Norio

Supervisors: Prof. Arturo Falaschi
Dr. Mauro Giacca

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

TRIESTE

CONTENTS

4 DISCUSSION

INTRODUCTION

1.1 ORIGINS OF REPLICATION: A SIMPLE MODEL

Ten years after the identification of the DNA structure by Watson and Crick, a model was proposed to explain the replication of this molecule in by Jacob, Brenner and Cuzin (Jacob *et al.* 1963). At the beginning of the 1960's very little was known about the mechanisms controlling DNA replication; the enzyme isolated by Arthur Kornberg was able to reproduce *in vitro* copies of a DNA template in the presence of a primer but without any apparent control. On the other hand, pieces of the bacterial chromosome, when inserted in new recipient cells were able to replicate only after integration in the host chromosome. Likewise, it was observed that independent genetic elements such as phages or sex factors could lose by mutation their ability to multiply autonomously.

In their article "On the regulation of DNA replication in bacteria" Jacob, Brenner and Cuzin proposed the existence of a genetic element able to replicate autonomously within a cell; the name given to this unit is *replicon*. It was assumed that the capacity to behave as a replicon depended on the presence of at least two specific functional determinants: a) a structural gene controlling the synthesis of a specific protein, or *initiator*, involved in the initiation of DNA replication and b) an origin of replication, or *replicator*, representing a target sequence recognized by the *initiator* protein (Figure 1.1.1).

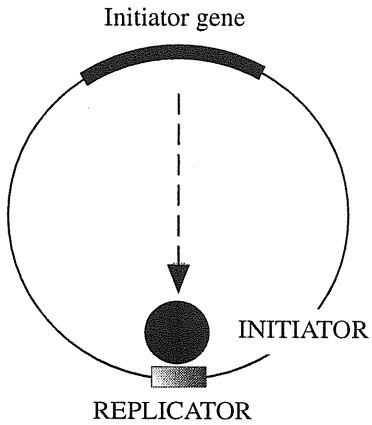


Figure 1.1.1 The *replicon* model according to Jacob, Brenner and Cuzin.

This model was superbly confirmed in its basic structure by the subsequent studies on bacterial and phage DNA replication with the addition of further control systems to ensure a response to various signals, linking replication with cell growth and division. Eukaryotic viruses and more recently simpler eukaryotes such as yeast, have been proven to follow the same strategy.

Bramhill and Kornberg proposed an important extension of the replicon model (Bramhill and Kornberg, 1988) in which the *initiator* protein serves three different and essential functions: 1) origin recognition; 2) induction of local unwinding within the origin and 3) recruitment of other replication enzymes by protein-protein interactions.

In other words, replication begins at or near the sites where specific proteins bind specific DNA sequences to initiate DNA unwinding (Kornberg and Baker 1992; DePamphilis 1993). This process allows the entrance of specific helicases which extend the unwound DNA tract and enable the assembly of the replication machinery. This event is followed rapidly by initiation of DNA synthesis on one or both of the DNA templates. This synthesis is usually carried forward by the standard replication fork mechanism as shown in Figure 1.1.2: synthesis is continuous on the forward arm (leading strand) while that on the retrograde arm (lagging strand) is discontinuous.

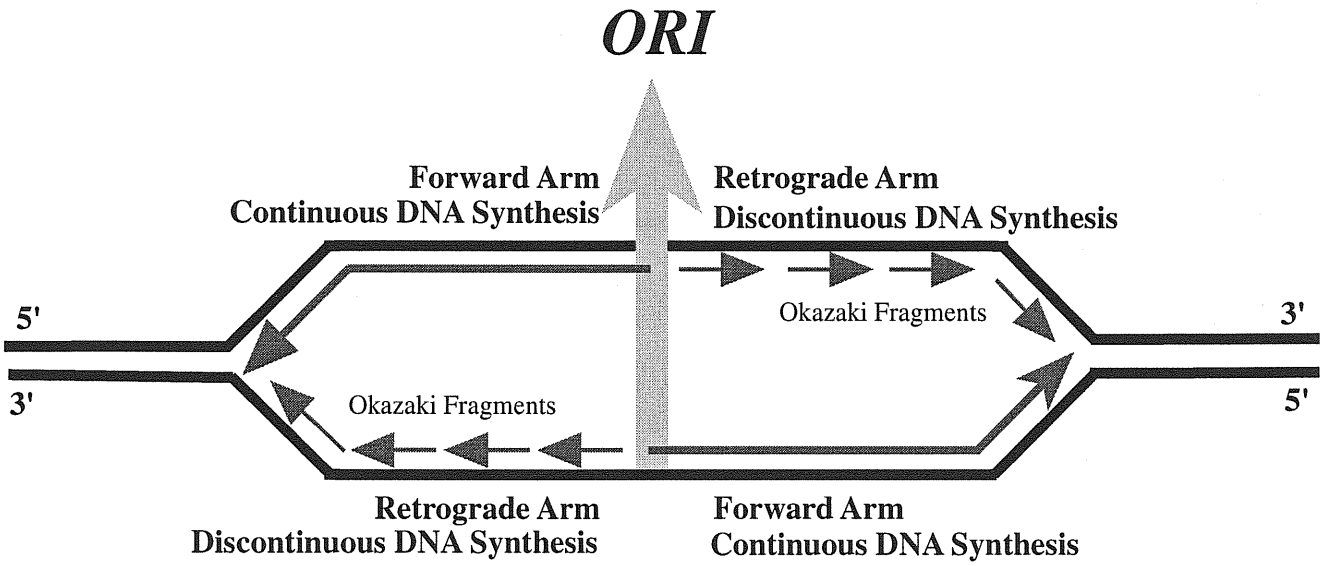


Figure 1.1.2 The replication fork model

The above described melting of DNA initiates replication and gives rise to the concept of *origin of replication* (*ori*) as cis-acting site where the DNA synthesis begins.

Presently, in the light of the available information on the best-understood replication systems (*E. coli*, SV40, *S. cerevisiae*), we can draw the common anatomy of the origins of replication. For all these units the following elements can be recognized: a sequence recognized by *initiator* proteins also referred as origin recognition element (ORE), a DNA-unwinding element (DUE) required for initial unwinding of the double helix and one or more binding sites for specific transcription factor(s) with an auxiliary function in the origin activity (Figure 1.1.3). Together, ORE and DUE constitute what is normally referred to as the *ori* core. Spacing, orientation and arrangement of these three components are usually critical for *ori* function.

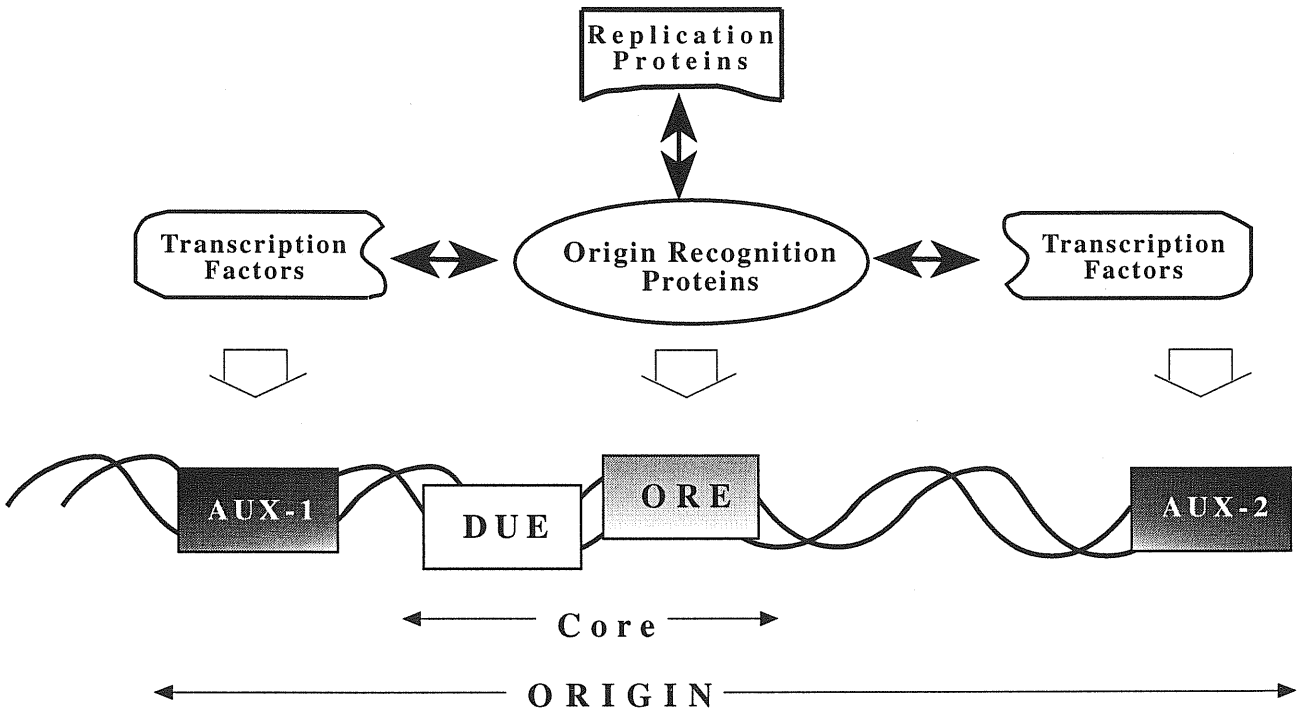


Figure 1.1.3 Origin structure in simple genomes.

Another feature that can be expected to be present in any replication origin is the presence of an element able to anchor the origin of replication to some "solid" cell structures, which will eventually allow the separation of the daughter chromosomes into separate cells. This is assured in bacteria by the binding of the *ori* to a specific site of the membrane and is most likely assured in eukaryotes by the attachment of *ori* to the chromosome scaffold.

Examples of classical origin recognition proteins are represented by dnaA protein, EBNA1 and Large T-antigen in *Escherichia coli*, Epstein-Bar and SV40 viruses respectively. All these molecules have the ability to bind the DNA in a sequence specific manner as predicted by the replicon model (one or more binding site for these proteins are normally present in the ORE's).

On the contrary, the DUE is not a well defined sequence like ORE. The origin unwinding can be favored by different DNA sequences depending on base-stacking interactions forces. In the case of SV40 it has been demonstrated that the DUE precisely corresponds to the *ori* element (Hay *et al.* 1982; Hay *et al.* 1984; Guo *et al.* 1991).

Ori auxiliary (AUX) components consist of transcription factor-binding sites that facilitate *ori* core activity. Experimental evidence shows that *ori* auxiliary components must bind specific transcription factors in order to facilitate *ori* activity (an example of these factors are Sp1 in the SV40 origin of replication and ABF1 in some yeast ARSs). The presumed involvement of specific protein-protein interactions between transcription factors and the origin recognition proteins could explain the transcription factor requirement. Transcription factors that activate one origin do not necessarily activate another and the ability of a transcription factor to stimulate a promoter does not necessarily reflect its ability to stimulate an origin (for review: DePamphilis 1993; Donovan *et al.* 1996).

1.2 THE THOUSAND AND ONE REPLICONS: COORDINATED REGULATION OF MULTIPLE ORIGINS OF REPLICATION IN EUKARYOTES

As genomes began to be more large and complex the evolution of a mechanism to rapidly replicate large amounts of DNA probably became a formidable task. The evolutionary answer to this question was the increase of the number of replicons per genome thus opening the way to the expansion of genome size and consequently to the explosion of morphological and functional diversity seen among eukaryotes. An effect of this multiplication was the need to coordinate activity of all the replicons during the cell cycle.

Whilst budding yeast, containing about four times the amount of DNA per haploid cell as *E. coli*, replicates its genomic DNA from 250-400 origins distributed on 16 chromosomes, the *E. coli* genome contained in a single chromosome, is replicated from a single replication origin (oriC). Thus the amount of DNA replicated from an individual replication origin is considerably smaller in eukaryotes than in prokaryotes (the size of the eukaryotic replicon is approximately 100 kb in metazoans and 10-20 kb in yeast while the *E. coli* replicon is about 5000 Kb).

The genomes of higher eukaryotes consist of approximately 10^4 to 10^5 tandemly arranged replication units of which the activity needs to be coordinated during the cell-cycle (Cairns 1966, Huberman and Riggs 1968; McKnight *et al.* 1977; Bozzoni *et al.* 1981). Activation is a highly ordered process which follows reproducible and developmentally regulated spatial and temporal programs (Fangman and Brewer, 1992; Diller and Raghuraman, 1994). In some cases, activity is triggered early in the S-phase of the cell cycle, while it is triggered late in other cases (Huberman and Riggs 1968; Marx 1995).

Various experimental observations shows that not all replicons are regulated identically. The length of S phase in cells of different animals can vary by more than two orders of magnitude. This great variation in length is not due to differences in the rate of DNA synthesis but appears to reflect differences in both the number of activated origins and the timing of activation of clusters of origins during the S-phase. Fewer functioning origins result in an increase in inter-origin distance and consequently, a longer time is required for adjacent replication forks to

meet and a longer S phase is necessary (Callan 1973, Blumenthal *et al.* 1973).

Defined regions within the chromosomes replicate at distinct and very different times during the S-phase. The simplest interpretation of these observations is that chromosomes contain clusters of many replication origins which are activated at different times during S phase (Hand 1978). Experiments performed in replicating nuclei (reviewed in Mills *et al.* 1989) have shown that replication sites are not located diffusely throughout the nucleus but rather are confined to several hundred discrete foci, each focus containing several hundred replication forks. In this process the nuclear structure may have an organizational function, pulling together colinear arrays of replication origins into clusters that function synchronously.

Besides these general features, the timing of replication of every origin is developmentally regulated. The genes within a DNA segment encompassing several hundred kilo bases can replicate late and sequentially in one tissue and early and coordinately in another tissue type (Dhar *et al.* 1988, Hatton *et al.* 1988). Generally, genes that replicate late in S phase are transcriptionally inactive, while actively transcribed genes are often early replicating (Hatton *et al.* 1988). Changes in the expression of genes either in different tissue types (Dhar *et al.* 1988) or in response to experimentally manipulated factors, such as 5-azacytidine treatment (Paterno *et al.* 1985) are often accompanied by changes in the timing of replication.

These results suggest that the timing of activation of origin function during S-phase is controlled primarily by chromosomal context and does not appear to be specified by the origin itself.

1.3 THE YEAST REPLICON AS A PARADIGM FOR EUKARYOTIC REPLICATION ORIGINS

From the studies and the considerations reported above, it is clear that most of the models we have on replication origins derive from prokaryotes or from relatively simple systems such as the eukaryotic DNA-virus replicons. These simple genomes are characterized by the presence of a single replicon per DNA molecule and the replication origin is invariably a precisely defined, constant site.

In this context, the eukaryotic genomes with their multiple replicons present particular problems.

The model proposed by Jacob, Brenner and Cuzin assumes that the replicon is an independent unit, thus any segment of DNA that contains a *replicator* would be expected to replicate autonomously when introduced in cells that produce the appropriate *initiator*. This is the rationale that led to the identification of plasmids containing autonomously replicating sequences (ARS) elements in yeast (Stinchomb *et al.* 1979; Struhl *et al.* 1979; Stinchomb *et al.* 1980; Chan *et al.* 1980). These elements were isolated tanks to their capacity to increase the plasmid efficiency of transformation of yeast cells by a factor of 100-1000.

In 1987, two groups using two-dimensional gel techniques (2D-gels) to examine replication intermediates, demonstrated that bidirectional DNA replication actually initiates specifically within or very close to these ARS elements (Brewer and Fangman, 1987; Huberman *et al.* 1987). It was also demonstrated that whenever origins have been mapped in yeast an ARS element was always present. Furthermore, mutations that abolish ARS function also abolish chromosomal origin activity (Deshpande and Newlon 1992). However, not all chromosomal sequences functioning as ARSs in the plasmid assay serve as chromosomal origins. As the removal of an ARS from its chromosomal location and its placement on a plasmid can activate origin function, it is clear that flanking chromosomal sequences can influence origin function. Several reports revealed that chromosomal position is an important determinant affecting both the efficiency of origin use and the time in S phase when an origin is used (Brewer *et al.* 1993). It is safe to state that when inserted in plasmids, ARSs act both as genetically defined *replicator*s and biochemically defined origins of DNA replication.

Biochemical and genetic methods provide the means for the dissection of the yeast *ori*s, the establishment of their functional domains and the identification of a short consensus sequence (ARS Consensus Sequence, ACS also described as A element) shared by all ARSs (Fangman and Brewer 1991; Marahrens and Stillman 1992). The ACS is a 11 bp sequence that, in all cases tested, resulted to be essential for both ARS and the origin of DNA replication functions. Nevertheless the ACS by

itself is not sufficient for origin function, and in all cases tested, a flanking sequence is also essential (Figure 1.3.1).

This flanking sequence is generally composed of several modules that contribute to origin function. As a general principle, mutations of a single module normally only reduce the ARS activity. However, when these elements are mutated in pairwise combination, ARS function is dramatically reduced to the level of mutations in the A element (Bell *et al.* 1993b). This modular structure has been described for several budding yeast origins but, despite the similarity in the overall configuration, scarce sequence similarity between the various modules was found (Rao *et al.* 1994). In a functional complementation assay in ARS containing plasmids, the ability of the modules to substitute only for similarly positioned elements was demonstrated. This result suggests either that the location of these elements is critical for their function or that they perform distinct functions. The importance of the flanking modules was also confirmed in the chromosomal context where mutations in the B elements gave the same effect as that of mutations in the ACS (Bell *et al.* 1993b; Rao and Stillman 1995).

The existence of such sequences strongly suggest that the replicon model might be helpful in understanding the initiation of DNA replication in eukaryotic chromosomes.

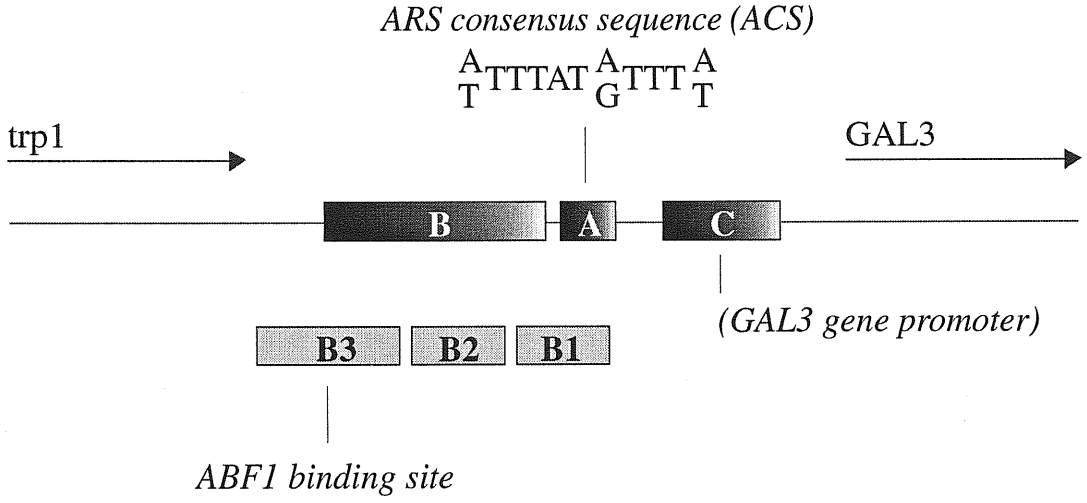


Figure 1.3.1 The *S. cerevisiae* Autonomous Replication Sequence 1 (ARS1). Earlier studies showed that three regions were important for ARS activity (A, B and C). Later, other studies demonstrated that it was possible to subdivide the B region in three functionally distinct subregions.

1.4 AUTONOMOUS REPLICATION STUDIES IN HIGHER EUKARYOTES

During the last two decades considerable efforts have been dedicated to the identification of *replicators* in higher eukaryotes. If cis-acting sequences are required for origin function in cellular chromosomes, the best way to identify and characterize them is by functional assay.

Fiber autoradiographic studies in the 1960's vividly demonstrated the presence of origins of bidirectional replication at around 100 kb intervals in mammalian DNA (Huberman and Riggs, 1968). However, it has been extremely difficult to prove that these sites actually correspond to genetic *replicators*. The difficulty in identifying OBRs (Origins of Bidirectional Replication) in mammalian cells derives from the non-functionality of the conventional methods of *ori* search based on the ability to sustain autonomous replication (Biamonti *et al.* 1985, Gilbert and Cohen, 1989; Burhans *et al.* 1990; Caddle and Calos, 1992).

Despite reports that some cloned sequences may be more efficient *replicators* than others in ARS assays (Frappier *et al.* 1987; Sudo *et al.* 1990; Mc Whinney and Leffak 1990; Virta-Pearlman *et al.* 1993; Berberich *et al.* 1995), it has yet to be shown that mutations in any such elements alter origin function. Indeed, most mammalian sequences seem to replicate to some extent when re-introduced into mammalian cells (Heinzel *et al.* 1991) but the replication event in these cases remains at an unacceptably low signal-to-noise ratio.

The reasons for the failure of the functional assays in mammalian cells could be several. The high frequency transformation assay for the identification of origins of replication is limited by its ability to quantify ARS function only in relation to the transformation activity which is not necessarily correlated in all organisms. In addition, while all yeast origins of replication are ARS elements, not all ARS elements function as origins in their chromosomal location (Fangman and Brewer 1991). Thus, also in yeast, the ARS assay alone is not sufficient to identify chromosomal origins of DNA replication that probably have more severe constrains to work as origins.

The higher complexity of the metazoan cell could further restrict the ability of any sequence to behave as an autonomous origin making the ARS assay completely unsuccessful. Some factors that could be important in the establishment of episomal autonomous replication in

mammalian cell are: the length of the construct, the association with the nuclear matrix, suitable DNA ends and possibly an epigenetic priming for the replication of the template (as binding of protein factors not always available, modifications of the DNA, etc.).

1.5 *REPLICATORS* IN HIGHER EUKARYOTES

The usefulness of the replicon model in describing the eukaryotic DNA replication was dealt a serious blow by the experiments of Harland and Laskey (1980), which showed that any kind of DNA molecule was able to replicate efficiently and in a once-per-cell-cycle manner when microinjected into *Xenopus leavis* eggs. The lack of need for specific *replicator* sequences is not unique to *Xenopus* but is seen in, at least, some other early metazoan embryos.

The apparent lack of sequence requirement for *Xenopus replicators* may not be a universal feature of higher eukaryotes. Physical mapping of the replication start site has shown that initiation is not random in mammalian chromosomes (as described below). The distinction between *origins* as biochemically defined replication start sites and *replicators* as genetically defined sequence elements required for initiation becomes critical here because, although mammalian replication origins have been identified, there is little genetic information indicating that these sites contain DNA sequences that are essential for replication. The lack of a functional assay for higher eukaryotes is a further indication in this sense.

On the other hand, there are also indications that *replicators* and *origins* could be at least partially overlapping also in higher eukaryotes. An origin of replication has been mapped in the β-globin locus by several techniques in a region between the δ- and the β-globin genes (Kitsberg *et al.* 1993; Aladjem *et al.* 1995). Initiation no longer occurs in this region in cells derived from patients suffering from β-thalassemia in which an 8 Kb sequence of DNA containing the β-globin origin has been removed. In these cells, the β-globin locus is replicated passively from an uncharacterized origin outside of the β-globin locus. Presumably, some DNA sequence within this region is essential for specifying an origin at this location. The expression of the β-globin genes requires the locus control region (LCR), a sequence more than 50

Kb upstream from the β-globin gene. Besides its function in transcription, the LCR proved also to be essential for the firing of the β-globin origin because the β-globin locus is replicated passively in mutants where the LCR has been deleted.
These results indicated that also in higher eukaryotes sequences located well outside of a biochemically defined origin can be essential for that origin to fire. Moreover these results could also explain the difficulties in identifying mammalian *replicator*s by a functional assay.

1.6 CHROMOSOMAL MAPPING OF ORIGINS OF DNA REPLICATION

The failure of functional studies to identify higher eukaryotes *replicators* led to the devising of alternative strategies (Tribioli et al. 1987; Falaschi *et al.* 1988; Falaschi *et al.* 1993; Falaschi and Giacca 1994; Hamlin and Dijkwel 1995).
Origin-mapping is, at the moment, the only way to identify sequences that could be involved in the *replicator* function of higher eukaryotes, being the sequences where replication actually starts in the chromosomal contest of complex genomes.
The techniques available today for origin-mapping in metazoan cells can be basically divided, depending on the properties of the OBR (*ori*) they study into the following three categories (Figure 1.6.1 A-C): structural analysis of the replication intermediates, analysis of the leading/lagging strand polarity and chromosomal mapping of nascent DNA strands. These categories can be further divided according to the principle adopted for mapping.

Structural analysis of the replication intermediates (Figure 1.6.1 A)

Two complementary two-dimensional (2-D) gel electrophoretic mapping techniques were developed to obtain an intimate view of replication intermediates. Both techniques rely on the unique migration pattern of replication intermediates in agarose gels and utilize defined molecular probes to examine these intermediates in a genomic region of interest. They were devised originally to analyze yeast replicons and were used to show that ARS elements usually, but not always, serve as origins in the chromosome (Brewer and Fangman 1991; Newlon and Theis 1993).

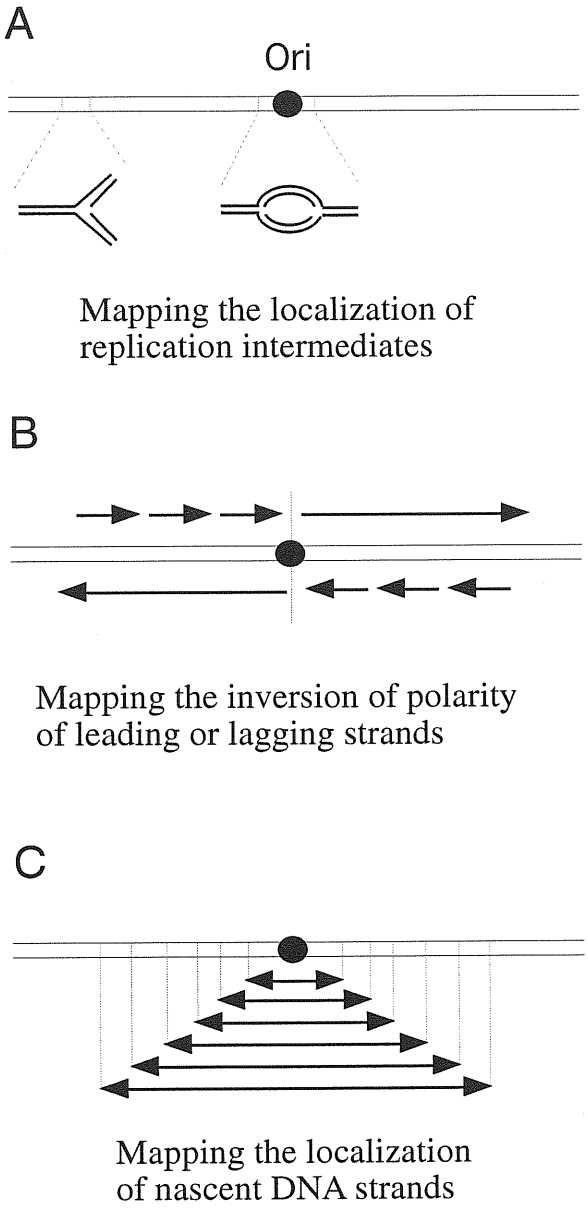


Figure 1.6.1 Techniques available for origin mapping in metazoan cells

A) Neutral/neutral 2-D gel

In the neutral/neutral 2-D gel method (Brewer and Fangman 1987), replication intermediates are separated in the first dimension on the basis of molecular mass (using low gel concentrations), which varies from 1n (unreplicated) to 2n (fully replicated) and in the second

dimension on the basis of both mass and shape (using high gel concentrations).
Characteristic arcs are traced by fragments containing either a single fork (Figure 1.6.2a), an origin (Figure 1.6.2b, c) or a termination structure (Figure 1.6.2d). The replication intermediates are then visualized, in a region of interest, by hybridization with an appropriate radioactive probe.

B) Neutral/alkaline 2-D gel

The neutral/alkaline 2-D gel method (Nawotka and Huberman, 1988) determines the direction of fork movement through a chromosomal locus. A restriction digest of replicating DNA is separated by mass in the first dimension. The nascent strands are then released from their templates with alkali and are separated by size in the second dimension. The direction of fork movement through any given restriction fragment can be determined by sequentially probing with adjacent probes. An origin-proximal probe traces a complete diagonal, whereas an origin distal probe detects only the nascent strands (Figure 1.6.3). Logically it can be deduced that a region between two divergently moving forks must contain an origin.
In principle these techniques can potentially localize an origin of replication within hundreds of base pairs but in practice, their application is seriously limited by the complexity of the genome studied.

2-D gel based origin-mapping methods have been successfully used to map replicons in simple genomes (Vassilev and DePamphilis 1992; Benard *et al.* 1996, Dhar *et al.* 1996; Little *et al.* 1995). However, when applied to complex genomes, 2-D gel based methods become problematic due to their relatively low resolution limits and the potential ambiguities in the interpretation of the results (Linskens and Huberman 1990). However a few convincing reports exist on origin localization using these methods (Table 1.6.1). Such reports are puzzling since in most cases, the hybridization patterns obtained indicate the presence of an origin-region rather than a fixed *ori*-site.

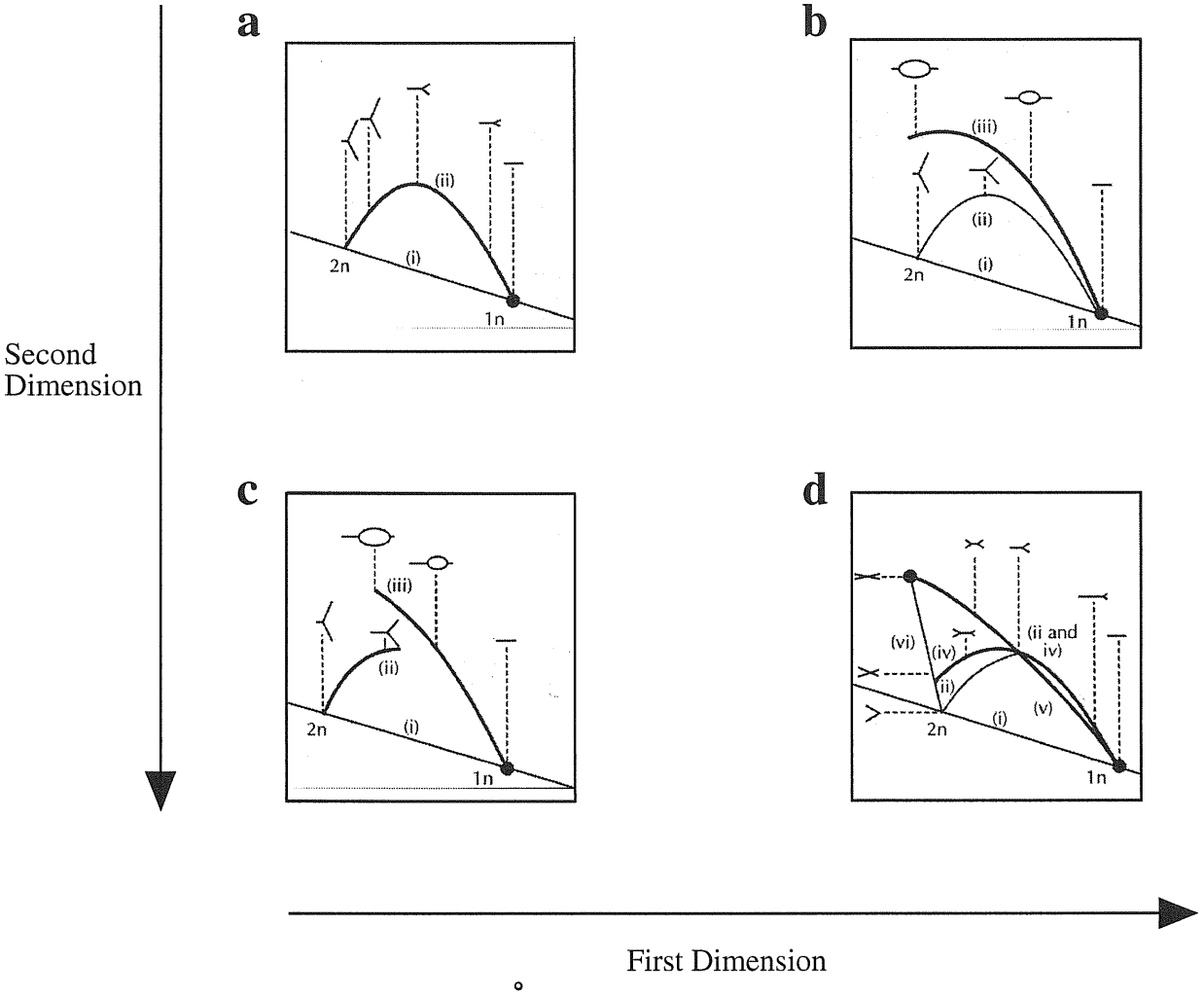


Figure 1.6.2. The principle of the neutral/neutral 2-D gel replicon-mapping technique. Idealized autoradiographic images obtained when a digest of replicating DNA is hybridized to probes for fragments that contain different intermediates. Curve (i) represents the diagonal of non-replicating fragments from the genome as a whole. **a)** A complete single fork or simple Y-arc (ii) from a fragment replicated passively from an outside *ori*. **b)** A fragment with a centered *ori* traces a curve (iii) that arches over the single fork arc (curve ii). **c)** A fragment containing an off-centered *ori* traces the beginning of a bubble arc (curve iii) when the replication fork crosses one restriction site. **d)** A fragment containing two approaching forks. If they meet asymmetrically or symmetrically, curves iv and v are obtained, respectively. If there is a fixed terminus in a fragment, the collected X-shaped structures would result in a concentrated spot somewhere on curve vi, on which any recombination structures present would fall also.

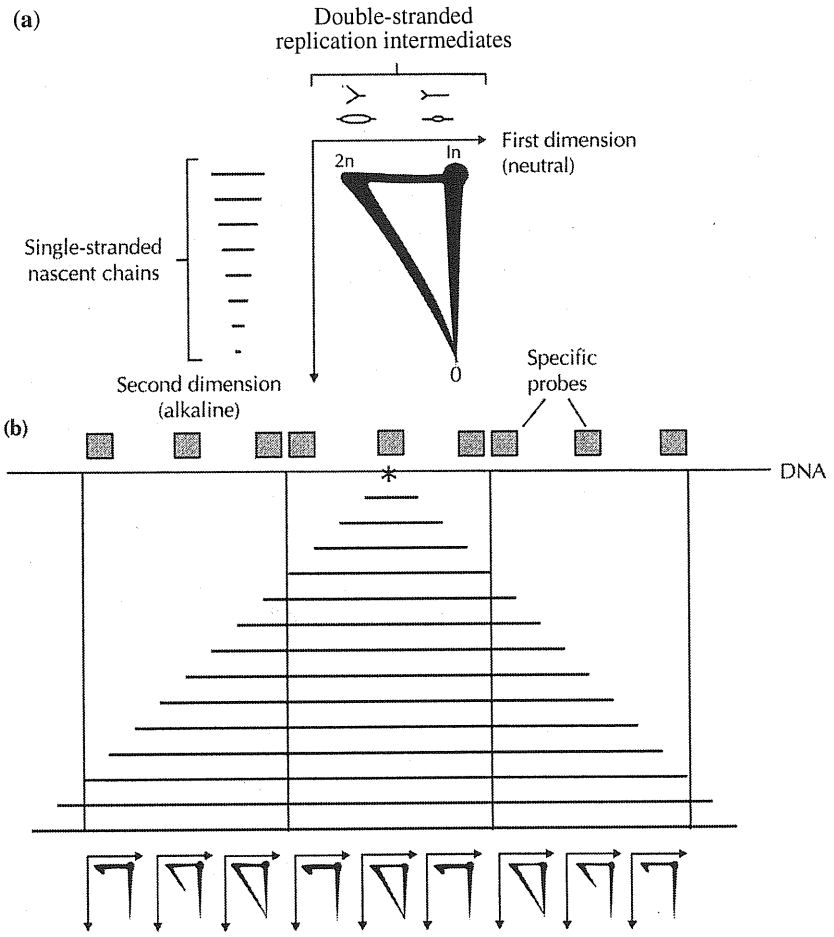


Figure 1.6.3 Principle of the neutral/alkaline 2-D gel replicon-mapping technique. (a) Idealized patterns of migration of double-stranded replication intermediates in the first (neutral) dimension and the single stranded nascent chains in the second (alkaline) dimension. (b) Three adjacent restriction fragments are demarcated by the vertical lines. An ori is positioned in the central fragment (asterisk), and nascent strands of different sizes are indicated. The panels below show the autoradiograms that would be obtained when each of the indicated probes (shaded boxes) is used to illuminate its cognate restriction fragment.

The resulting hybridization patterns obtained in 2-D gel-based methods are usually more complex to analyze than those expected from the model, suggesting the possible presence of unusual DNA structures responsible for the unexplained hybridization spots.

In higher eukaryotes, the 2-D gel mapping approach has been best applied to highly amplified loci or to single copy loci utilizing synchronization procedures, typically with aphidicolin, to enrich for the newly synthesized DNA. Aphidicolin, an inhibitor of at least three eukaryotic DNA polymerases (Wang 1991), blocks the progression but not the initiation of DNA synthesis (Huberman 1981; Mosca et al. 1992). Moreover, inhibition of DNA replication can cause reinitiation events (Mariani and Schimke 1984), possibly leading to multiple-branched initiation intermediates and to branch migration of nascent DNA (Dijkwel et al. 1991). To rule out these potential artifacts, another synchronizing agent, mimosine (Lalande 1990) has been used in studies that still tend to confirm the delocalization of origins (Dijkwel and Hamlin 1992; Mosca *et al.* 1992). It is however important to note that the mechanism of action of mimosine in cell synchronization is yet unknown (Kalejta and Hamlin 1997). Moreover, use of the 2-D gel approach in animal cells requires previous enrichment of replication intermediates by DNA purification procedures, including benzoylated naphthoylated(BND)-cellulose chromatography (Levine *et al.* 1970) or selective isolation of the nuclear matrix (Dijkwel *et al.* 1991). These procedures can easily alter the structure of replication intermediates, for example, by introducing breaks into replication bubbles consequently masking the presence of an active *ori* (Linskens and Huberman 1988). The 2-D gel mapping results, should be therefore interpreted with cautious skepticism, especially when applied to cell of higher eukaryotes.

A good example is the DHFR *ori* in the Chinese hamster ovary (CHO) cells where origin mapping technique, other than 2-D gel, mapped a DNA origin or replication to a precise location downstream from the DHFR (dihydrofolate reductase) gene, in a Chinese hamster cell line with multiple copies of the DHFR region (Handeli *et al.* 1989; Burhans *et al.* 1990; Pelizon *et al.* 1996).

Table 1.6.1. Metazoan *oris* mapped with methods based on 2-D ge electrophoresis

Region	Organism	Main conclusion	References
DHFR locus	Chinese hamster	55 kb initiation zone with preferred delocalized *ori*-sites	Vaughn *et al.* 1990 Dijkwel *et al.* 1991
Chorion locus	Drosophila	*ori* contained in 1 kb region	Heck *et al.* 1990
Histone gene	Drosophila	multiple initiation sites in the 5 kb histone transcription unit	Shinomiya *et al.* 19
rDNA locus	Human	multiple initiation sites in 31 kb non transcribed spacer (NTS)	Little *et al.* 1993
	Drosophila	initiation-zone in the NTS	Hyrien *et al.* 1995
Puff II/9A	Sciara sp.	around 1 kb initiation zone	Liang *et al.* 1993 Liang *et al.* 1994
Polymerase α gene	Drosophila	10 kb initiation zone 3' to pol α gene	Shinomiya *et al.* 19

The application of the 2-D gel methods, on the contrary, showed a delocalization of origins in a broad initiation zone of 55 kb (Vaughn *et al.* 1990; Dijkwel and Hamlin 1992)

Analysis of the leading/lagging strand polarity (Figure 1.6.1B)
Since DNA polymerase synthesizes DNA only in the 5' to 3' direction, the OBR is the site where continuous DNA synthesis on the leading strand and discontinuous DNA synthesis on the lagging strand, switch template and invert polarity. This switch can be obviously detected following both the Okazaki fragment or the leading strand.

A) Okazaki fragment polarity
This technique (Burhans *et al.* 1990) relies on the hybridization of radiolabeled Okazaki fragments to cloned strand-specific DNA templates that span the initiation site of replication from a region of interest. Although this method has been successfully used to map some metazoan origins (Table 1.6.2), it suffers an important drawback. The

Okazaki fragments are scattered uniformly along the genome and thus it is not feasible to use this method to detect *ori* in single copy loci in physiologically growing cells. In order to obtain a sufficiently high signal-to-noise ratio the cells need to be synchronized, permeabilized and labeled with a very high amount of ^{32}P-dNTPs (Burhans *et al.* 1990), raising the same problems earlier mentioned for the 2-D gel techniques.

B) Imbalanced DNA synthesis

On the other hand, a leading strand polarity-switch based assay relies on the observation that the protein synthesis inhibitor, emetine, selectively inhibits lagging strand synthesis (Burhans *et al.* 1991).

Table 1.6.2. Metazoan *oris* mapped with methods based or leading/lagging strand polarity switch

Region	Organism	Main conclusion	References
DHFR locus	Chinese hamster	*Ori* site within a 15 kb region	Handeli *et al.* 1989
APRT locus	Chinese hamster	*Ori* site within a 6 kb region	Handeli *et al.* 1989
ADA gene	Mouse	11 kb origin region 5' to ADA gene	Carroll *et al.* 1993
b-globin	Human	*Ori* mapped in a 2 kb fragment	Kitsberg *et al.* 1993
Ribosomal protein S14 gene (RPS14)	Chinese hamster	origin mapped in a 2.5 kb region overlapping the gene	Tasheva and Roufa19
CAD gene	Syrian and Chinese hamsters	origin localized in a 5 kb region in the CAD transcription unit	Kelly *et al.* 1995
c-myc	Human	origin mapped within 2.5 kb region upstream of the gene	Waltz *et al.* 1996

By using emetine-resistant BrdUrd-labeled (leading strand) DNA as a probe on positive and negative strands of recombinant M13 clones from the region in question, it is possible to determine the switch site for the leading strand synthesis and hence the origin. The main drawback of this method is the lack of knowledge on the exact mode of emetine action and whether such treatment, which considerably alters cellular metabolism, also affects the physiological pattern of DNA replication initiation.

Chromosomal mapping of nascent DNA strands (Figure 1.6.1C)
The initiation of DNA replication is characterized by the production of short nascent DNA strands. Besides their length, they can be distinguished from the bulk DNA on the basis of their replication timing having been synthesized earlier than any other DNA sequence within the same replicon. Since these nascent strands represent an extremely small portion of the total genome, the methods based on such an approach usually aim to obtaining a considerable enrichment of nascent strands sufficient enough to give a detectable signal for origin localization.

A) Replication timing
Replication timing studies are based on the principle that, within a single replicon, the segments of DNA closest to the origin should replicate earlier than others. Therefore, DNA fragments labeled with a ^{14}C or a ^{3}H DNA precursor after release of a cell population synchronized at the G1/S border, should represent the sequences of the origins activated in the early S.
The use of the replication timing studies in mapping mammalian *ori*s is particularly limited by the resulting low resolution. Although several studies have been conducted on the replication timing of contiguous loci in selected chromosomal regions and excellent information on the replication properties of wide regions of the genome could be obtained, yet no *ori* could be mapped precisely (Calza *et al.* 1984; Brown *et al.* 1987; Dhar *et al.* 1988; Taljanidisz *et al.* 1989; Gale *et al.* 1992; Selig *et al.* 1992; Spack *et al.* 1992).
Increased resolution in such studies can be obtained when applied to

highly amplified regions where there is a naturally higher abundance of the nascent strands as compared to the single copy regions. Such studies on the CHOC 400 cells (a Chinese hamster cell line with more than 1000 copies of an approximately 240 kb long genomic region encompassing the DHFR gene) suggested the presence of two origins of replication spaced about 22 kb apart downstream from the DHFR gene (Ma *et al.* 1990). The method consisted of synchronizing the cells at G1/S border and labelling the nascent DNA fragments with a radioactive nucleotide precursor immediately after the cells entry into S phase. The radiolabeled products were then analyzed either by direct visualization after specific restriction enzyme digestion (Heintz and Hamlin 1982) or used as probes against cloned DNA fragments of the region (Burhans *et al.* 1986). Recently, using a similar method, the nascent-strand abundance in not-synchronously growing cells was quantitatively evaluated by densitometric scans of the resulting southern-hybridization signals. This method led to the identification of an *ori* within the naturally repeated human ribosomal DNA (Yoon *et al.* 1995). Using an in-gel renaturation procedure to eliminate most of the background due to non amplified genomic sequences the technique can be further refined (Leu and Hamlin 1989). However the major drawback of such replication timing studies i.e. their low resolution in *ori* mapping, continues to exist. Even in the case of highly amplified loci, like that of DHFR in CHOC 400 cells, the results obtained with these techniques still need to be complemented with other kinds of origin mapping methods (Hamlin and Dijkwel 1995).

B) Strand extrusion

Enrichment of nascent strands can also be obtained by trapping newly synthesized DNA between two neighboring psoralen crosslinks flanking the origin and subsequently extruding it by alkaline denaturation (Russev and Vassilev 1982). This method once again confirmed in the CHOC 400 cells the presence of the two *ori*s identified earlier by the replication timing studies (Anachkova and Hamlin 1989). While the advantage of this method lies in its applicability to an asynchronous cell population, its major drawback is the potential alteration of the physiological controls of DNA replication upon psoralen treatment.

Employing high temperature to extrude nascent DNA, labeled with BrdUrd, from replication bubbles in synchronized cells (Zannis-Hadjopoulos *et al.* 1981) it is possible to purify the nascent double-stranded population by CsCl density gradient centrifugation. This method has been used to map an *ori* region near the chicken a-globin gene (Razin *et al.* 1986). A similar approach with Hg-dCTP labelling and metal-affinity chromatography as purification step, was utilized to construct a library of monkey DNA sequences putatively enriched in *ori* (Kaufmann *et al.* 1985).

C) *In vitro* run-off

An alternative approach to localize origins of replication and determine the direction(s) of replication is offered by *in vitro* run-off

This technique consists in allowing isolated nuclei to elongate DNA chains already initiated *in vivo* in the presence of a heavy DNA precursor. Since DNA in isolated nuclei does not undergo reinitiation events, the extent of incorporation of the label in each molecule is directly proportional to its distance from the *ori*. This method has been the basis of reports where, although at low resolution, some *ori*s have been localized (James and Leffak, 1986; Heintz and Stillman, 1988; Trempe *et al.* 1988; Leffak and James, 1989; McWhinney and Leffak, 1990).

D) PCR of nascent DNA strand

This technique is based on the fact that in an unsynchronized population of cells, the size of replication bubbles generated by each origin of replication will vary from small bubbles just initiated to large bubbles that are about to complete replication. Identify the sequence of the shortest nascent DNA molecule means to map the origin of replication (Figure 1.6.4).

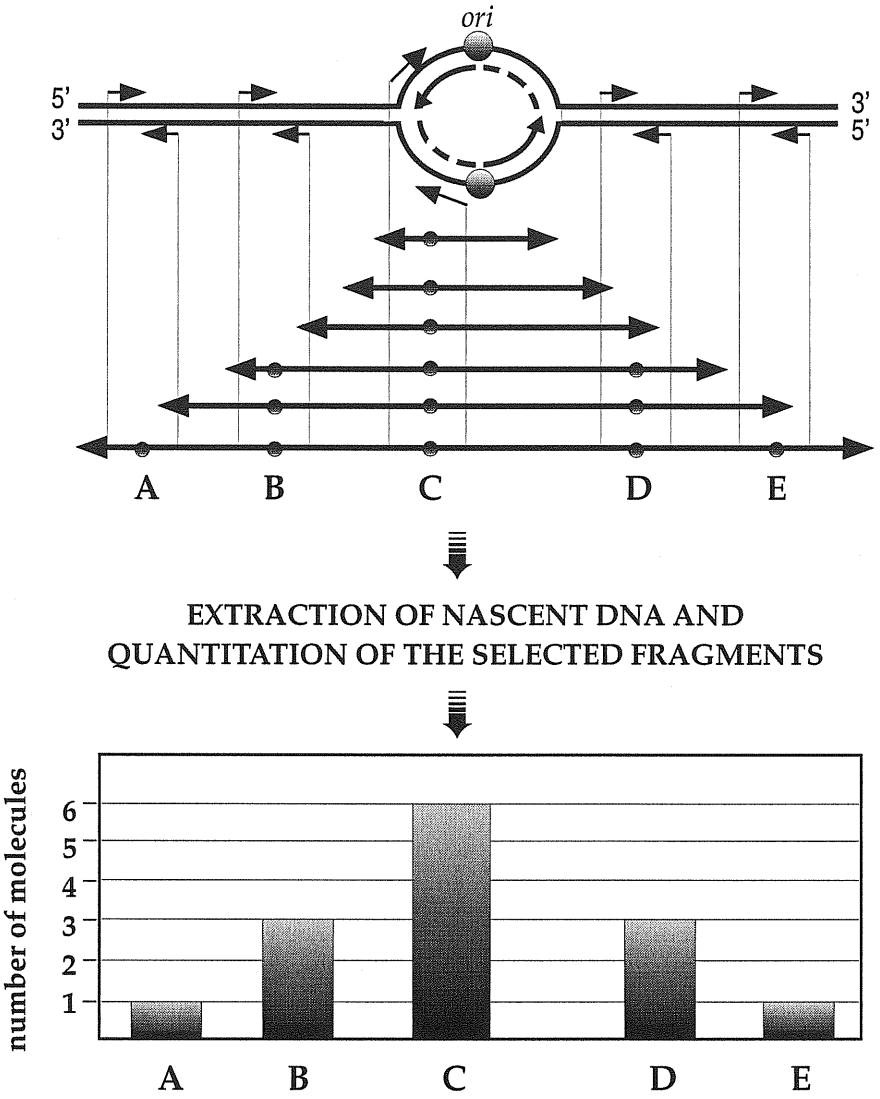


Figure 1.6.4 Ori-mapping procedure based on nascent DNA strand analysis. The short fragments of nascent DNA can be isolated and separated from the bulk DNA. Within this population of newly synthesized DNA molecules, the abundance of selected markers scattered within a genomic region (A through E) identifies the active origin of replication as the fragment that shows the highest enrichment.

The identification of these sequences cannot be achieved by direct hybridization, due to the low abundance of nascent DNA, but can be easily done by hybridization after some cycles of PCR amplifications (Vassilev *et al.* 1990a; Vassilev *et al.* 1990b).

In practice, for any genomic DNA region of a known sequence, containing a putative origin of replication, it is possible to synthesize pairs of oligonucleotides encompassing the region in order to amplify some specific segments of it. Each fraction of nascent DNA (labeled with BrdUrd) obtained, after denaturation, size fractionation in alkaline sucrose gradients and immunoprecipitation with anti-BrdUrd Ab, from unsynchronized cells, can be used as template for PCR amplification with all the primer sets.

Slot blotting and hybridization of the amplification reactions, with probes specific for the amplified regions, reveals that reactions containing the shortest fractions of nascent DNA hybridize preferentially to the probes nearest to the origin of replication. On the contrary, amplification reactions containing the longest fractions of nascent DNA hybridize equally well with all probes, allowing the detection of the fork progression direction.

The advantage of this method is that it avoids the use of metabolic inhibitors, does not require synchronization of cells and can also detect replication through single copy sequences in complex genomes.

However, it is worth noting that this PCR technique is not truly quantitative and, considering the variability of the experimental condition, it is extremely prone to artifactual results and difficulties in reproducibility. If these drawbacks could be overcome, PCR-based techniques would prove very advantageous in origin mapping.

Competitive PCR on nascent DNA, developed in our laboratory, is a technique that addresses the problems inherent in origin mapping by PCR based technology.

The development of this technique forms part of the work presented in this thesis and will be described later (Results and Discussion).

Table 1.6.3 lists various reports on origin-localization using methods based on analysis of nascent DNA strand size.

Table 1.6.3. Metazoan *ori*s mapped with methods based on nascent DNA strand size analysis

Region	Organism	Major Conclusions	References
DHFR locus	Chinese hamster	2.5 kb *ori* (*ori* b) from CHOC 400 (multiple DHFR copies) cell line 0.8 kb *ori* (*ori* b)from CHO K1 (single DHFR copy) cell line	Handeli *et al.* 1989 Pelizon *et al.* 1996
c-myc gene	Human	2 kb region 5' to the gene contains an *ori*	Vassilev *et al.* 1990
Histone gene repeat	Drosophila	multiple *ori* sites in the 5 kb histone gene repeat unit.	Shinomiya *et al.* 1993
Immunoglobulin heavy chain enhancer (Em)	Human	*ori* localized within a 1-2 kb region *ori* localized within 1 kb	Ariizumi *et al.* 199 Iguchi-Ariga *et al.* 1993
ADA locus	Mouse	1-2 kb *ori* region in murine fibroblasts with amplified ADA domain.	Virta-Pearlman *et al.* 1993
Ribosomal protein S14 gene (RPS14)	Chinese hamster	origin mapped in a 2.5 kb region overlapping the gene	Tasheva and Roufa 1994
Aldolase B locus	Rat	<1.3 kb initiation zone in synchronized hepatoma cells	Zhao *et al.* 1994
HSP 70 locus	Human	0.4 kb *ori* within the promoter of the gene	Taira *et al.* 1994
DNA polymerase α gene	Drosophila	*ori* mapped in a <10 kb region	Shinomiya *et al.* 1994
rDNA repeat	Human	*ori* within few kb upstream of the gene in HeLa cells	Yoon *et al.* 1995
	Human and rat	two *ori* s, one upstream and other downstream of the gene in both the organisms.	Gencheva *et al.* 199
Lamin B2 gene	Human	0.4 kb *ori* 3' to the gene in HL60 cells same *ori* functional in various other human cell lines	Giacca *et al.* 1994 Kumar *et al.* 1996

1.7 DNA-PROTEIN INTERACTIONS AT THE REPLICATION ORIGINS

The ARS binding factor 1

As yeast ARS elements meet all the criteria for a *replicator* sequence, a great deal of effort has been directed at the identification of proteins that specifically recognize the ARS elements in the hope that they would represent the corresponding *initiator* protein. Genomic footprinting of the yeast ARS1 revealed that the ACS and flanking sequences are occupied *in vivo* by a protein factor (Diffley and Cocker 1992).

The yeast transcription factor ABF1 was identified for its ability to bind the B3 element of ARS1 (Buchman *et al.* 1988; Diffley and Stillman 1988; Sweder *et al.* 1988). However because the B3 element is not essential for the ARS1 function, ABF1 cannot be a *initiator*.

Interestingly, the critical function of ABF1 appears to involve the localization of a transcriptional activation domain to the origin. When the ABF1 site was replaced by a GAL4-binding site, replication frequency became GAL4-dependent (Marahrens and Stillman 1992). In view of the weak affinity of RPA for duplex DNA, certain double-strand DNA-binding proteins may help to sequester RPA and to stabilize a pre-initiation complex at the point in the cell cycle prior to the formation of single strands. RPA itself seems to bind either to the acidic activation domain of a transcription factor or to a single strand DNA (Dutta *et al.* 1993; Li and Botchan 1993). Thus the formation of two alternative complexes may oppose each other; an unwound complex with RPA bound to the single strand DNA or a duplex DNA tethered to a transcription factor bound to RPA. In this context the function of ABF1 or of other transcription factors at the origins of replication could facilitate the step of DNA unwinding, increasing the probability for an origin-recognition element (ORE) to work as a replication origin.

Another possible role for ABF1 is suggested by studies on the ARS1 chromatin structure both on the chromosome and on plasmids carrying wild type or mutated ABF1 binding sites (Venditti *et al.* 1994). Nucleosome mapping and micrococcal nuclease *in vivo* footprinting showed that the ABF1 protein could act as a boundary element of chromatin structure by limiting the nucleosome invasion towards the

essential A-domain.

The ORC complex in yeast

A multiprotein complex (Origin Recognition Complex, ORC) has been purified and shown to bind *in vitro* to the ACS of ARS1 in an ATP-dependent manner and to generate footprints very similar to those obtained *in vivo* (Bell and Stillman 1992; Diffley and Cocker 1992; Rowley *et al.* 1995). Whereas ACS only is required for ORC to recognize DNA, ORC also protects additional sequences in the B domain of ARS elements both *in vivo* end *in vitro.*

ORC is a tight complex of six polypeptides which binds to the ACS of all origins tested to date. The binding of ORC to ARS's correlates with their functional activity. Mutations within the ACS, that eliminate ARS function *in vivo,* eliminate ORC binding *in vitro.* This renders ORC a strong candidate for a eukaryotic *initiator* protein complex (Bell *et al.* 1993a; Bell *et al.* 1993b; Li and Herskowitz 1993). Because these genomic footprints were performed on chromatin from logarithmically growing cells and the footprint over the ACS was nearly saturated, these results suggested that ORC must be bound at ARS1 during a large fraction of the cell cycle.

The genes encoding all six ORC subunits have now been isolated and all six are essential for viability in yeast (Bell *et al.* 1993a; Bell *et al.* 1995; Foss *et al.* 1993; Li end Herskowitz 1993; Micklem *et al.* 1993; Loo *et al.* 1995; Hardy 1996). A few of the subunit-encoding genes have been analyzed in more detail and all of the available evidence indicates that ORC is essential for initiating DNA replication. For example, *orc2*, *orc3* and *orc5* mutants show cell cycle aberration consistent with defects in DNA replication and lose ARS-containing plasmids at high frequency (Foss *et al.* 1993; Micklem *et al.* 1993; Fox *et al.* 1995; Loo *et al.* 1995; Hardy 1996). Additionally, direct examination of replication intermediates in both *orc2* and *orc5* mutants has indicated reduced initiation frequency from chromosomal origins (Liang *et al.* 1995; Loo *et al.* 1995).

These results all argue that ORC fulfills the criteria of an *initiator* protein in yeast DNA replication. It specifically recognizes *replicators* both *in vitro* and *in vivo* and it is essential for initiating DNA

replication. Subsequent genomic footprinting experiments, suggest that ORC is bound during most of the cell cycle as well as during the periods of quiescence. In particular, ORC remains bound at origins after DNA replication during the G2 and M phases of the cell cycle (Diffley *et al.* 1994). Therefore, the binding of ORC to the ACS, although essential is not sufficient to drive the initiation of DNA replication *in vivo*.

Another interesting observation is that, regardless of the fact that ACS remain protected during most of the cell cycle, neighbor sequences adjacent to the ACS show a protection pattern that change during the cell cycle. In particular, the pattern of protection observed at the ACS during the S and M phases (closely resembling those produced *in vitro* by the purified ORC complex) is substituted by an extended pattern during G1 (Diffley *et al.* 1994). The first interaction was named *post-replicative state* while that more extended was called *pre-replicative state*. This modification of the protection pattern suggests that the recruitment of other factors to the origin is probably a fundamental step in activating the origins which have been committed by the binding of ORC.
Several works identified candidate proteins involved in the assembly of the *pre-replicative complexes* (pre-RCs) and in interactions with the ORC complex in general (Dowell *et al.* 1994; Liang *et al.* 1995; Cocker *et al.* 1996; Leatherwood *et al.* 1996; Hardy 1996; Rowles *et al.* 1996; Zou *et al.* 1997). Nevertheless, whatever factors are responsible for the assembly of the pre-RCs they are probably cell cycle regulators that are required for DNA replication but not for origin selection. A detailed discussion on the role of these cell cycle regulators in the initiation of replication is beyond the scope of this thesis and for this reason it will not be treated.

ORC proteins in higher eukaryotes
Protein homologous of the ORC proteins were isolated from several eukaryotes (Ehrenhofer-Muray *et al.* 1995; Gavin *et al.* 1995; Gossen *et al.* 1995; Muzi-Falconi and Kelly 1995; Carpenter *et al.* 1996; Leatherwood *et al.* 1996; Rowles *et al.* 1996). Multisubunit complexes were isolated from *Xenopus* and *Drosophila*, suggesting that an ORC-like

complex is present also in higher eukaryotes.
Functional studies performed using cell-free systems for DNA replication showed that an origin recognition complex is not only present but appears to be, in these systems, essential for DNA replication (Carpenter *et al.* 1996; Leatherwood *et al.* 1996; Rowles *et al.* 1996; Romanowsky *et al.* 1996). Unfortunately, in these cell-free systems, *replicator* DNA sequences are not required for DNA replication suggesting that ORC is probably not able to confer alone the sequence specificity observed in yeast.
An explanation of this phenomenon could be that the ORC complexes used in the cell-free extracts (obtained form *Xenopus* egg extracts or from early *Drosophila* embryos) represent particular protein forms, evolved to allow an accelerated rate of replication, that do not require sequence specificity to bind the DNA. Alternatively, the eukaryotic ORC could be a sequence-specific DNA-binding protein that, when present in high concentrations in early embryos, is also able to bind DNA independently from its sequence. Finally, metazoan ORC in both early embryos and adults might be a nonspecific DNA-binding protein. In this last model, the apparent origin specificity seen in adult cells would come not from sequence specificity of ORC itself but from other features of the chromosome. The presence of sequence-specific DNA-binding proteins to facilitate the ORC binding could be one of these features. In this case, transcription factors might only be required for replication in nonembryonic cells, when the concentration of *initiator* factors became lower.

A recent experiment (Walter and Newport 1997) seems to offer a different explanation. Once a specific number of cells have been produced in the early Xenopus laevis embryo, replicon size during the S phase of the cell cycle increases. A similar increase in replicon size occurs when the concentration of nuclei in replication-competent *Xenopus* egg extracts exceeds a critical threshold. In this system, the origin recognition complex (ORC) do not become stoichiometrically limiting for initiation and similar amounts of this complex binds to chromatin regardless of replicon size (one every 8000 bp on the average).

This means that the binding of ORC to the *replicator* sequence could be only the first step of origin selection. A second step for origin selection could be produced by some unidentified factor (as proposed by Walter and Newport) or by the particular features of the chromosomal structure. This second round of selection would be responsible for the activation of only a subset of the many preformed ORC-DNA complexes.
These interpretations are also in accordance with another observation (Santocanale and Diffley 1996) in which the presence of pre-initiation complexes in yeast is not sufficient for the origin activation, being present both in ARS that are active and non active as replication origins. Moreover, this experiment shows that the proteins required for the assembly of the pre-replicative complexes are not factors that participate in the second round of origin selection.

We may conclude that in adult cells of higher eukaryotes (where the amount of chromatine-bound ORC is lower than in embryos) the final pattern of origin activation will probably be the result of one or more competitive events. The first event should be the competition of the various putative initiator sites for ORC binding. After ORC binding, the competition of the various pre-replicative ORC complexes for the access to the general replication machinery will probably make a second round of selection. In this second step probably lies the important role played in the selection process by the chromatin structure that surrounds the origins.

Other origin-binding proteins in higher eukaryotes

Due to the scarce information on the replication origins in higher eukaryotes only a small number of proteins have been identified as origin-binding proteins.
Replication of the Chinese hamster dihydrofolate reductase gene (DHFR) initiates near a fragment of bent DNA able to bind multiple cellular factors. One of these proteins was identified as a nuclear protein that also binds to domain B of the yeast origin of replication (RIP60). This protein was purified with another protein (RIP100) that also contains an ATP-dependent DNA helicase activity (Dailey *et al.* 1990). Other reports indicated that RIP60 dimers and multiples of dimers assemble

at the OBR of the DHFR amplicon in Chinese hamster ovary cells (Mastrangelo *et al.* 1993). Unfortunately, other indications regarding the true involvement of these proteins in the initiation of chromosomal DNA synthesis in mammalian cells, are still missing.

The HeLa Pur factor (pur-alpha) is another candidate as an origin-binding protein. It is a single-stranded DNA binding protein identified thanks to its ability to bind a specific element, extremely rich in purine residues, conserved in gene flanking regions and origins of DNA replication (PUR element) (Bergemann and Johnson 1992). The PUR element is present at origins of replication and in gene flanking regions in a variety of eukaryotes from yeast to humans. In mammalians the consensus sequence GGNNGAGGGAGARRRR has been identified in two regions which have been reported as the initiation zones for DNA replication in the human c-myc and hamster dhfr loci.
This protein, which is also a potent transcriptional activator, is well conserved among vertebrates and is ubiquitously expressed in human cells. Sequence homologies were detected between pur-alfa and transforming proteins of DNA-tumor-viruses (as the SV40 T antigen) as well as with others cellular proteins in yeast and human cells that may be also involved in the initiation of DNA replication (Bergemann and Johnson 1994). Furthermore, human Pur-alpha associates with the retinoblastoma protein (Rb) and this association regulates binding of pur-alpha to the PUR element (Johnson *et al.* 1995). In this case also, further information seems to be necessary in order to draw a possible role for PUR proteins in DNA replication.

1.8 THE L30E EARLY-S REPLICATED REGION

Isolation of early-S replicated sequences

Some years ago a library of sequences that replicate immediately after the onset of the S-phase was isolated from human promyelocytic HL60 cells (Tribioli *et al.* 1987).
Cells were synchronized at the G1/S border by two aphidicolin blocks and replication was allowed to restart in the presence of BrdUrd and a subinhibitory concentration of aphidicolin which allows 2% residual synthesis rate.

Elongation of the DNA strands was allowed for 70 minutes, equivalent to approximately 1.5 minutes at the physiological rate of synthesis. Subsequently, the total DNA was extracted and fractionated by CsCl gradients. Sedimentation analysis of this DNA showed that it consisted of short DNA fragments, as expected for strands coming from origins of replication activated in the early S-phase and elongating at the expected rate.
The fractions (HH) and (HL) of the CsCl gradients, corresponding to newly synthesized BrdUrd labeled DNA chains, were purified and cloned in a pAT153 vector. About 200 different sequences were obtained (67 clones from (HL) DNA and 158 from (HH) DNA) corresponding to inserts ranging from 200 to 1500 bp without any enrichment for repeated sequences but with a significant increase in Cot DNA (snap-back DNA). The two largest (HH) DNA clones (pB48 [1560 bp] and pLC46 [716 bp]) were sequenced and further analyzed.
Firstly, these sequences were checked to ascertain if they were really synthesized at the onset of the S-phase. Synchronized HL60 cells were isolated at six different time intervals in the S-phase, the nascent DNA was labeled with BrdUrd, extracted and density fractionated as described above. A slot blot analysis using the two sequences as probes, demonstrated that the two sequences were in fact present only in the DNA synthesized within 40 minutes from the beginning of the S-phase.

Structure and properties of pB48

The pB48 sequence contains a number of interesting features that can be summarized as follows (figure 1.8.1; Falaschi *et al.* 1988; Giacca *et al.* 1989; Biamonti *et al.* 1992a):

- a 600 bp long CpG-island typically present upstream of house-keeping genes
- three possible, thermodynamically stable stem-loop structure
- a sequence homologous to a nuclear matrix attachment region of mouse DNA also present in one of the loops
- several putative binding sites for transcription factors and other known DNA-binding proteins (bHLH, Sp1, topoisomerase II)
- three sequence homologous to the central palindrome in the origin of the human Papova virus JCV

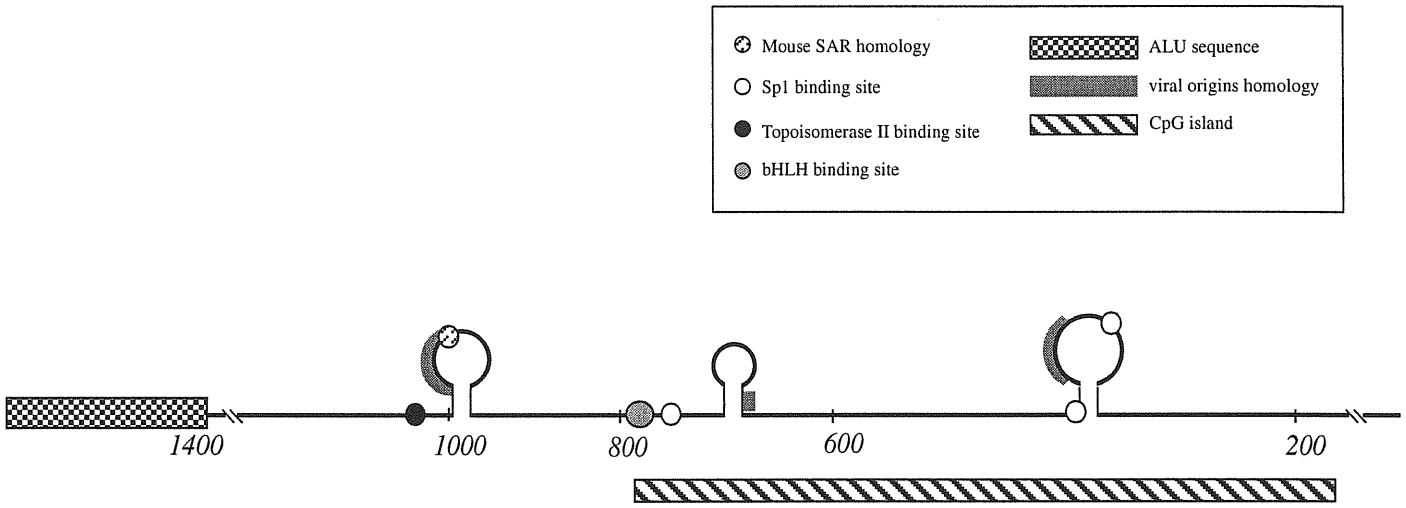


Figure 1.8.1 Features of the pB48 clone isolated from a library of sequences replicated in early S phase. The region contains a CpG island, an Alu sequence, some region of homology with viral origins of replication and several putative binding sites for known nuclear factors.

Isolation of the genomic clone L30E and chromosomal localization

A 14 Kb chromosomal clone, containing the pB48 fragment, was isolated from a human placenta DNA library (Biamonti *et al.* 1992a). Several subclones were obtained and most of the sequence was determined (Figure 1.8.2). Southern blot analysis, with L30E probes, revealed that this region is a single copy in the human genome and *in situ* hybridization showed that it is located in the short arm of chromosome 19, at the G-negative band of the subtelomeric region p13.3.

Sequence data analysis identified inside the L30E region most of the coding sequence and the 3' end of the human homologous of the mouse lamin B2 gene.

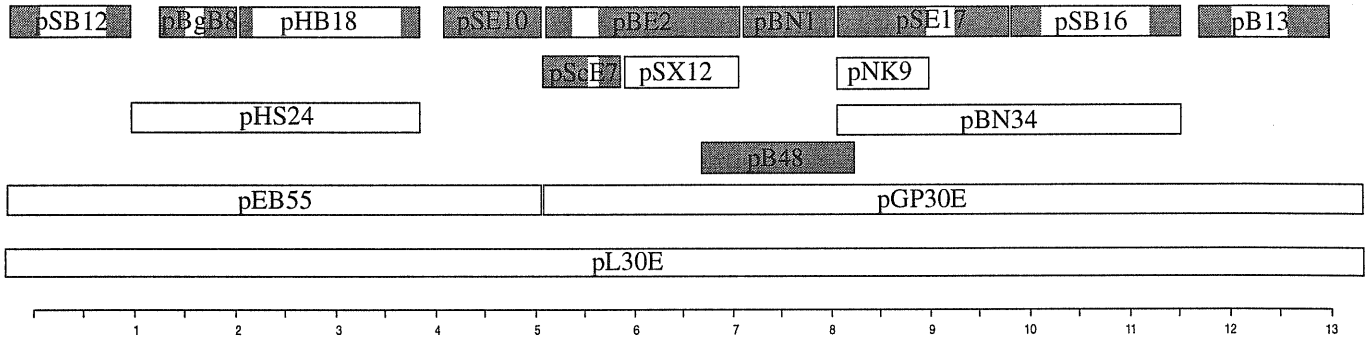


Figure 1.8.2 The 14 Kb long L30E chromosomal clone and subclones (boxed regions). Sequenced regions are coloured in grey.

Transcriptional pattern of L30E

Considering the close relationship between DNA replication and transcription generally observed both in prokaryotes and in eukaryotes, the transcriptional pattern of this region was studied in detail (Falaschi *et al.* 1988; Biamonti *et al.* 1992a; Biamonti *et al.* 1992b).

Two non overlapping transcription units, separated by 595 nt. were detected in HL60 cells. Northern blotting analysis suggest that transcription unit I contains at least two transcripts of 5000 nt. and 1450 nt. while transcription unit II contains 850 and 1150 nt. RNAs (figure 1.8.3).

Northern blots on total RNA from different cell lines, under several different conditions, gave the same results indicating that this transcriptional pattern is not restricted to HL60 cells. Further analysis also suggested that expression of all the transcripts is subject to proliferation dependent regulation, with high steady state levels of transcripts in proliferating cells and low levels in non replicating cells. In any case the transcripts remain detectable also after the induction of terminal differentiation in HL60 by retinoic acid treatment when the replication is permanently switched off (not shown).

The cDNAs of the transcripts of 5000 and 850 nt. were obtained. The 5000 nt. sequence was identified as the human homologue of the mouse lamin B2 gene, while the 850 nt. transcript (named ppv1) didn't show any homology to any known sequence.

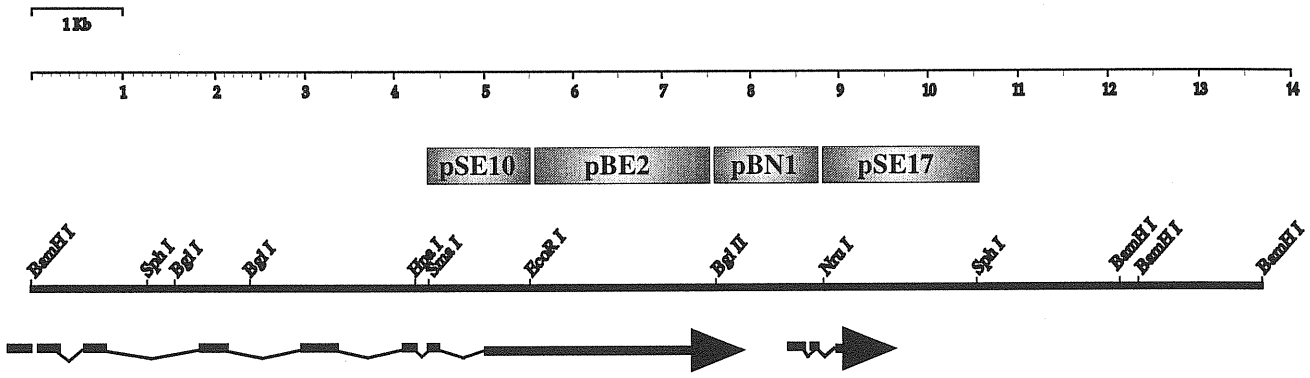


Figure 1.8.3 Schematic representation of the human genomic region analyzed in this study. The 13.7 kb region of the human chromosome 19 investigated contains the 3' end of the lamin B2 gene (left arrow) and another tandemly arranged small gene (ppv1; right arrow). Bold straight portions of the respective gene transcripts shown by arrowed lines represent exons while the vents represent introns. The gray boxes show the locations of four of the L30E subclones. The sequence of the core portion of this region (HUMLAMB2B) is available in GenBank (accession number M94363).

The aim of the research presented in this thesis is the development of a technique sensitive enough to map origins of replication in complex genomes, the identification of the DNA sequences required in *cis* for the origin activity in the lamin B2 gene locus and the identification of factors interacting with the *ori* region.

MATERIALS AND METHODS

2.1 COS-1 CELLS, TRANSFECTION AND LABELING OF NEWLY SYNTHESIZED DNA

Cells

COS-1 cells (monkey) were grown in DMEM medium (GIBCO) supplemented with 10% fetal calf serum, 2 mM L-glutamine and 50 µg/ml gentamycine in a 5% CO_2 atmosphere at 37°C

Plasmids

Three plasmids were used for transfections:

- pwtSV (~ 9 Kb) that carries the whole SV40 genome inserted in the *Bam*H I site of pAT153 (Anson *et al.* 1984).
- pSV2neo (5.6 Kb) that contains the minimal SV40 origin of DNA replication and the neomycin gene under the control of the early promoter of SV40
- pSV2neo-inv derived from pSV2neo by the inversion of the restriction fragment between the sites AccI and EcoRI containing the selective marker gene for ampicillin resistance

All these plasmids are able to replicate in COS-1 cells using the SV40 origin of replication.

Transfections and labelling

Six 10 cm tissue culture plates, containing about 10^6 COS-1 cells each, were transfected with 10µg of pwtSV by the calcium-phosphate precipitation technique. After 10 hr incubation in calcium-phosphate solution, cells were extensively washed and fresh medium was added, containing 10 nCi/ml ^{14}C-thymidine (52 mCi/mmol, Amersham, UK). After 18 hr incubation, BrdUrd (Boehringer Mannheim), 100 µM final concentration, and ^{3}H-deoxycytidine (21.5 Ci/mmol, Amersham), 1µM final concentration, were added. After 1 min. incubation, cells were killed by the addition of sodium azide and DNA was extracted.

All the subsequent steps were performed in the dark or with an orange safety light to prevent photodamage to BrdUrd-substituted DNA.

The same protocol was used for pSV2neo and pSV2neo-inv.

2.2 HL60 CELLS, SYNCHRONIZATION, LABELING OF NEWLY SYNTHESIZED DNA AND DIFFERENTIATION

Synchronized HL60 cells

HL60 (human promyelocytic leukemia) were grown in RPMI 1640 medium, supplemented with 10% fetal calf serum, 2 mM L-glutamine and 50 μg/ml gentamycine in a 5% CO_2 atmosphere at 37°C.

For the extraction of the neosynthesized DNA $2.5x10^8$ HL-60 cells were uniformly labeled for three days with ^{14}C-thymidine (52 mCi/mmol, Amersham, UK) and synchronized at the G1/S border with aphidicolin as follows. A medium containing 2μg/ml aphidicolin (Sigma Chemical Co. USA) was added to the cells. After 24 hours, cells were washed with medium without aphidicolin to remove the drug, centrifuged and resuspended in RPMI at a density of 200,000 cells/ml and incubated for additional 12 hours. Aphidicolin was then added at 5μg/ml and incubated for 12 hours.

Cells were finally collected by centrifugation, washed with fresh medium without the drug and immediately used in labelling experiments.

After release from the block, newly synthesized DNA was labeled for 10, 20, or 30 min. in the presence of subinhibitory concentration of aphidicolin (1 μg/ml), by the addition of 1 μM (final concentration) ^3H-deoxycytidine (21.5 Ci/mmol, Amersham) and 100 μM (final concentration) cold BrdUrd (Boehringer Mannheim). After the incubation times, cells were killed by the addition of sodium azide and DNA extracted.

Since synchronization causes a certain level of cell death (50-70%), the cells were layered onto Lymphocyte Separation Medium (density 1.077 g/ml, Boehringer Mannheim), centrifuged, and the band of cells at the interface collected. This treatment reduced the level of dead cells to 25-35%.

Unsynchronized HL60 cells

For the extraction of the neosynthesized DNA, $2.2x10^8$ HL60 cells in exponential phase of growth were labeled for three days with ^{14}C-thymidine (52 mCi/mmol, Amersham, UK) and then pulse labeled for 10

min. with 1μM ^{3}H-deoxycytidine (21.5 Ci/mmol, Amersham) in the presence of 100 μM cold BrdUrd (Boehringer Mannheim). At the end of the incubation, cells were killed by the addition of sodium azide and DNA extracted.

Differentiated HL60

For the induction of HL60 terminal differentiation, retinoic acid (RA, Sigma) was added to a final concentration of 10^{-6} M to a culture of $5x10^{5}$ cells/ml, followed after 24 hours by the addition of 60 mM dimethylformamide (DMF, Merck) for 72 hours (Barker *et al.* 1988). Viability was assessed by the ability of the cells to exclude trypan blue dye.

The neosynthesized DNA was labeled and extracted as described for the undifferentiated HL60 cells.

2.3 OTHER CELL LINES; LABELING OF NEWLY SYNTHESIZED DNA

HeLa cells (epithelial), were cultured in Dulbecco's modification of Eagle's minimal essential medium (DMEM, GIBCO) in monolayer culture or in Joklik's modification of Eagle's minimal essential medium (GIBCO) in spinner culture both with 10% fetal calf serum.

IMR90 (lung fibroblasts) and SKNBE (neuroblastoma) cells were cultured in Dulbecco's modified Eagle medium with 12% fetal calf serum. IMR32 (neuroblastoma) and SKNMC (neuroblastoma) cells were cultured in RPMI 1640 with 15% fetal calf serum. In all cases media were supplemented with 2 mM glutamine and 50 μg/ml gentamicin.

For the extraction of the neosynthesized DNA, 10^{8} cells were pulse-labeled for 10 min. at 37°C either with 1 μM ^{3}H-deoxycytidine (21.5 /mmol, Amersham, UK) and 100 μM cold 5'-BrdUrd (Boehringer, Germany) or with 1 μM ^{3}H-thymidine only (15.1 Ci/mmol, Amersham) as was the case in the experiments done without immunoaffinity chromatography to purify the neosynthesized DNA (IMR90, IMR32, SKNMC, SKNBE and HeLa).

2.4 EXTRACTION AND PURIFICATION OF NEWLY SYNTHESIZED DNA

DNA extraction and size fractionation

Genomic DNA was extracted according to Sambrook *et al.* (1989) from nuclei purified as described by Challberg and Kelly (1979). After heat denaturation in boiling water for 15 min. the DNA was loaded on neutral sucrose gradients for size-fractionation using concentrations below 7.5 µg/ml of gradient.

In the experiment involving transfection of plasmid pwtSV, pSV2neo and pSV2neo-inv, the extracted DNA was fractionated on four 5 to 20% (wt/vol.) linear sucrose gradients (5 ml each) for 210 min at 20°C in a Beckman SW55Ti rotor at 55 Krpm; 24 fractions of 200 µl were collected.

In the experiment with synchronized HL-60 cells, DNA was fractionated on eight 5 to 30% sucrose gradients (5 ml each) for 5 hr at 20°C in a Beckman SW55Ti rotor at 45 Krpm; 24 fractions of 200 µl were collected.

In the experiments with asynchronous cells, DNA was size fractionated on four 5 to 30% sucrose gradients (38 ml each) for 20 hr at 20°C in a Beckman SW28 rotor at 26 krpm; 12 fractions of 3 ml were collected.

The corresponding fractions from the parallel gradients of each of the above described experiments were combined; fractions were pooled as described for each experiment and dialyzed for at least 12 hours. Dialysis was carried out in TBSE (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1 mM EDTA) for those samples further purified by immunoaffinity chromatography and in TE (10 mM Tris-HCl pH 7.5, 0.1 mM EDTA) for those directly analyzed by competitive PCR.
The molecular weight of the different fractions was estimated by running parallel reference gradients with size marker DNA fragments of known sedimentation coefficients.

Immunoaffinity chromatography

Immunoaffinity chromatography was performed as described by Contreas *et al.* (1992). Goat anti-mouse IgG (Zymed Laboratories Inc. USA) was coupled to CNBr activated Sepharose 4B (Pharmacia, Sweden) according to the manufacturer's protocol (2.5 mg protein/ml gel). 0.5 ml of the coupled Sepharose was poured into a Poly-Prep 2 ml disposable chromatography column (Bio-Rad, USA) and washed with TBSE. Mouse anti-BrdUrd monoclonal antibody (Becton Dickinson, USA) was added at a final concentration of 3µg/ml and the column was closed and incubated at room temperature with slow agitation. After 2 hours of incubation the column was washed with TBSE to remove unbound antibody.

The size fractionated DNA from the various experiments, equilibrated in TBSE, was denatured for 2 min. in boiling water, chilled on ice and added to the column. After 2 hours incubation at room temperature with slow agitation, unbound DNA was allowed to drain out. One aliquot was assayed for radioactivity (^{3}H and ^{14}C) and the column was then washed with TBSE and bound ^{3}H-BrdUrd-DNA-anti-BrdUrd-mAb complexes were eluted with 150 mM NaCl adjusted to pH 11.5 with NH_4OH. The eluate was neutralized and treated with proteinase K (200 mg/ml) at 37 ^{o}C overnight. An aliquot of the eluate was assayed for radioactivity (^{3}H and ^{14}C) while the remaining was extracted with phenol:chloroform:isoamylic alcohol (25:24:1), once with chloroform:isoamylic alcohol (24:1) and precipitated with ethanol.

In order to wash out the residual DNA completely, the coupled Sepharose was washed with 5 gel volumes of 150 mM NaCl (pH 11.5 with NH_4OH) and with 5 gel volumes of 100 mM glycine/HCl pH 2.5, followed by re-equilibration with 30-40 ml of TBSE. The column was stored at 4 ^{o}C after addition of 100µl of 1% thimerosal (0.05% final, Sigma). The same goat anti-mouse IgG Sepharose can be used for at least 25-30 purifications without any loss of activity.

Results from different experiments demonstrate that purification of BrdUrd DNA by immunoaffinity chromatography is quantitative and specific, since the recovery of ^{3}H-BrdUrd DNA in the eluate is greater than 90% of the total ^{3}H-BrdUrd DNA loaded with virtually no ^{14}C-thymidine DNA present in the purified sample.

In the experiments performed with IMR90, IMR32, SKNMC, SKNBE and HeLa the step of labelling with BrdUrd was eliminated and consequently also the purification by immunoaffinity was omitted.

2.5 OLIGONUCLEOTIDE SYNTHESIS AND PCR AMPLIFICATION

Oligonucleotide primers used for amplifications were chemically synthesized with an Applied Biosystems 380B DNA synthesizer apparatus and reagents from the same company. The sequence of the oligonucleotides, the product size and the map position are indicated in the tables showed below.

A set of four primers were synthesized for each DNA region amplified. The first two (external) primers of each set were chosen in order to amplify DNA fragments of 100-300 bp with an AT content ranging from 32 to 63%; the last two primers of each set consist of two common 5' tails of 20 nt. (unrelated to genomic sequences; tail 1 : 5'-ACCTGCAGGGATCCGTCGAC-3'; tail 2 : 5'-GTCGACGGATCCCTGCAGGT-3') linked to the specific sequences reported (complementary to genomic targets) on the 3'-end; they were utilized for the construction of competitors according to the procedure reported in the section 2.6 "Construction of competitive templates".

The PCR cycle profiles were as follows: denaturation at 94°C; annealing at 52°C for the primer sets SB12, B48, SE17 and PCO, 56°C for B13, INS and SV13, 60°C for BE2, B48BIS, BN1, 62°C for SV5, SV12, SV14, and 64°C for B48TER; extension at 72°C. The time for each step was 30 sec. Between 35 and 50 cycles were performed for every reaction.

PCR reactions were performed in a buffer containing 10 mM Tris-HCl pH 8.3; 50 mM KCl; 1.5 mM $MgCl_2$; 0.2 mM each of dATP, dCTP, dGTP and dTTP; 1 µM of each oligonucleotide primer; 2.5 U of Taq polymerase (Perkin Elmer Cetus) in a final volume of 50 µl. Mineral oil was overlaid on every reaction mixture. Amplifications were carried out in a Perkin Elmer Cetus DNA thermal cycler 480. After amplification, reactions were extracted once with chloroform and analyzed in 8% polyacrylamide gels.

Table 2.5.1 Oligonucleotides used in the origin mapping on the SV40 derived plasmids

Region	Template Size (bp)	Primer name	Sequence	Position (nt.) and source
SV1	285	SV1SXX	5'-GAGAAAATACCGCATCAGGC-3'	589-608
		SV1DX	5'-TATAGTCCTGTCGGGTTTCG-3'	873-854
		SV1PERSXX	5'-tail1-GTTATCCCCTGATTCTGTGG-3'	721-702
		SV1PERDX	5'-tail2-GCAGGAAAGAACATGTGAGC-3'	722-741
				XXU02434
SV1bis	177	SV1bisSXII	5'-CCTCGCGCGTTTCGGTGATG-3'	328-347
		SV1bisDXII	5'-TACACTCCGCTATCGCTACG-3'	416-435
		SV1bisPERSX	5'-tail1-CCGGCATCCGCTTACAGACA-3'	415-396
		SV1bisPERDX	5'-tail2-GAGCAGACAAGCCCGTCAGG-3'	505-486
				XXU02434
SV2	265	SV2SX	5'-ATGCGAAACGATCCTCATCC-3'	5358-5377
		SV2DX	5'-CGGATGAATGTCAGCTACTGG-3'	5623-5603
		SV2PERSX	5'-tail1-GCGGTTTTATGGACAGCAAG-3'	5521-5502
		SV2PERDX	5'-tail2-CCAGTCTAGTCATCGCCATG-3'	5522-5541
				XXU02434
SV4	274	SV4SX	5'-TCGAAATCTCGTGATGGCAG-3'	4501-4520
		SV4DX	5'-ATCATGGTGGAAAATGGCCG-3'	4774-4755
		SV4PERSX	5'-tail1-GATTCGCAGCGCATCGCCTT-3'	4612-4593
		SV4PERDX	5'-tail2-GGGAGCGGCGATACCGTAAA-3'	4613-4632
				XXU02434
SV5	261	SV5SX	5'-TGGTGCTGGAAAACCCATTC-3'	1921-1940
		SV5DX	5'-CTGCTCATCAAGAAGCACTG-3'	2251-2232
		SV5/C-	5'-tail1-ATGAACACTGACCACAAGGC-3'	2071-2052
		SV5/C+	5'-tail2-CTGCTGACTGTCAACTGTAG-3'	2072-2091
				SV4CG
SV12	302	SV12SX	5'-CGAAGCAGTAGCAATCAACC-3'	4740-4759
		SV12DX	5'-TCATCCTGATAAAGGAGGAG-3'	5041-5022
		SV12PERSX	5'-tail1-AGGCTTCTGGGATGCAACTG-3'	4939-4920
		SV12PERDX	5'-tail2-CCAAAGTCAGGTTGATGAGC-3'	4940-4959
				SV4CG
SV13	259	SV13SX	5'-TGCTACTGTGTCTGAAGCTG-3'	597-616
		SV13DX	5'ATACCCCACTTGAGCAACAG3'	855-836
		SV13PERSX	5'-tail1-GTGAGGCCTATAGCAGCAAT-3'	749-730
		SV13PERDX	5'-tail2-TCCACAGGCCTATGCTGTGA-3'	750-769
				SV4CG
SV14	282	SV14SX	5'-ACTCCAATTCCCATAGCCAC-3'	2944-2963
		SV14DX	5'-CAGTGTGCCTAAAACACTGC-3'	3225-3206
		SV14PERSX	5'-tail1-ATTGTGGAGTGGAAAGAGAG-3'	3032-3013
		SV14PERDX	5'-tail2-TCTGCTCTGAATACTTTGAG-3'	3033-3052
				SV4CG

Note: Tail1 = 5'-ACCTGCAGGGATCCGTCGAC-3'; Tail2 = 5'-GTCGACGGATCCCTGCAGGT-3'

Table 2.5.2 Oligonucleotides used in the lamin B2 origin mapping study

Region	Template Size (bp)	Primer name	Sequence	Position (nt.) and source
SB12	109	SB12SX	5'-CACACCTTGGCCTGGTAGGT-3'	156-175
		SB12DX	5'-GAGTACGACTTCAAGATGGC-3'	264-245
		SB12+/1	5'-tail1-GTCGTGCTGGCTCCGCAGCT-3'	212-231
		SB12+/2	5'-tail2-AGCAAGTGCGGCTCTACAAG-3'	264-245 (unpublished)
SE10	139	SE10SX	5'-GTTCTGGCAGTTCGCTTAGA-3'	1368-1349
		SE10DX	5'-AGGCTGCTACGTGATGTGAA-3'	1230-1249
		SE10+/1	5'-tail1-GAAAGGTGTGTGGATGAGGA-3'	1276-1257
		SE10+/2	5'-tail2-TTTACCCAGAGCCACTGAAA-3'	1277-1296 HUMLAMB2B
BE2	193	BE2SX	5'-ACTTTCTGAAGGAGGCTCTC-3'	2500-2481
		BE2DX	5'-TCAGTCCAGCTTCCCAATGG-3'	2327-2308
		BE2+/1	5'-tail1-GCTCCCTGACCTGTGTACTG-3'	2409-2390
		BE2+/2	5'-tail2-TCACGCCAGGGCGTTGAGGA-3'	2410-2429 HUMLAMB2B
BN1	205	BN1SX	5'-GCGTATTTGGGGACAAATGC-3'	3421-3402
		BN1DX	5'-CTACATTTGCCTTGGGTGGA-3'	3217-3236
		BN1+/1	5'-tail1-ATGAACCTCCAATGACCCAC-3'	3328-3309
		BN1+/2	5'-tail2-TGAGAATTGAGTCCTTTGGA-3'	3329-3348 HUMLAMB2B
B48 BIS	180	B48 BISSX	5'-TCCGTTTTTGCAGGTTGTGC-3'	3822-3803
		B48 BISDX	5'-CCAGAATCCGATCATGCACC-3'	3643-3662
		B48 BIS+/1	5'-tail1-TAACAAAGAACTGCCGCGTG-3'	3726-3707
		B48 BIS+/2	5'-tail2-AAGATCTGAGGGACTCCTCA-3'	3727-3746 HUMLAMB2B
B48	225	B48SX	5'-TAGCTACACTAGCCAGTGACCTTTTTCC-3'	4104-4077
		B48DX	5'-AGATGCATGCCTAGCGTGTTC-3'	3880-3900
		B48+/1	5'-tail1-ACTCAGAGGCAGAACCTAAA-3'	3999-3980
		B48+/2	5'-tail2-TTATTCCTGAGGGGAAGCTC-3'	4000-4019 HUMLAMB2B
B48 TER	205	B48 TERSX	5'-GAGGACTACAATCCCCACAC-3'	4502-4483
		B48 TERDX	5'-GTGAGTACAAAGTGATCGGC-3'	4298-4317
		B48 TER+/1	5'-tail1-CACGAGCTACCCGTGGTTGC-3'	4402-4383
		B48 TER+/2	5'-tail2-TAGGTAACGGCAGGTCCAGG-3'	4403-4422 HUMLAMB2B
SE17	208	SE17SX	5'-AATTTCCACTCCACAGCCGT-3'	6462-6443
		SE17DX	5'-TAGGAGTGACCCTAGTGACT-3'	6255-6274
		SE17+/1	5'-tail1-CCACCTTTGATTGAGCCCAT-3'	6326-6307
		SE17+/2	5'-tail2-CAGACATTTATGGACACCCG-3'	6327-6346 HUMLAMB2B
B13	164	B13SX	5'-CCTCAGAACCCAGCTGTGGA-3'	54-73
		B13DX	5'-GCCAGCTGGGTGGTGATAGA-3'	217-198
		B13+/1	5'-tail1-CTCCAGAGGCGTGTTTTCCT-3'	153-172
		B13+/2	5'-tail2-ACAGCAGGTGGGTATGGGAC-3'	152-133 (unpublished)

β-GLOBIN	110	PCO4	5'-CAACTTCATCCACGTTCACC-3'	62981-62962
		PCO3	5'-ACACAACTGTGTTCACTAGC-3'	62872-62891
		PCO+/1	5'-tail1-GGAGAAGTCTGCCGTTACTG-3'	62929-62948
		PCO+/2	5'-tail2-TCAGGAGTCAGGTGCACCAT-3'	62928-62909
				HUMHBB
Insulin	141	I-SX	5'-AAGACACACAGACGGCACAG-3'	2491-2510
		I-DX	5'-GTGGCATTGTGGAACAATGC-3'	2631-2612
		I+/1	5'-tail1-CTACTGCAACTAGACGCAGC-3'	2559-2578
		I+/2	5'-tail2-TTCTCCAGCTGGTAGAGGGA-3'	2558-2539
				HUMINS01

Note: Tail1 = 5'-ACCTGCAGGGATCCGTCGAC-3'; Tail2 = 5'-GTCGACGGATCCCTGCAGGT-3'

2.6 COMPETITIVE PCR

Construction of competitive templates

Competitors were built directly from the amplification products obtained by the overlap extension method (Higuchi *et al.* 1989; Ho *et al.* 1989) and modified as indicated in figure 2.6.1.

In this case as well, a set of four primers were synthesized for each DNA region to be amplified (figure 2.6.1a). The external primers of each set were chosen in order to amplify DNA fragments of 100-300 bp. The internal primers of each set consist of two common 5' tails of 20 nt. (further details in section 2.5 "Oligonucleotide synthesis and PCR amplification") linked to the specific sequences reported (complementary to genomic targets) on the 3'-end.

Pairs of internal and external primers were used in two separate PCR reactions for the construction of two intermediate products (figure 2.6.1b and 2.6.1b'). These intermediate products were eluted from acrylamide gels, mixed, denatured and annealed (figure 2.6.1c). Subsequently, after one round of extension, the hybrid product was amplified using the external primers to obtain the competitor molecule (figure 2.6.1d). Competitors have the same sequence as the genomic targets, except for the addition of 20 nt. in the middle.

Quantitation of competitors

In the early experiments quantitation of competitors was achieved by radioactive incorporation in PCR amplifications to obtain an absolute quantitation of the molecule number (as described below, section A).

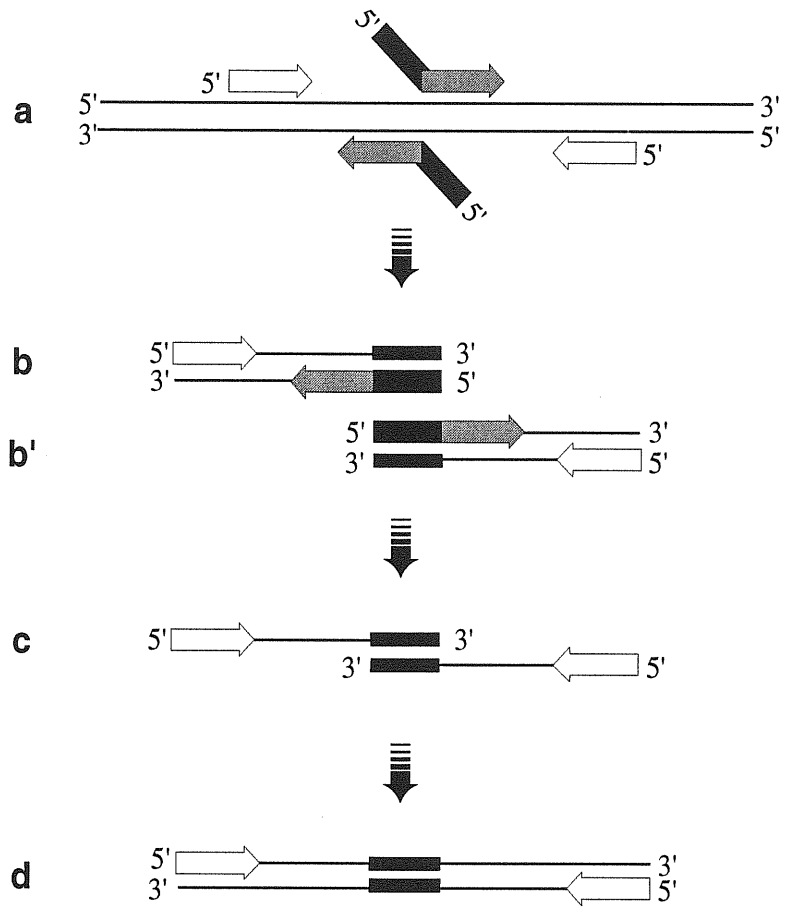


Figure 2.6.1. Construction of the competitor molecule for competitive PCR. Section 2.6 for description.

In the later experiments, for simplicity, a relative quantitation of competitors was chosen; using known amounts of plasmid DNA (for the SV40 experiments) or total genomic DNA (for the human DNA experiments) in co-amplification reactions with the competitors to quantify the molecule number (as described below, section B).

section A) For that which concerns the radioactive quantitation, a small amount of competitor DNA sample was taken from acrylamide gel by touching the competitor band with the tip of a needle. The tip was then immersed in 100 µl of distilled water and 1 µl of this solution was used as template in a PCR amplification in the presence of 10 nmoles of cold dCTP and 0.66 pmoles of ^{32}P-labeled dCTP (3000 Ci/mmole, 10 µCi/ml; Amersham, UK). The amplification product was eluted from polyacrylamide gel in water and a small aliquot was counted to evaluate the concentration of competitor in accordance with the expected specific activity of labelling and the number of dCTP nucleotides incorporated from each competitor.

section B) A precise quantitation of competitors was also obtained by the addition, in different PCR reactions, of increasing quantities of competitor to a fixed amount of plasmid (or total genomic) DNA. After amplification with the appropriate primer sets, the reactions were resolved by non-denaturing 8% polyacrylamide gel electrophoresis, the gel was stained with ethidium bromide and the intensity of the bands quantified by scanning densitometry.
Finding the reaction in which the ratio between the competitor band and the template band is 1, means to find the condition in which, before the amplification, the same amount of template and competitor molecules were present in the reaction. As a consequence, it is possible to know the precise number of plasmid/genomic DNA molecules used for every reaction by measuring the concentration of any preparation of competitor (more details are reported in section 3.1).
All the quantitations reported in this thesis represent an average of at least three different measurements.

Quantitation of BrdUrd DNA by competitive PCR
The quantitation on the nascent DNA was obtained as described for the relative quantitation of the competitors; using increasing amounts of competitors and fixed amounts of neosynthesized DNA.
All the preparations of nascent DNA were dialyzed extensively (2-3 days) in TE before being used for the PCR reactions.

2.7 CELLULAR AND NUCLEAR EXTRACTS

Several kinds of cellular and nuclear extracts were used for band-shift and DNaseI-footprinting analysis or for protein purification purposes.
In all cases, the cells (whether fresh or frozen) were collected by centrifugation, washed once with cold phosphate-buffered saline (PBS), resuspended in 5 pellet volumes of hypotonic buffer (buffer A: 10 mM Hepes-KOH pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT, 0.5 mM PMSF) and allowed to stand for 10 min. on ice. Subsequently cells were centrifuged for 10 min. at 1,000 g, resuspended in other 2 volumes of buffer A and lysed by 10 strokes in a glass-glass Dounce homogenizer (pestle B). After a 10 min. of centrifugation at 1,000 g, 0.11 volumes of buffer B (0.3 mM Hepes-KOH pH 7.9, 30 mM $MgCl_2$, 1.4 M KCl) was added to the supernatant and again centrifuged at 100,000 g for 1 hour to obtain the **S-100 cytoplasmic extract**, while the nuclei contained in the pellet of the 1,000 g centrifugation were differently extracted to obtain different kinds nuclear extracts.

The nuclei obtained in this way were extracted with buffer C (20 mM Hepes-KOH pH 7.9, 25% glycerol V/V, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 10 mM KCl, 1 mM DTT, 1 mM PMSF, 1 μM leupeptin, 1 μM pepstatin A + KCl or NaCl). Using 0.42 M KCl or 0.42M NaCl, the nuclear extract is termed **Dignam[K] extract** or **Dignam[Na] extract** respectively (Dignam *et al.* 1983). In other cases different concentration of NaCl were used during the extraction with buffer C to obtain the **0.4 M, 0.8 M** and the **1.2 M extracts**. The pellet of nuclear debris obtained after the Dignam extraction was further extracted using buffer C containing 1M KCl and the non ionic detergent NP-40 (1% final) to obtain the **1+1 extract**. All the extracts were centrifuged for 30 min. at 25,000 g to remove the particulate material and were dialyzed in dialysis buffer 1 (20 mM Hepes-KOH pH 7.9, 10% glycerol V/V, 0.5 mM $MgCl_2$, 0.2 mM EDTA, 100 mM KCl, 0.5 mM DTT, 0.5 mM PMSF, 1 μM leupeptin, 1 μM pepstatin).

2.8 GEL RETARDATION ASSAYS AND SUPERSHIFTS

Oligonucleotides and plasmids

The following oligonucleotides were synthesized to perform the gel retardation assays:

USF-BS up: 5'-CCGGTCGCATCACGTGACGAAGAG-3'
USF-BS down: 5'-CTCTTCGTCACGTGATGCGACCGG-3'

USF-BS up: 5'-CTGACTCTTCGTCACGTGATGCGACCGGGCTCCGCCCCGG-3'
USF-BS down: 5'-GACTGAGAAGCAGTGCACTACGCTGGCCCGAGGCGGGGCC-3'

NRF1-BS up: 5'-TGCCTCGTGCGCATGCGCGGCAGGC-3'
NRF1-BS down: 5'-GCCTGCCGCGCATGCGCACGAGGCA-3'

hOCBS-75 up: 5'-AGCTTAATATACATTTTATGACTGGAAACTTTTTTGTACAACACTCCAATA AACATTTTGATTTTAGGTTCTGCC-3'
hOCBS-75 down: 5'-TCGAGGCAGAACCTAAAATCAAAATGTTTATTGGAGTGTTGTACAAAAAAG TTTCCAGTCATAAAATGTATATTA-3'

SN1: 5'-GATCTGAATATACATTTTATGACTGGAAACTTTTTTGTACAACACTCC-3'
SN2: 5'-GATCGGAGTGTTGTACAAAAAAGTTTCCAGTCATAAAATGTATATTAC-3'

Some probes were also obtained from DNA-binding sites appropriately cloned in plasmids:

- pCR-B48 280: this plasmid derives from the plasmid pCR II^{TM} (Stratagene, USA) by cloning the PCR-amplification product (3795-4074 HUMLAMB2B). The hOCBS-280 probe was obtained digesting the plasmid in its polylinker with SpeI and EcoRV.
- pCR-B48 120: this plasmid derives from the plasmid pCR II^{TM} by cloning the PCR-amplification product (3880-4003 HUMLAMB2B). The hOCBS-120 probe was obtained digesting the plasmid as reported for pCR-B48 280.
- pCR-4SR: this plasmid derives from the plasmid pCR II^{TM} by cloning the PCR-amplification product (4112-4408 HUMLAMB2B). The 4SR probe was obtained by digestion with EcoRI.
- pCR-B48mut40: was obtained from the plasmid pCR II^{TM} by cloning the corresponding region used to produce pCR-B48 280 (3795-4074 HUMLAMB2B) with a 40 bp mutation in the region 3917-3956

obtained by PCR using two internal mutated oligonucleotides. The sequence of the mutation is a permutation of the original 40bp; 5'-TTATTGTTGTCGGAATTCCATTTAAAAGTACTAAATTTTT-3' (coding strand).

Probes labelling

The oligonucleotide probes for gel retardation assay (USF-BS, NRF1-BS, SN, hOCBS-75) were labeled at their 5' end with γ^{32}P-ATP (Amersham, U.K.; 3000 Ci/mmol; 10 mCi/ml) using polynucleotide kinase (New England Biolabs, USA) according to the indications of the manufacturers. After ethanol precipitation, probes were mixed with the complementary oligo, denatured 10 min. in boiling water and allowed to anneal by gradual cooling to room temperature (annealing buffer: 60 mM TRIS-HCl pH 8, 12 mM $MgCl_2$, 1 mM EDTA).

In the case of probes obtained by restriction digestion of cloned sequences (hOCBS-280 and hOCBS-120), plasmids were digested with the first restriction enzyme (5' protruding), labeled with α^{32}P-dATP (Amersham, U.K.; 3000 Ci/mmole; 10 mCi/ml) using DNA-polymerase I Klenow fragment (New England Biolabs, USA) and cut again with the second restriction enzyme to generate the particular probe.
All the labeled probes were purified from 8% polyacrylamide gels by cutting the appropriate radioactive band, eluting the DNA O/N at RT in elution buffer (0.5 M CH_3COONH_4, 1 mM EDTA) and then precipitating the probes in ethanol.

Binding reactions and running conditions

The protein-DNA binding conditions used for the putative binding sites of known transcription factors was taken from the literature (Csordas *et al.* 1993 for USF and Evans and Scarpulla 1990 for NRF1) while for GST-HOX C13 the optimal binding requisites were determined after several trials under different conditions.

The reaction mixture for the NRF1 binding assay consist of a binding buffer containing 25 mM Tris-HCl pH 7.5, 100 mM KCl, 1 mM EDTA, 0.5 mM DTT, 0.01% Triton X-100 in presence of 1000ng of non specific competitor such as poly d(A-T) and various amounts of specific

competitors as well as protein extract. For every reaction 1ng of NRF1 probe (10000 dpm) was used. After 30 min. incubation at RT the DNA-protein complexes were resolved at room temperature by high-ionic strength (TGB 1X; Sambrook *et al.* 1989) 4% polyacrylamide gels run at 5 volt/cm for 2 hours after 1 our of prerunning.

The reaction mixture for the USF binding assay consist of a binding buffer containing 20 mM Hepes-NaOH pH 7.3, 50 mM NaCl, 0.2 mM EDTA, 4 mM DTT, 5% Glycerol in presence of 1000 ng of non specific competitor such as poly d(I-C) and various amounts of specific competitors as well as protein extract. For every reaction 1 ng of USF probe (10000 dpm) was used. After 30 min. incubation at RT the DNA-protein complexes were resolved at room temperature by low-ionic strength (TBE 0.5X; Sambrook *et al.* 1989) 5% polyacrylamide gels run at 5 volt/cm for 2 hours after 1 our of prerunning.

For GST-HOX C13 the best binding buffer contained 20 mM Tris-HCl pH 8, 100 mM KCl, 0.1 mM EDTA, 4 mM $MgCl_2$, 0.2 mM $ZnCl_2$, 4 mM spermidine, 0.25 mM DTT, 0.1 mg/ml BSA and 10% Glycerol. All the reaction were carried out for 1 hour on ice using 1 ng of hOCBS-75 probe (10000 dpm) in the presence of 3000 ng poly d(I-C) used as aspecific competitor and 10 ng of GST-fusion HOX C13 protein. The DNA-protein complexes were resolved by low-ionic strength (TBE 0.5X; Sambrook *et al.* 1989) 10% polyacrylamide gels run at 5 volt/cm for 2 hours, at 4 ^{O}C, after 1 our of prerunning The same conditions were also used for His-HOX C10.

Super-shift assay

Gel retardation super-shifts were obtained for the NRF1-BS probe by further incubating the binding reaction mixtures (prepared as described in "Binding reactions and running conditions") for additional 30 min. with a polyclonal anti-NRF1-serum (gift of R. C. Scarpulla) and resolving the complexes as described for the gel retardation assay.

2.9 DNASE I *IN VITRO* FOOTPRINTING ASSAYS

Probes and labelling

The probes used for the *in vitro* DNaseI footprinting assay were obtained from regions of the lamin B2 origin cloned in plasmids:

- hOCBS-upper was obtained from the plasmid pCR-B48 280 by labelling the SpeI site of the SpeI-EcoRV restriction fragment with α^{32}P-dATP (Amersham, U.K.; 3000 Ci/mmole; 10 mCi/ml) and DNA-Polymerase I Klenow fragment (N.E.B. USA).
- hOCBS-lower was obtained from the plasmid pCR-B48 280 by labelling the XbaI site of the XbaI-BamHI restriction fragment with α^{32}P-dATP and DNA-Polymerase I Klenow fragment.
- 4SR-lower was obtained by labelling the SpeI site of the SpeI-EcoRV restriction fragment from the plasmid pCR-4SR with DNA-Polymerase I Klenow fragment. The pCR-4SR plasmid, on the other hand, derived from the pCR II^{TM} (Stratagene, USA) by cloning the PCR-amplification product (4112-4408 HUMLAMB2B).
- hOCmut-lower was obtained from the plasmid pCR-B48mut40 by labelling with α^{32}P-dATP the SpeI site of the SpeI-EcoRV restriction fragment.

Binding reactions

To identify protein-DNA interactions *in vitro* different binding conditions were used. In the case of the ppv1 promoter region (4SR-lower probe) binding buffer F (20 mM Hepes-KOH pH 7.9, 30 mM NaCl, 0.1 mM EDTA, 6 mM $MgCl_2$, 0.2 mM $ZnCl_2$, 2 mM DTT, 0.1 mg/ml BSA and 5% Glycerol) was used . The same buffer was also used for the hOC binding site (hOCBS-lower and hOCBS-upper probes) adding to the reaction ribonucleotides and deoxyribonucleotides as reported for the DNA binding of yeast Origin Recognition Complex by Bell and Stillman, 1992 (4 mM ATP, 0.1 mM CTP/GTP/UTP, 0.06 mM dCTP/dGTP/dTTP, 0,01 mM dATP).

For GST-HOX C13 we used the same conditions reported for band shift assay slightly modified by use of higher concentrations of $MgCl_2$ (20 mM) and $ZnCl_2$ (0.5 mM).

All the binding reactions were carried out at RT for 20 min. followed by

a 1 min. digestion with 0.5 units of DNaseI (Promega, USA) in DNaseI buffer (final concentration: 4 mM Tris-HCl pH 8, 5 mM NaCl, 4 mM $MgCl_2$, 0.4 mM $CaCl_2$). The reactions were stopped by addition of one volume of Stop Solution (50 mM EDTA and 2% SDS), extracted two times with phenol:chloroform:isoamilic alcohol (25:24:1), ethanol precipitated, resuspended in 4 µl of loading buffer for sequencing gels (Sambrook *et al.* 1989), denatured and resolved on 6% denaturing polyacrylamide gels. A G+A chemical cleavage sequence ladder was obtained from the same fragment as described by Maxam and Gilbert, 1980.

2.10 PROTEIN PURIFICATION

DNase I footprinting was used as monitoring assay in protein purification. Unfortunately, due to the characteristic of the assay, a clear quantitation of the DNA-binding activity is not possible also if the minimal number of molecules present in the extract can be roughly estimated measuring the lowest amount of extract able to give a total protection of the probe. In this situation the number of protein molecules present cannot be lower to the number of DNA-probe molecules.

Protease inhibitors (1 mM PMSF, 1 mM sodium metabisulfite, 1 µM pepstatin, 1 µM leupeptin, and 1 mM DTT) were added to all the buffers immediately before use. All procedures for protein purification were performed at 4°C.

Protein elution from all the columns was monitored by continuous UV absorption at 280 nm. Protein contents of different column fractions were analyzed on 12% SDS polyacrylamide gels stained with Coomassie Brilliant Blue or BioRad silver reagent (BioRad, USA). Protein concentrations were determined by the Bradford assay using a BioRad protein assay reagent.

Ammonium sulfate precipitation

The crude nuclear extracts were differentially precipitated adding different amounts of an ice-cold saturated ammonium sulfate solution (697 g/l) adjusted to pH 8 with NH_4OH. Each protein pellet was resuspended in a small volume of dialysis buffer 1 (20 mM Hepes-KOH pH 7.9, 10% glycerol V/V, 0.5 mM $MgCl_2$, 0.2 mM EDTA, 100 mM KCl, 0.5

mM DTT, 0.5 mM PMSF, 1 μM leupeptin, 1 μM pepstatin) and dialyzed against the same buffer for 24 hours at 4 °C with three changes.

BioRex 70 column

Cation exchange BioRex 70 resin (BioRad, USA), equilibrated at 4 °C in dialysis buffer 1 (20 mM Hepes-KOH pH 7.9, 10% glycerol V/V, 0.5 mM $MgCl_2$, 0.2 mM EDTA, 100 mM KCl, 0.5 mM DTT, 0.5 mM PMSF, 1 μM leupeptin, 1 μM pepstatin), was run according to the conditions recommended by the manufacturers loading no more than 3 mg of protein extract/ml of wet resin. The eluted fractions were collected in pools and dialyzed against dialysis buffer 2 (20 mM Tris-HCl pH 8, 10% glycerol V/V, 0.5 mM $MgCl_2$, 0.2 mM EDTA, 60 mM KCl, 0.5 mM DTT, 0.5 mM PMSF, 1 μM leupeptin, 1 μM pepstatin).

Q-Sepharose column

Anion exchange Q-Sepharose column (Pharmacia, Sweden), equilibrated at 4 °C in dialysis buffer 2, was run according to the conditions recommended by the manufacturers loading no more than 0.5 mg of protein extract/ml of wet resin. The eluted fractions were collected in pools and dialyzed against dialysis buffer 1.

DNA-affinity purification

The 280 bp PCR-amplification product (3795-4074 HUMLAMB2B) biotinilated at one of the extremities was coupled to avidin coated magnetic beads (Dynal) according to the manufacturer's protocol. The beads were incubated for 1 hour at 4 °C with the protein extracts using the same conditions described for the DNase I *in vitro* footprinting. After two washes with the same buffer, bound proteins were eluted by several washes in binding buffer with increasing concentration of KCl.

2.11 GENOMIC FOOTPRINTING

Isolation of Nuclei and DNA from HL60 Cells

HL60 cultured cells were collected by centrifugation and washed once with cold phosphate-buffered saline (PBS). These were resuspended in standard reticulocyte buffer (RSB: 10 mM Tris/HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$) at $25x10^6$ cells/ml and incubated on ice for 5 min.

Subsequently an equal volume of 0.2% Nonidet P-40 in RSB was added and the cell suspension was incubated on ice for additional 3-5 min. after gentle mixing. Nuclei were recovered by centrifugation, washed once with RSB and resuspended, in the same buffer, at $25x10^6$ nuclei/ml. DNA was isolated as described (Miller *et al.* 1988).

Nuclease Treatment of HL60 Nuclei

Aliquots of the isolated nuclei were adjusted to a final concentration of 25 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$ and 1 mM $CaCl_2$. DNase I (Boehringer, Germany) was added to the nuclei (1-10 units/ml) and digestion was allowed to proceed for 10 min. on ice. The reactions were stopped by the direct addition of lysis buffer (50 mM Tris/HCl, pH 8, 0.4 M NaCl, 50 mM EDTA, 1% SDS, 600 µg/ml Proteinase K, final concentrations) and DNA was extracted by the same procedure as for untreated nuclei.

Nuclease Treatment of Naked DNA

In vitro reactions contained 50-100 µg of purified HL60 DNA in 1 ml of 1x DNase I buffer and different concentrations (0.05-0.5 units) of DNase I. Reactions were stopped by the addition of EDTA to 20 mM, samples were extracted with phenol/chloroform and DNA was precipitated with ethanol.

DMS Treatment of HL60 Cells

Aliquots of cells were resuspended in RPMI medium at $2.5x10^7$ cells/ml and treated with increasing concentrations of DMS (0.05-0.5%) for 2-10 min. at room temperature. The reactions were stopped by the addition of 10 volumes of cold PBS, cells were collected by centrifugation, washed once with cold PBS and DNA was prepared directly from cells or from nuclei prepared as described above.

***In vitro* Methylation of HL60 DNA**

Purified HL60 DNA (50-100 µg) was resuspended in 100 µl of DMS buffer (50 mM sodium cacodylate, pH 8, 0.1 mM EDTA) and 1 µl of DMS was added for 1-5 min. at room temperature. Reactions were stopped by the addition of 35 ml of cold DMS stop buffer (1.5 M sodium acetate, pH 7, 1

M β-mercaptoethanol) and DNA was precipitated with ethanol. Piperidine cleavage of methylated DNA, as well as the G+A and the C+T sequencing reactions were performed as described (Saluz and Jost 1987; Ausubel *et al.* 1994).

Oligonucleotide Primers and LMPCR

The nucleotide sequences and the amplification conditions for each primer are described in the table reported below. Primer 1 from each set is used in the initial primer extension reaction, primer 2 - in the amplification step, and the radioactively labeled primer 3 - in the final linear amplification reaction. The LMPCR reactions were performed with 5 µg of DNA (methylated or DNase I-digested, both *in vivo* or *in vitro*) from HL60 cells as described (Ausubel *et al.* 1994).

set	primer	length [bp]	sequence	genomic position	annealing temperature [°C]
A	1	25	5'-TTACCTACACGAGCTACCCGTGGTT-3'	4385-4409	60
	2	22	5'-TGGTTGCGACTCCGCGGGAAGA-3'	4368-4389	68
	3	27	5'-TCCGCGGGAAGAGGGAGGCCCTGACTT-3'	4353-4379	76
B	1	28	5'-GGCTAGTGTAGCTAGTGTAAACAGGACC-3'	4091-4118	60
	2	25	5'-GTAAACAGGACCCAGGCGATGCATG-3'	4107-4131	67
	3	27	5'-CAGGACCCAGGCGATGCATGGGACCCT-3'	4112-4138	76
C	1	25	5'-TCGCATCACGTGACGAAGAGTCAGC-3'	4179-4203	60
	2	25	5'-GAGTCAGCTTGTGCAACAGCGTCGG-3'	4162-4186	66
	3	27	5'-GCTTGTGCAACAGCGTCGGAGGCTCAC-3'	4154-4180	72
D	1	25	5'-GTCACAGCACAACCTGCAAAAACGG-3'	3795-3821	60
	2	25	5'-CAAAAACGGAGCTGGGCTGCAGCTG-3'	3813-3837	70
	3	26	5'-GGGCTGCAGCTGGGGCTGGCATGGAC-3'	3826-3851	76
E	1	25	5'-GGGGTGGAGGGATCTTTCTTAGACA-3'	4049-4074	60
	2	27	5'-GACATCCGCTTCATTAGGGCAGAGGCC-3'	4026-4052	68
	3	27	5'-TCATTAGGGCAGAGGCCCGGCTCGAGC-3'	4016-4042	76
SL*		11	5'-GAATTCAGATC-3'		
LL		25	5'-GCGGTGACCCGGGAGATCTGAATTC-3'		68

*SL and LL are the short and long linker primers, respectively

2.12 ONE-HYBRID SCREENING

The screening was performed on a cDNA-fusion expression library of HeLa cells in which the proteins of the library are fused to the end of a constitutively expressed GAL4 activation domain.

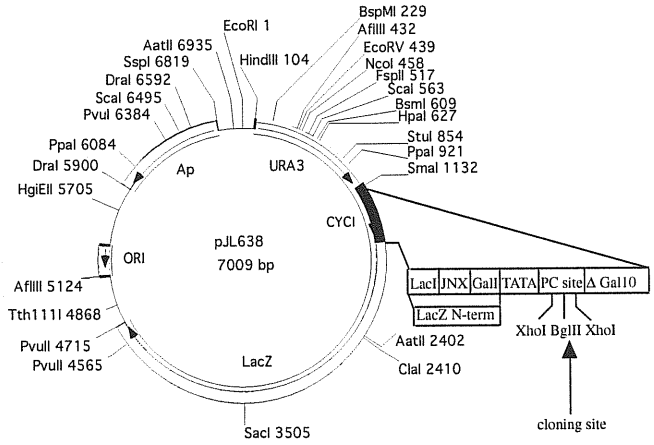


Figure 2.12.1 Plasmid pJL638d. This plasmid was used as a vector for a one-hybrid screening in yeast to identify the c-DNA of proteins able to interact with the hOC region of the human lamin B2 gene origin.

As reporter vector the pJL638 plasmid was used (figure 2.12.1). This vector contain the ampicillin resistance gene as selectable marker in bacteria, the URA3 gene as selectable marker in yeast and a UAS-less LacZ gene as reporter downstream to the cloning site for the target sequence.

The target sequence used for the screening was the head to tail trimer of the 73 bp sequence of the hOC binding site (3921-3993 HUMLAMB2B) cloned in the BglII site of pJL638. K699 was the yeast strain used for the assay. The protocol used for the screening was developed by Li end Herskowitz (1993).

The plasmid pGEX-2T (Pharmacia, Sweden) was used to produce the GST-fusion protein GST-HOX C13 according to the indications of the manufacturers while pTrcHis A (Invitrogen, The Netherlands) was used to produce the histidine-tag-HOX C10 fusion protein.

2.13 PRODUCTION OF POLYCLONAL ANTIBODIES

Polyclonal antibodies against HOX C13 were produced inoculating two rabbits with 200 μg of GST-tagged protein in complete Freund's adjuvant followed after tree weeks by four successive boosts of 150 μg

in incomplete Freund's adjuvant with the same time interval one from another. The specific response of the rabbits against the antigen was tested by western blotting assay that was also used to determine the antibody titer using different antiserum dilutions.

Western blotting

Proteins were separated by SDS-PAGE and transferred onto nitrocellulose filter by electroblotting for 15 hours. The filter was incubated in 10% milk-TBS buffer (10% w/v non-fat dried milk in 125 mM NaCl, 10 mM Tris-HCl, pH 7.4) for 1 hr at 37°C. Incubation with the appropriate rabbit antibody was performed in 5% milk-TBS for 2 hours at room temperature. The filter was then washed in TBS-0.1% Tween 20 and incubated in 5% milk-TBS with alkaline phosphatase-conjugated goat anti-rabbit immunoglobulins diluted 1:2000 for 1 hr at room temperature. After several washes, bound antibodies were revealed using BCIP/NBT color development solution (BioRad).

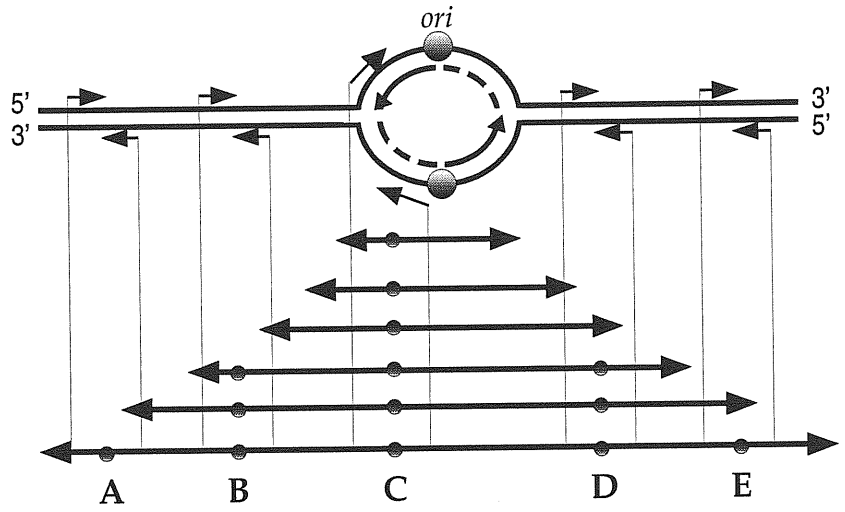
RESULTS

3.1 A NOVEL METHOD TO MAP ORIGINS OF DNA REPLICATION

In the absence of a functional assay, the first step to identify the DNA sequences required in *cis* during the initiation of the DNA replication is the development of a technique that is able to map the origin activity in the genome of interest.

As discussed in the introduction, several methods were developed to map origin activity in several laboratories but none of them were suitable to easily map, with a good resolution, an origin contained in a non-amplified region of a mammalian genome. For this purpose we devised a new mapping technique based on a modification of the "PCR of nascent DNA strand" technique originally described by Vassiliev (as described the Introduction section 1.6; Vassilev *et al.* 1990a; Vassilev *et al.* 1990b).

This method make use of competitive PCR to map the relative abundance of various sets of DNA markers in the nascent DNA pool. As shown in Figure 3.1.1, an origin of bidirectional DNA replication in a cell culture chronically labeled with ^{14}C-thymidine is pulse labeled with ^{3}H-deoxycytidine and cold BrdUrd. The pulse labeled DNA is isolated, denatured and fractionated on neutral sucrose density gradients. The density gradient separation serves to isolate the nascent DNA, rich in origin-sequences, from bulk DNA since. These fragments derive from the origins recently activated, are very short. BrdUrd substituted nascent DNA fraction is then purified from the fragmented (unreplicated) DNA through a column containing an immobilized monoclonal antibody against BrdUrd. This process yields more than 95% of ^{3}H-labeled DNA and practically none with ^{14}C-label (Contreas *et al.* 1992). These fragments are subsequently assayed for the abundance of selected markers with quantitative PCR, in which the relative abundance of a marker in the nascent DNA is directly proportional to its closeness to the origin (figure 3.1.1).



BrdUrd LABELING OF NASCENT DNA IN GROWING CELLS

DNA EXTRACTION AND SIZE FRACTIONATION BY NEUTRAL SUCROSE GRADIENTS

PURIFICATION OF THE BrdUrd-LABELED DNA BY IMMUNOAFFINITY CHROMATOGRAPHY WITH ANTI-BrdUrd ANTIBODIES

QUANTITATION OF THE NASCENT DNA BY COMPETITIVE PCR USING DIFFERENT PRIMER PAIRS AND COMPETITORS

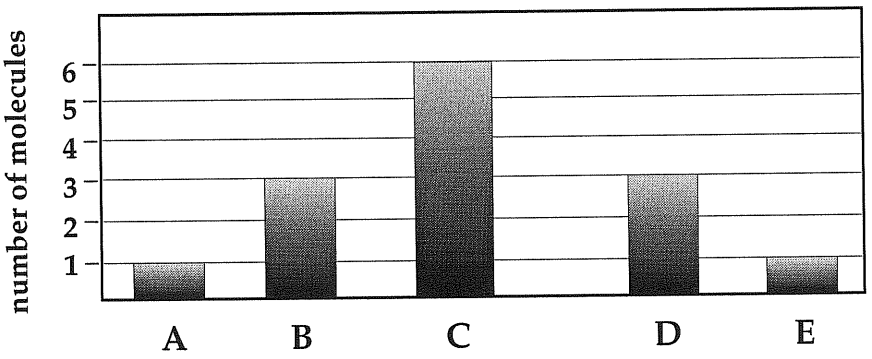


Figure 3.1.1 The novel ori-mapping procedure based on nascent DNA strand analysis with quantitative-PCR. Short fragments of nascent DNA, synthesized in the presence of BrdUrd, can be isolated and separated from the bulk DNA on a neutral sucrose gradient. Further purification of this DNA can be done by immunoaffinity chromatography with anti-BrdUrd antibodies. Within this population of newly synthesized DNA molecules, the abundance of selected fragments (A through E) scattered within a genomic region can be precisely known by quantitative-PCR. The pair of primers (small arrows on the topmost figure) amplifying the highest number of molecules in the sample, is closest to the origin (marker C).

The quantitative PCR method devised in our lab takes advantage of the addition of a competitor molecule to the amplification reaction (construction and quantification of the competitor molecules are reported in section 2.6). In this way, any kind of predictable or unpredictable variations affecting the reaction for a given template would affect the competitor in the same way. This because the competitor has the same sequence of the template except for a negligible insertion of 20 base pairs. Such an insertion makes also the competitor molecule distinguishable from the template molecule when resolved on a polyacrylamide gel.
The amount of specific template molecules present in the sample can be easily calculated coamplifying a fixed amount of sample with known amounts of competitor since the template/competitor ratio remains constant throughout the reaction (figure 3.1.2).

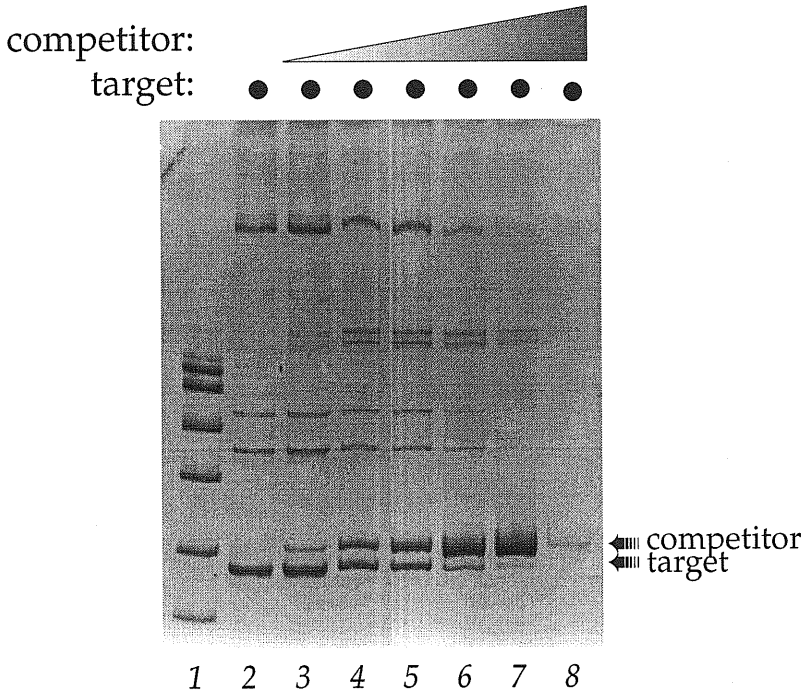


Figure 3.1.2 Example of competitive PCR. Coamplification of a fixed amount of target molecules in the presence of known amounts of competitor DNA. The reaction in which the target band has the same intensity as the competitor band represents a situation of equivalence. The equivalence can be reached when, in the starting reaction, the number of target molecules is equal to the number of competitor molecules (known). Lane 1, molecular weight marker; lanes 2-8, fixed amount of target DNA in the presence of increasing amounts of competitor.

Quantitation of nascent DNA by competitive PCR is easy and reproducible. In this way, to map an origin, it is possible avoid the use of metabolic inhibitors or cell-synchronization procedures and origins can be detected through single copy sequences in complex genomes.

3.2 COMPETITIVE PCR ON NASCENT DNA CAN BE USED TO MAP A KNOWN VIRAL ORIGINS OF REPLICATION

The validity of the above described method in mapping an origin efficiently was ascertained by its use to map an already well known origin, that of the SV40. The experiment was performed using the plasmid pwtSV. This plasmid contains the whole genome of SV40 virus and is able to replicate as an episome when transfected in COS cells. Figure 3.2.1 shows the flow chart of the mapping experiment.

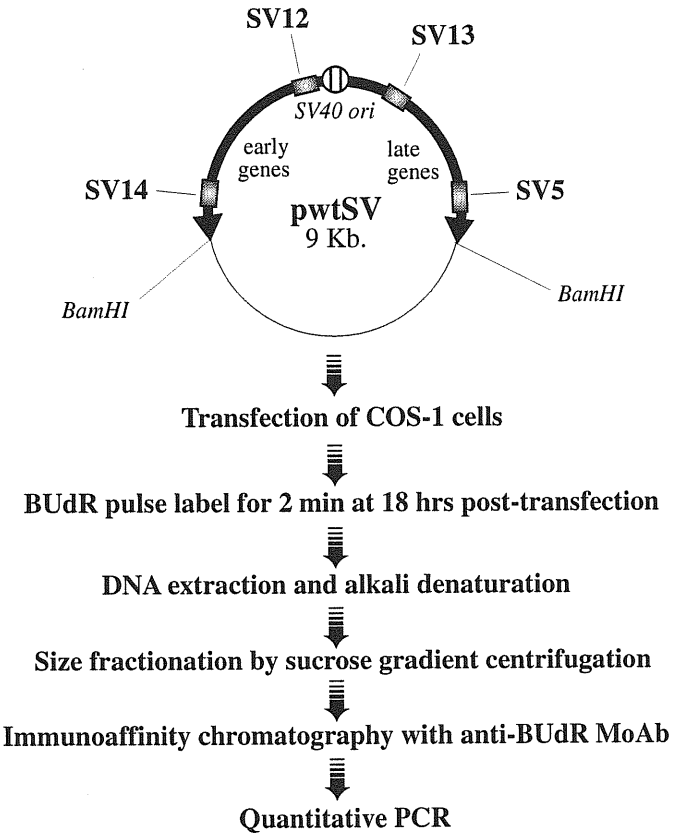
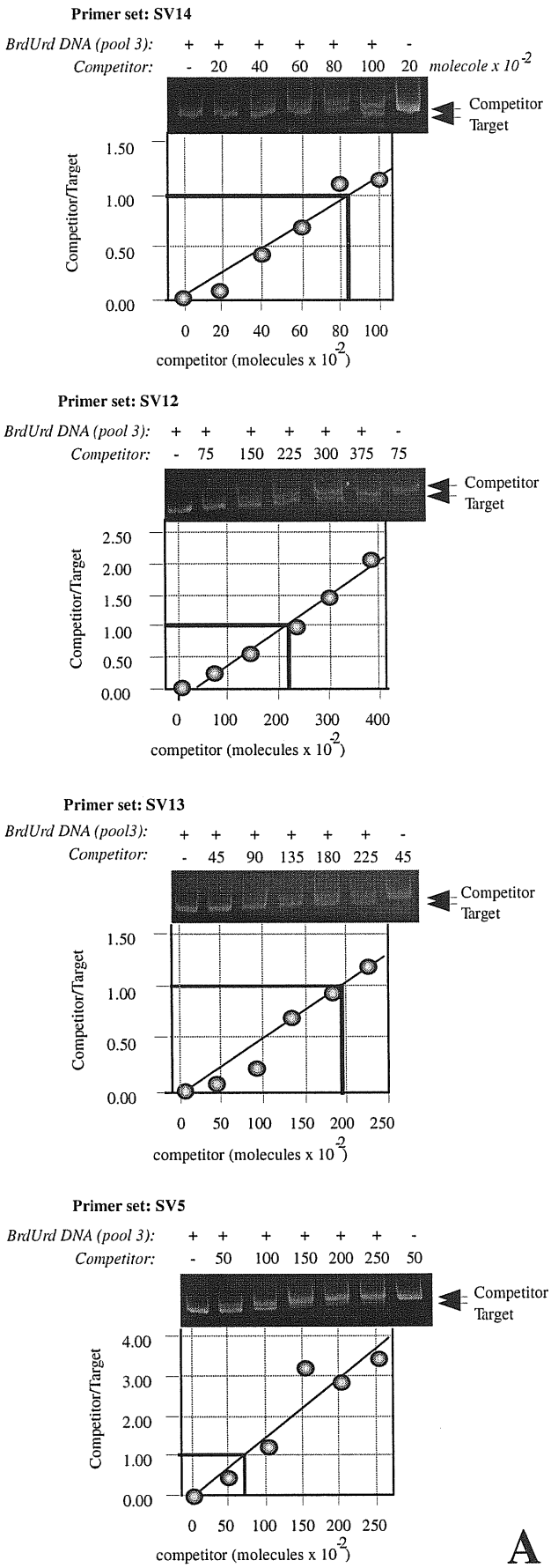


Figure 3.2.1 Mapping of the SV40 origin of DNA replication. Flow chart and position of the regions amplified by the various primer sets in pwtSV.

Six 10 cm tissue-culture plates, containing 10^6 COS-1 cells each, were transfected with 10 µg of plasmid by the Ca-phosphate precipitation technique. After 10 hr. incubation in Ca-phosphate solution, cells were extensively washed and fresh medium added, containing 10 nCi/ml [^{14}C]thymidine. After 18 hr. incubation, BrdUrd, 100 µM final concentration, and [^{3}H]-deoxycytidine, 1 µM final concentration, were added. After 1 min. incubation, cells were killed by the addition of sodium azide; total DNA was extracted, divided in four aliquots (700 µl final volume), alkali-denatured and fractionated by sedimentation through four parallel 5 ml neutral 5 to 20% (wt/vol.) linear sucrose gradients for 210 min. at room temperature in a Beckman SW55Ti rotor at 55 krpm.
Fractions of 200 µl each were collected from the top; corresponding fractions from the four identical gradients were combined. Size distribution of nascent DNA was determined by agarose gel electrophoresis of DNA molecular weight markers run in parallel gradients. Pools of 2-4 fractions were dialyzed against TBSE and concentrated down to an appropriate volume (1.5 to 4 ml) with polyethylenglycol 6000.

BrdUrd-substituted DNA was further purified from three fraction pools (average length 700, 1400, and 2300 nt., respectively) by immunoaffinity chromatography with anti-BrdUrd antibodies.
Four pairs of primers were used for competitive PCR experiments: primer set SV12; primer set SV13; primer set SV14 and primer set SV5. The localization of these primer sets with respect to SV40 origin in pwtSV is shown in figure 3.2.1. All the amplifications were carried out as described in "Materials and Methods" and the quantifications for pool 3 (2300 nt.) are shown in figure 3.2.2 panel A.
The results of the quantifications for each primer set within the three BrdUrd pools are plotted in figure 3.2.2 panel B against their distance from the SV40 origin. The number of molecules for every amplified region is expressed in an arbitrary scale: number of primer specific molecules amplified in a fixed volume of the ssDNA pool tested.



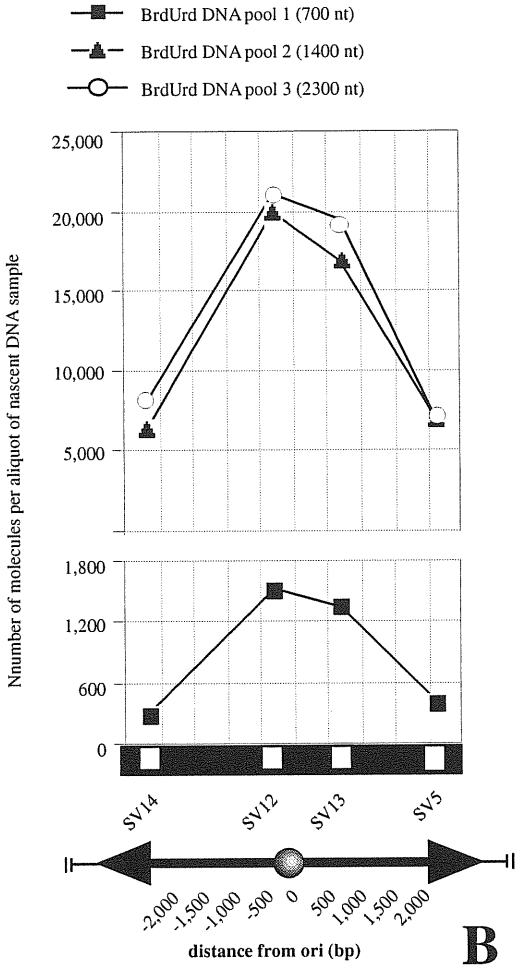


Figure 3.2.2 Mapping of the SV40 origin of DNA replication. **Panel A**, result of the competitive PCR reactions obtained for pool 3 using primer sets SV5, SV12, SV13 and SV14. On the Y axis is reported the ratio competitor/target obtained for every reaction while on the X axis is reported the number of competitor molecules added to every reaction. The ratio competitor/target=1 allows to calculate the number of target molecule present in any analyzed sample; bold lines. The **Panel B**, results of the competitive PCR ractions for all the three pools of nacent DNA. The white boxes on the black bar indicate the position of the amplified marker sequences in comparison with the positions of the SV40 *ori* in the plasmid. The number of molecules for every amplified region is expressed as the number of primer specific molecules amplified in a fixed volume of the ssDNA pool tested.

In the BrdUrd-containing molecules there is a clear enrichment for the target sites amplified by primer sets SV12 and SV13 more evident in pools 2 and 3.
These observations demonstrated that competitive-PCR on the nascent DNA can be used as a tool to map active origins of replication.

3.3 COMPETITIVE PCR ON NASCENT DNA CAN ALSO BE USED TO DEFINE THE DIRECTION OF REPLICATION

The same procedure used for pwtSV was also used to map the *ori* activity in plasmids containing the SV40 minimal origin (pSV2neo and pSV2neo-inv).
For pSV2neo nascent BrdUrd-substituted DNA was purified from three fraction pools (average length 270, 600, and 1000 nt., respectively) by immunoaffinity chromatography. Five pairs of primers were used for competitive PCR experiments: primer set SV1; primer set SV1bis; primer set SV2; primer set SV4 and primer set SV5. The localization of these primer sets with respect to SV40 origin in pSV2neo is shown in figure 3.3.1.

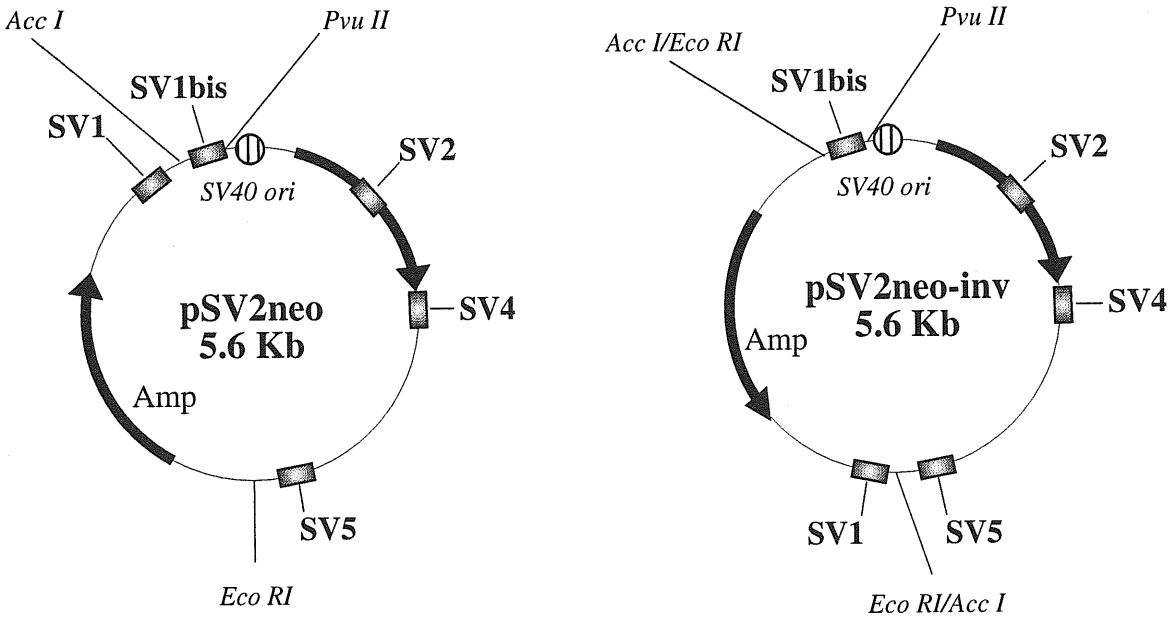
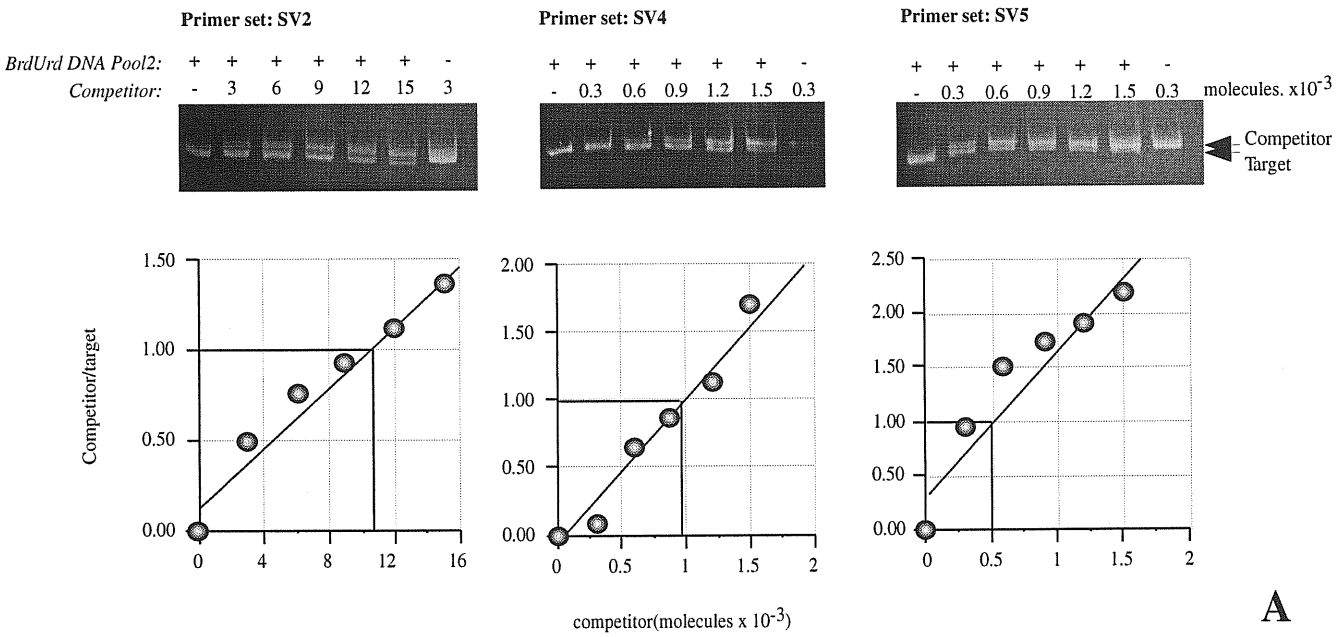


Figure 3.3.1 Position of the regions amplified by the oligonucleotide primer sets SV1, SV1bis, SV2, SV4 and SV5 in pSV2neo and pSV2neo-inv.



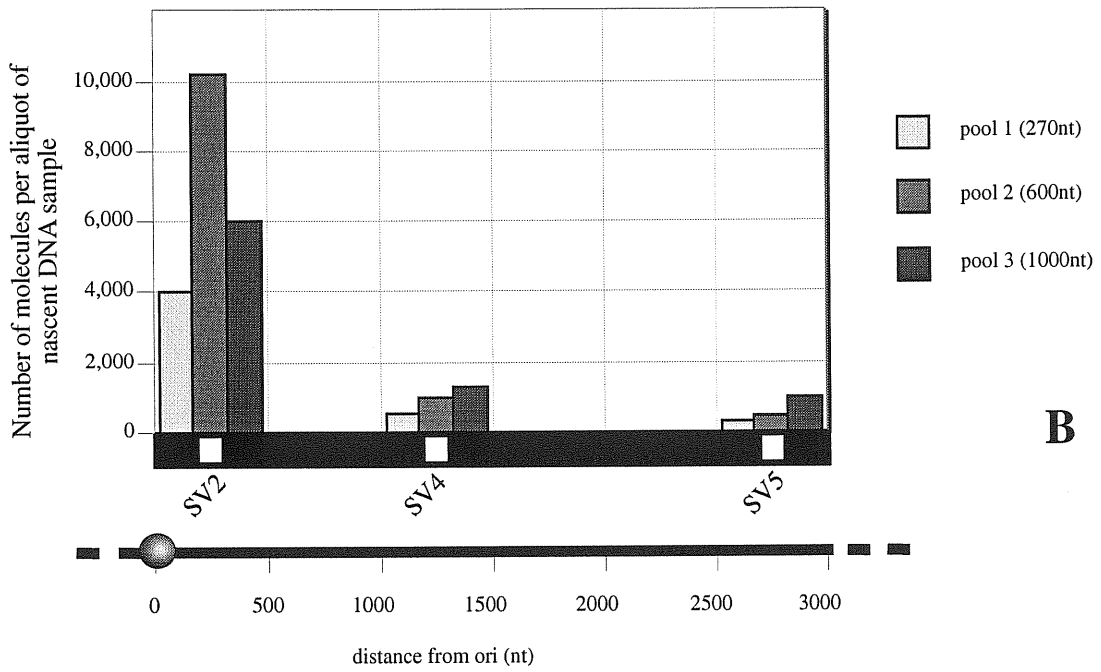


Figure 3.3.2 Mapping of the DNA origin of replication of pSV2neo. **Panel A**, result of the competitive PCR reactions obtained for pool2 using primer sets SV2, SV4 and SV5. On the Y axis is reported the ratio competitor/target obtained for every reaction while on the X axis is reported the number of competitor molecules added to every reaction. **Panel B**, comparison of the quantitation for the pools 1, 2 and 3. The white boxes on the black bar indicate the position of the amplified marker sequences in comparison with the position of the SV40 *ori* in the plasmid. The number of molecules for every amplified region is expressed as the number of primer specific molecules amplified in a fixed volume of the ssDNA pool tested.

Figure 3.3.2 panel A show the results of the quantifications for pool 2 (600 nt.) using primer sets SV2, SV4 and SV5. Also in this case, the number of molecules for every amplified region is expressed in an arbitrary scale. The number of molecules, for each primer set, within the three BrdUrd pools are plotted in figure 3.3.2 panel B against their distance from the SV40 origin. BrdUrd-containing molecules contain up to ten times more target sites for primer set SV2 as compared to the other primer sets confirming that our technique is able to map the SV40 minimal origin activity.
The experiment was repeated with another preparation of nascent DNA using all five primer set for the quantitations (SV1, SV1bis, SV2, SV4 and SV5; figure 3.3.3 panel A). All target sites previously tested (SV2, SV4 and SV5) showed to be enriched in comparison to primer sets SV1 and SV1bis. Considering the relative distance between the SV40-minimal-origin ad the various primer sets, the result was unexpected. In fact, primer sets SV1 and SV1bis lie closer to the origin compared to SV4 and SV5. This behavior was interpreted as a monodirectional replication activity starting from the SV40 origin.

The experiment was repeated with a plasmid derived from pSV2neo by the inversion of the restriction fragment EcoRI-AccI in order to clarify if the monodirectionality was a characteristic of the SV40 minimal origin or if it was a consequence of other features of the pSV2neo vector (as the presence of sequences that inhibits the progression of the replication forks in one direction or the presence of two transcription units oriented in the same direction). The plasmid obtained in this way (pSV2neo-inv) contains the full ampicillin gene coding sequence with its promoter in the opposite orientation compared to pSV2neo and shifts the SV1 primer set in proximity to SV5 (figure 3.3.1). Figure 3.3.3 panel B shows the result of the quantifications obtained with the five primer sets for the 600 nt. pool of nascent DNA. In this case, primer set SV1bis shows the major enrichment immediately followed by primer set SV2.

This data demonstrates that the monodirectional replication is not a characteristic of the SV40 minimal origin but is a consequence of features of the pSV2neo plasmid; moreover it confirms that competitive-PCR mapping can also be used to define the direction of the replication forks from a known DNA origin of replication.

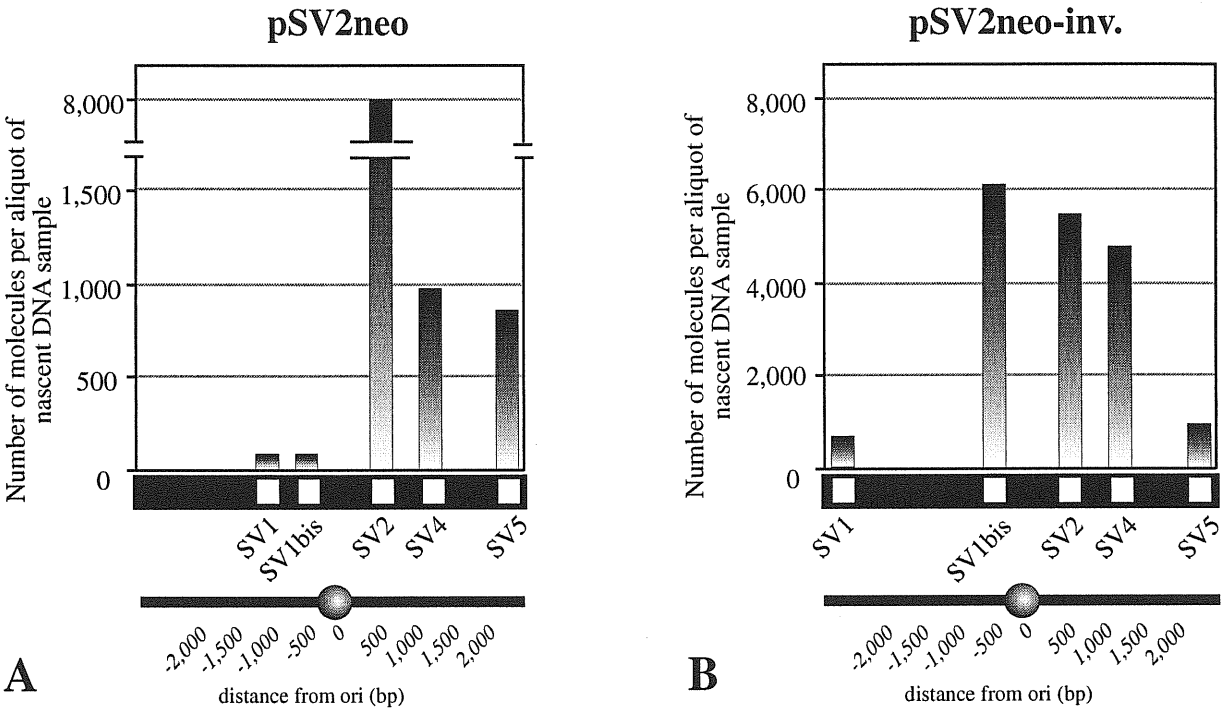


Figure 3.3.3 Mapping of the DNA origins of replication in pSV2neo and pSV2neo-inv. plasmids using primer sets SV1, SV1bis, SV2, SV4 and SV5. **Panel A**, result of the competitive PCR reactions for pSV2neo. **Panel B**, result of the competitive PCR reactions for pSV2neo-inv. On the Y axis is reported the number of molecules amplified in a fixed volume of the nascent DNA pool tested. The white boxes on the black bar indicate the positions of the amplified marker sequences in comparison with the position of the SV40 *ori* in the plasmids.

3.4 MAPPING OF THE LAMIN B2 GENE DOMAIN ORIGIN OF REPLICATION IN SYNCHRONIZED HL60 CELLS

After validation on the SV40 model this method was used to map an origin of replication in a non amplified region of the human genome. As described in section 1.8, the 14 Kb long lamin B2 gene domain was chosen for its characteristic of being early-S replicated in HL60 cells. Nine different primer sets were chosen along the lamin B2 gene locus (figure 3.4.1).

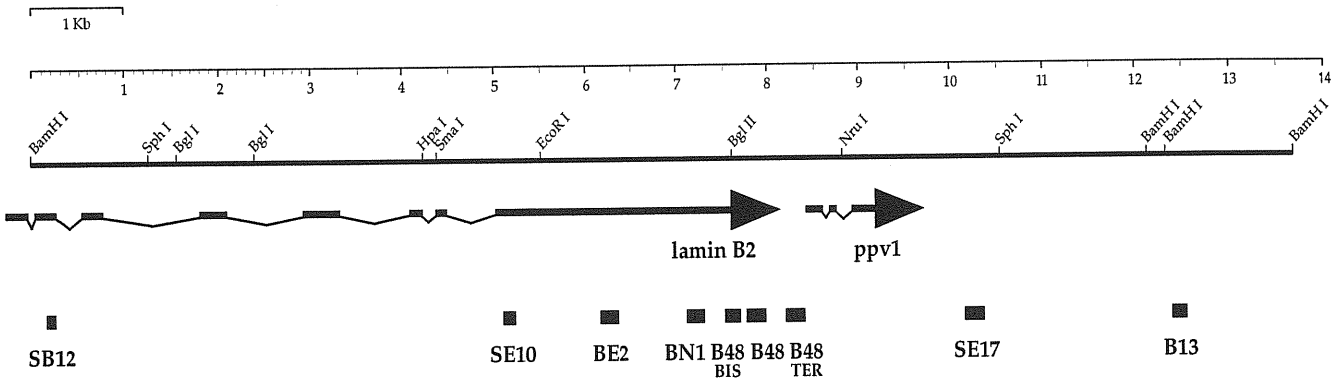


Figure 3.4.1 Positions of the regions amplified by the nine oligonucleotide primer sets (black boxes) used to map the lamin B2 origin of replication.

The first experiments were performed using HL60 cell synchronized at the G1/S border by two aphidicolin blocks, then replication was permitted to start again in the presence of a subinhibitory concentration of aphidicolin allowing a slow synthesis rate. The DNA synthesized at the onset of the S-phase was labeled with BrdUrd plus tritiated deoxycytidine and allowed to elongate for 10, 20 and 30 minutes corresponding to a few minutes at the physiological rate. The total DNA was extracted, heat denatured and fractionated on neutral 5 to 30% sucrose gradients. Figure 3.4.2 shows the profile of the gradient obtained at 20 min. after release from the aphidicolin block.

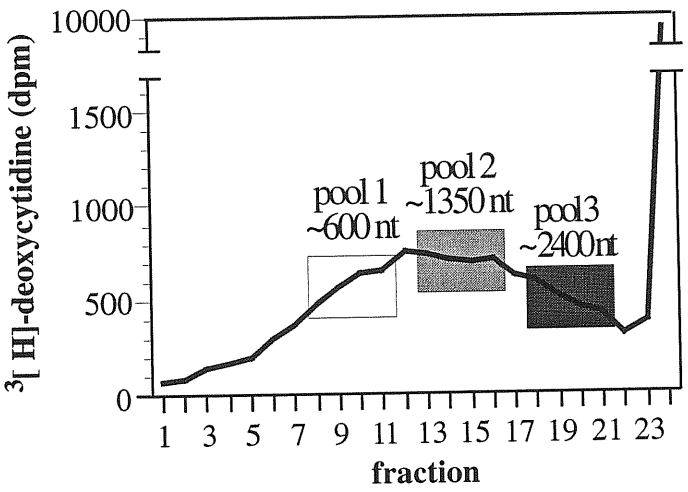


Figure 3.4.2 Profile of the sucrose gradient obtained loading the nascent DNA recovered from HL60 cells labeled for 20 min. after release from the aphidicolin block. The gradient fractions are reported on the X axis while the dpm counted for ^{3}H in every fraction are reported in the Y axis. Three pools of nascent DNA were used for immunoaffinity purification of the BrdUrd-labeled DNA (boxed regions).

The peak of radioactivity present in the low molecular weight fractions represents the labeled-neosyntesized DNA while the peak present toward the bottom of the gradient represents the radioactivity incorporated from the cells escaped to the synchronization or incorporated as result of the DNA-repair synthesis. Newly synthesized, BrdUrd-labeled, DNA was recovered by immunoaffinity chromatography from pools of fractions with various molecular weights (figure 3.4.2 and 3.4.3). Competitive PCR analysis was performed to measure the relative abundance of the target sequences indicated by the white boxes in the black bars of figure 3.4.3.

The results of the experiments at 20 min. and 30 min. after release from the aphidicolin block are shown in figure 3.4.3 A and B respectively. Values on the Y axis depict the number of primer specific molecules amplified in a fixed volume of the ssDNA pool tested. The ß-globin coding sequence is a region that is transcriptionally inactive in HL60 cells and represents a portion of the genome that is not replicated during the early S-phase.

These results indicate that the number of nascent DNA molecules deriving from the primer set B48 is higher than that deriving from the flanking sequences, thus suggesting that the B48 region contains an active origin of DNA replication activated in the first minutes of the S-phase in HL60 cells.

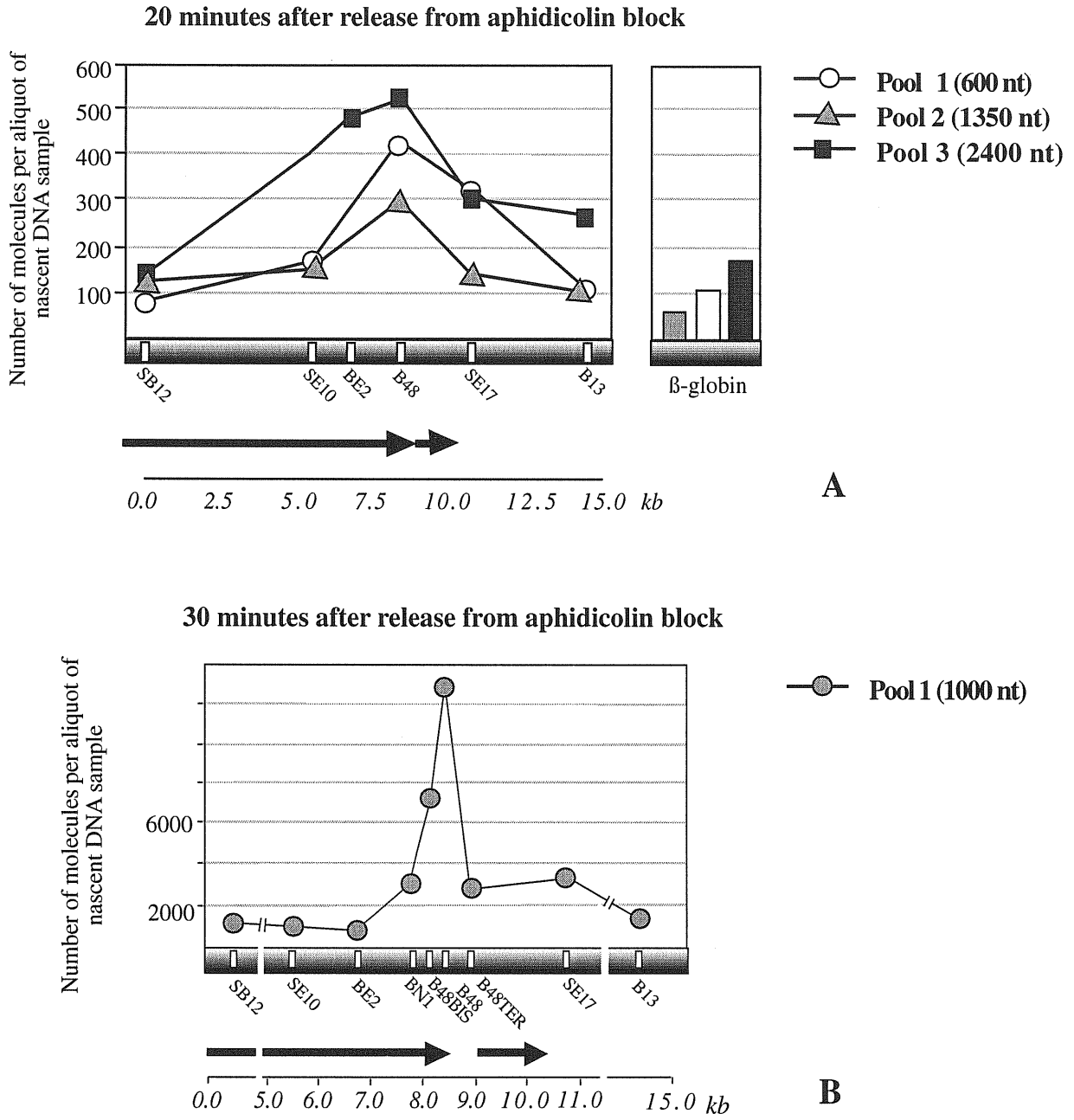


Figure 3.4.3 Mapping of the lamin B2 origin of DNA replication in synchronized HL60 cells. **Panel A**, result of the competitive PCR reactions for three pools of nascent DNA obtained from HL60 cells labeled for 20 min. after release from aphidicolin block. **Panel B**, result of the competitive PCR reactions for a pool of nascent DNA obtained from HL60 cells labeled for 30 min. after release from aphidicolin block. The white boxes on the black bar indicate the positions of the amplified marker sequences in comparison with the positions of the two transcription units in the lamin B2 gene locus. The number of molecules for every amplified region is expressed as the number of primer specific molecules amplified in a fixed volume of the nascent DNA pool tested. Quantification of the β-globin gene copy number was obtained by the primer set PCO.

3.5 MAPPING OF THE LAMIN B2 ORIGIN IN UNSYNCHRONIZED HL60 CELLS

Since the competitive PCR technique employed is very sensitive (theoretically down to the threshold of one molecule), it was possible to investigate the initiation of replication also from unsynchronized cell, avoiding the problems and the possible artifacts relative to the synchronization protocols.

Unsynchronized HL60 cells were labeled for a 10 min. pulse with BrdUrd plus tritiated deoxycytidine. Genomic DNA was extracted, denatured and size fractionated on neutral 5 to 30% sucrose gradients. Figure 3.5.1 panel A shows the profile of the sucrose gradient.

During the pulse-labelling of unsynchronized cells most of the radioactivity is incorporated by the replicons that are already active when labelling starts and only those origins activated during the time pulse produce low molecular weight DNA. As a consequence, the amount of ^{3}H incorporated in the low molecular weight fractions is very low while most of the radioactivity lies towards the bottom of the gradient.

Newly synthesized, BrdUrd-labeled, DNA was recovered by immunoaffinity chromatography from a pool of fractions with an average molecular weight of 1500 nt. Competitive PCR analysis was performed to measure the relative abundance of the same sequences described for synchronized cells. The result of the quantitations is showed in figure 3.5.1 panel B.

A portion of the insulin gene coding sequence (representing a genomic region that is transcriptionally inactive in HL60 cells and far from known origins of replication) was used as reference.

Also in the case of unsynchronized HL60 cells, the number of nascent DNA molecules deriving from the primer set B48 is higher than that deriving from the flanking sequences. This result confirms that the B48 region contains an active origin of DNA replication.

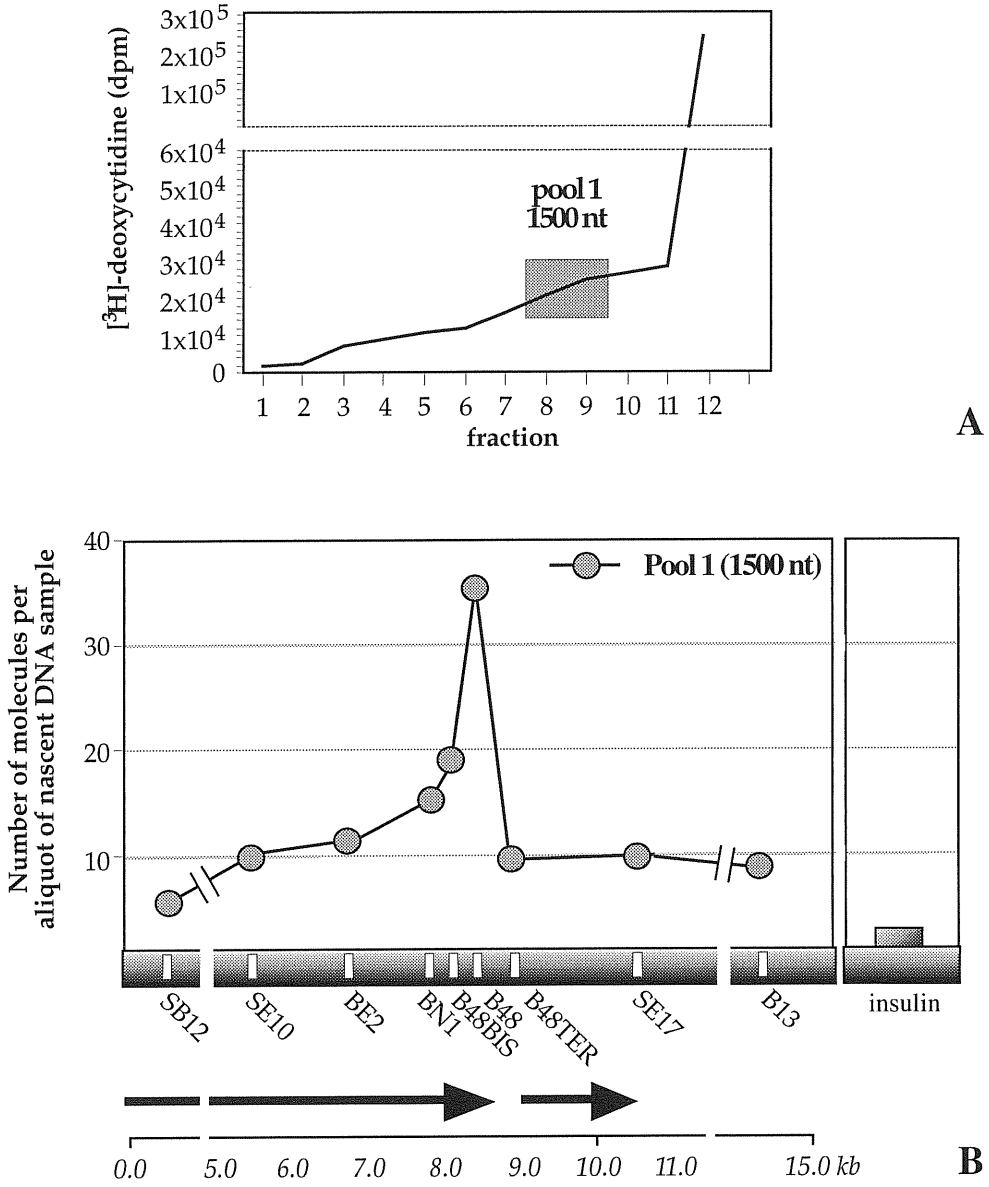


Figure 3.5.1 Mapping of the lamin B2 origin of DNA replication in unsynchronized HL60 cells. **Panel A**, Profile of the sucrose gradient obtained loading the nascent DNA recovered from HL60 cells labeled for 30 min. after release from the aphidicolin block. The gradient fractions are reported on the X axis while the dpm counted for ^{3}H in every fraction are reported in the Y axis. The pool of nascent DNA used for immunoaffinity purification of the BrdUrd-labeled DNA is boxed. **Panel B**, result of the competitive PCR reactions for the pool of nascent DNA.The white boxes on the black bar indicate the positions of the amplified marker sequences in comparison with the positions of the two transcription units in the lamin B2 gene locus. The number of molecules for every amplified region is expressed as the number of primer specific molecules amplified in a fixed volume of the nascent DNA pool tested.Quantification of the insulin gene copy number was obtained by the primer set I.

In order to rule out possible biases whilst estimating the abundance of selected DNA fragments introduced by the density labelling and/or purification procedures, an HL-60 culture was uniformly labeled for 24 hr with BrdUrd; nuclear DNA was extracted, sonicated to approximately 1000 bp, denatured, purified as reported above, and assayed for the abundance of five lamin B2 markers. No significant enrichment for any sequence (including B48) was observed (not shown).
To further control that the primers used did not introduce any bias in the analysis of marker abundance, the relative content for the same segments probed in this work was measured in a sample of DNA from non-proliferating HL60 cells differentiated by treatment with retinoic acid and dimethylformamide. As expected, no enrichment for B48 is detected in this sample.

3.6 SIMPLIFICATION OF THE MAPPING PROCEDURE AND MAPPING OF THE LAMIN B2 ORIGIN IN DIFFERENT CELL LINES

The original method for DNA replication origin mapping outlined until now is a long procedure that consist of several steps. First, cells must be labeled with BrdUrd, than nascent DNA is purified from bulk DNA (deriving from all regions in the genome, the cells being unsynchronized) according to its size. Finally, BrdUrd-labeled DNA is further purified by affinity chromatography (figure 3.6.1 protocol A).
Since the origin mapping procedure by fragment abundance detection, originally developed by Vassilev, relies in principle on size selection only, we attempted a simplification of this procedure by omitting the affinity chromatography step.
Competitive PCR mapping of the lamin B2 gene domain origin of replication was repeated using exponentially growing asynchronous cells with and without affinity purification of the nascent strand (not shown). Omission of the BrdUrd-purification step did not give any appreciable differences neither in the delectability of DNA segments by competitive PCR nor in the relative abundance of the markers. On the contrary, omission of this step, besides simplifying the procedure, also avoids the introduction of a possible bias in fragment selection on the anti-BrdUrd affinity column and reduces the possibility of DNA fragmentation by photodamage (Kumar *et al.* 1996).

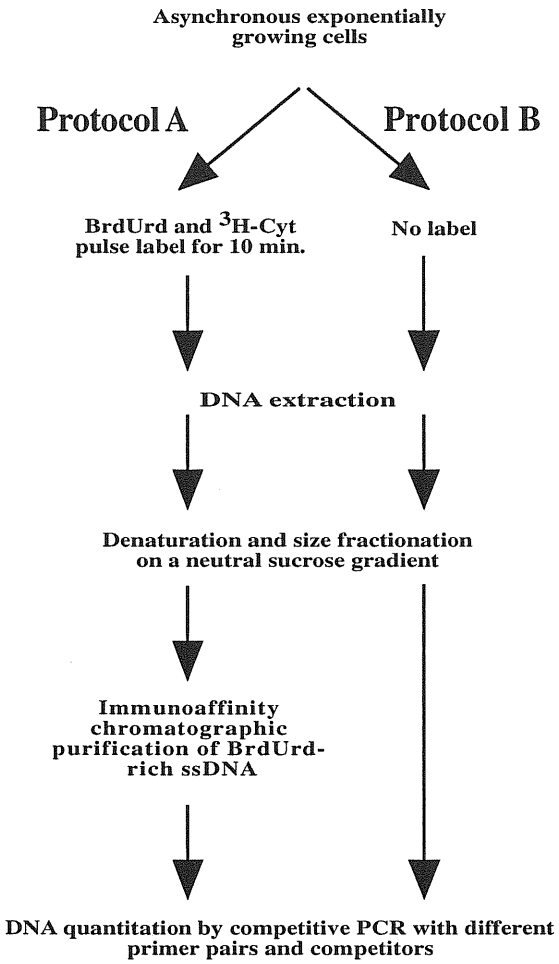


Figure 3.6.1 Flow chart diagram showing the original (protocol A) and the simplified (protocol B) procedures for preparing newly replicated DNA fractions for *ori*-mapping studies.

BrdUrd labelling was no longer utilized, and the simplified procedure of Figure 3.6.1 protocol B was used for all the quantifications that I will follow in this thesis.

Tanks to this simplification, the mapping of the origin activity was make easier and faster. This made possible the mapping of the lamin B2 origin besides in HL60 cells (human myeloid leukemia cell line) also in other human cell lines with different histological derivation. These include the IMR32, SKNBE and SKNMC neuroblastoma, the HeLa epithelial and the IMR90 lung fibroblast cell lines. The result of this mapping for pools of nascent DNA of 1000 nt. is shown in figure 3.6.2.

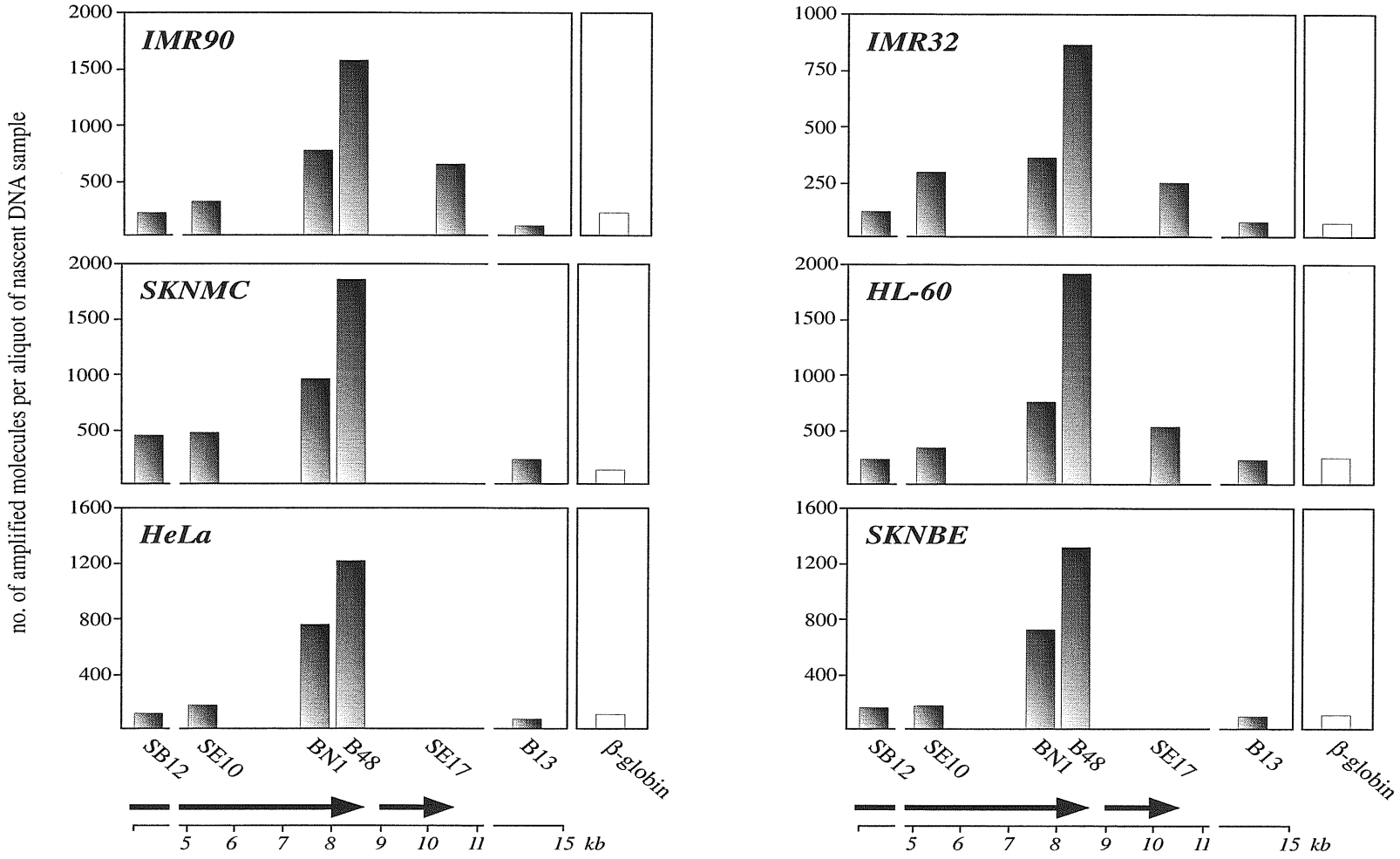


Figure 3.6.2 Graph representation of the results obtained by the quantification of different segments in the lamin B2 gene domain by competitive PCR of nascent DNA samples from the indicated cell lines. PCR quantitation results are given in terms of an arbitrary value, i.e. number of primer specific molecules amplified in a fixed volume of the ssDNA pool tested for each cell line. Quantification of the β-globin gene copy number was obtained by the PCO primer set.

In all these cells, a clear enrichment could be detected for the genomic fragment corresponding to the B48 marker, scoring 6-10 fold higher than the markers localized >5 kb apart on each side (SB12 and B13), and those of the β-globin gene control region.
This distribution of DNA segment abundance in samples of nascent DNA clearly suggests that the lamin B2 *ori* is active in all the analyzed cell lines behaving as a constitutive DNA origin of replication. This result is in agreement with the opinion that the lamin B2 gene is an essential gene and probably lies in a genomic region with a chromatin similarly organized in different cells and tissues.

3.7 *IN VIVO* DNA-PROTEIN INTERACTIONS AT THE LAMIN B2 ORIGIN

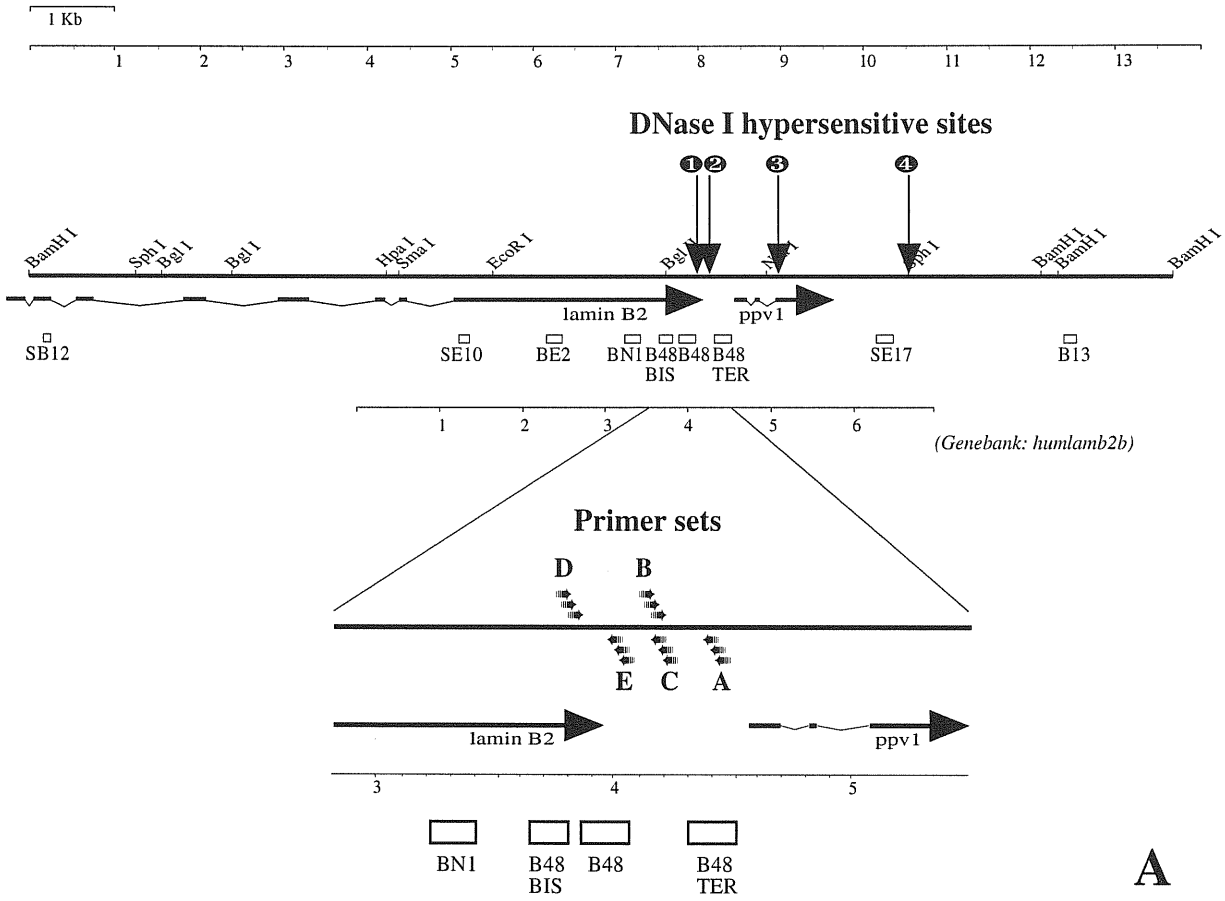
In all these cases, the observation that the same and rather narrowly defined region of the chromosome works as an origin in very different cell types, gives more weight to the conception that origins in higher organisms are constituted by well defined cis-acting sites interacting with specific trans-acting protein factors.
A study of the nature of the proteins involved in the definition and activation of the human lamin B2 origin was initiated by D. Dimitrova (Dimitrova *et al.* 1996) who analyzed the region by DNase-hypersensitivity and genomic footprinting assays. The study was continued by G. Abdurashidova who studied the dynamics of the *in vivo* DNA-protein interactions at the lamin B2 origin in the different phases of the cell cycle (unpublished results).
These studies have opened the way to the *in vitro* studies aimed at the identification of the protein factors responsible for these interactions which represent the second part of this thesis. For this reason a summary of the *in vivo* results will be reported in this session to allow a better comprension of the whole work.

Four DNaseI-hypersensitive sites (DHS) were revealed in the lamin B2 chromosomal region in nuclei from exponentially growing HL60 cells while they were not present in naked DNA (not shown). A schematic representation of the DHS is reported in figure 3.7.1 panel A.

Interestingly, all the hypersensitive sites are constitutive since they are also present in the chromatin of terminally differentiated HL60 cells. Two of the nuclease-hypersensitive sites (DHS-1 and DHS-2) map to the spacer between the lamin B2 and the ppv1 genes, exactly where the replication initiation site is located, suggesting the presence of an altered chromatin structure, probably due to the binding of non-histone protein factors.

High resolution analysis of protein-DNA interactions in the 600-bp region encompassing the 3' end of the lamin B2 gene and the ppv1 promoter was performed by ligation-mediated-PCR *in vivo* footprinting using the primer sets reported in figure 3.7.1 panel A. DNaseI and DMS were the probing reagents used to test the genomic organization of the region in both exponentially growing and differentiated (non-proliferating) HL60 cells according to the scheme reported in figure 3.7.1 panel B . Due to its relatively big size, DNase I cannot penetrate and cleave DNA within protein-DNA complexes and the footprints it produces are usually large and clear; furthermore, DNase I can also provide information about the nucleosomal organization of the analyzed region. In view of its size, which does not allow it to cross the cell membrane, DNase I footprinting can be performed only with isolated nuclei or permeabilized cells. To avoid potential artifacts deriving from the cell disruption step DMS was employed, in parallel, as a second probing reagent, i.e. a small molecule that can be used with intact cells. DMS modifies guanine residues at the N-7 position in the major groove and, to a much lesser extent, adenine residues at the N-3 position in the minor groove of DNA. This sequence specificity, however, reduces the usefulness of this probe in the case of proteins which recognize AT-rich sites in DNA. The high reactivity and small size of DMS can also be a potential drawback, since it can penetrate protein-DNA complexes and render them invisible. The combined use of the DNase I and DMS treatments, by providing complementary information, is best suited to obtain a fine description of protein-DNA interactions.

In growing HL60 cells, footprints homologous to binding sites for known transcription factors (members of the b-HLH family, NRF-1, Sp1 and UBF) were detected, using primer sets A and B, in the region corresponding to the ppv1 promoter (figures 3.7.2 and 3.7.4).



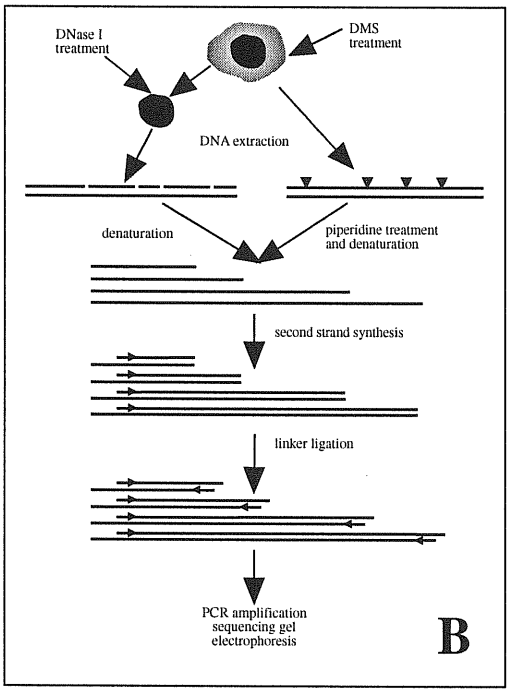


Figure 3.7.1 Panel A, localization of the hypersensitive site positions (arrows 1, 2, 3 and 4) and of the primer sets used for the genomic footprinting analysis (A, B, C, D and E) in the lamin B2 gene locus. **Panel B**, Flow chart of *in vivo* footprinting by ligation-mediated PCR.

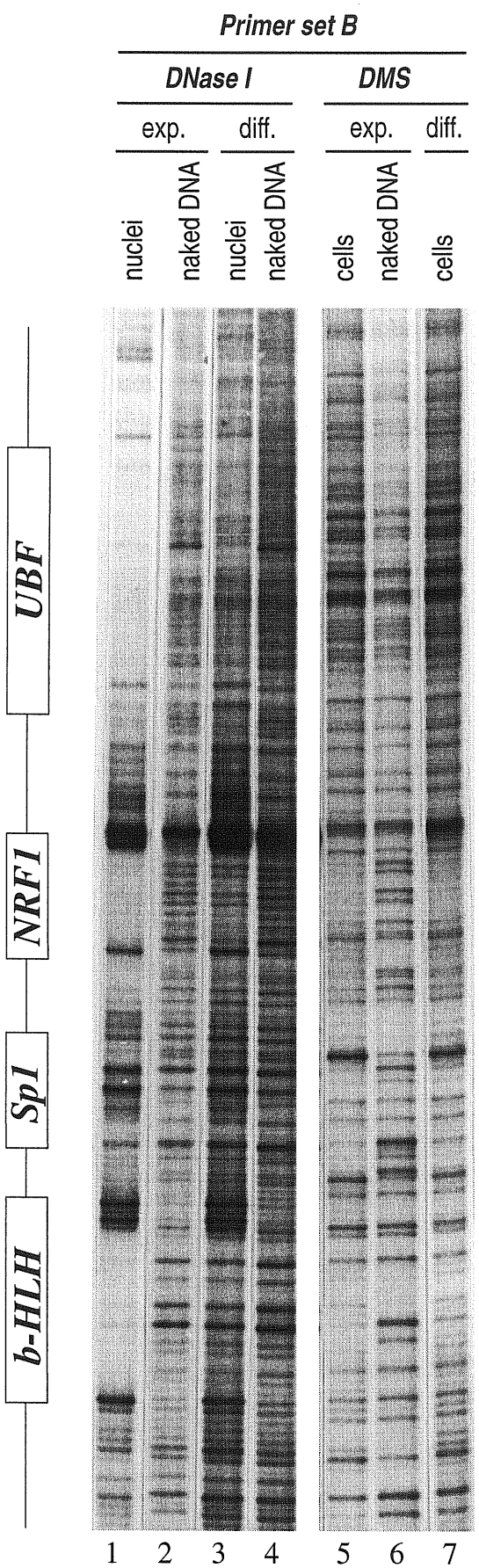
One of the protections pertains a 20 bp sequence (nt. 4186-4205 HUMLAMB2B, Genebank) containing the hexamer motif, 5'-CACGTG-3', recognized by proteins of the B subgroup of the basic-helix-loop-helix (b-HLH) family (including the Myc protein and the transcription factor USF). This region is protected on both strands from the action of DNase I and DMS.

Another footprint is adjacent to the b-HLH site, covering the region between nucleotides 4216-4231 on the upper strand and 4211-4230 on the lower strand. Close inspection of the protected element suggests that this footprint may actually be generated by the binding of two proteins, one over nucleotides 4211-4221, and the other over nucleotides 4221-4231. The left side (nt. 4211-4221) of the footprint contains a high-affinity binding site for transcription factor Sp1 (sequence 5'-GCTCCGCCCCGG-3'; Briggs *et al.* 1986).

Further downstream 4240-4270, a prominent footprint is evident on both strands in the DNase I- and DMS-cleavage patterns. It contains a sequence, 5'-TGCGCATGCGCG-3', which matches perfectly the binding site of the transcription factor Nuclear Respiratory Factor-1, NRF-1 (Virbasius *et al.* 1993).

Finally, protein binding *in vivo* was detected ~30 bp downstream from the NRF-1 motif, over nucleotides 4304-4352. This site contains three short, directly repeated sequences, 5'-(T/C)CGGCC-3', which are homologous to a binding site for the ubiquitous RNA polymerase I transcription factor Upstream Binding Factor, UBF (Bell *et al.* 1988). The 30-bp region between the NRF-1 and UBF sites contains numerous DNase I-hypersensitive nucleotides which suggests that the binding of one or both protein factors significantly distorts the DNA secondary structure.

Upon conversion to a non-proliferative state, a reduction in the intensity of these footprints was observed, that paralleled the diminished transcriptional activity of the region (figure 3.7.2). Although this change is suggestive of some alteration in the DNA-protein contacts in the region, it is obvious that the factors still bind *in vivo* to their recognition sites.



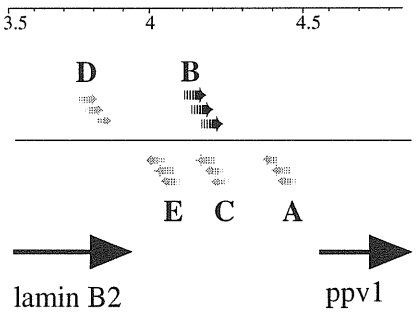


Figure 3.7.2 *In vivo* footprinting at the ppv1 promoter using primer set B. DNaseI and DMS *in vivo* footprinting in exponentially growing and differentiated HL60 cells. Lanes 2, 4 and 6, naked DNA; lanes 1, 3, 5 and 7, *in vivo* footprinting patterns. Four main DNA-protein interactions were identified. On the basis of their DNA sequences, these sites have been identified as putative binding sites for the indicated transcription factors (white boxes).

Examination of the region immediately upstream of the b-HLH motif, using primer set C, did not reveal any footprint in this area (not shown). All guanines are methylated to a similar extent *in vivo* and *in vitro* over the whole DNA length explored. In contrast, *in vivo* DNase I- treatment reveals multiple hypersensitive nucleotides, either single or grouped together in clusters. This difference between the *in vivo* and naked-DNA patterns indicates that DNA is engaged *in vivo* in some kind of interaction with nuclear proteins. Though it is difficult to precisely localize the DHSs described in the previous section, it is likely that the region analyzed with primer set C corresponds to the "cold spot" between DHS-1 and DHS-2. Taking into consideration the nucleosome size of the region (~190 bp), the absence of footprints and the DNase I-cleavage pattern, we speculate that a "wobbling" nucleosome may be present *in vivo* immediately upstream of the b-HLH box.

Extension of the LMPCR analysis to the replication initiation site, using primer set D and E, revealed a large and clear footprint (with both DNase I and DMS treatment) protecting a region of 70 nt. on the lower strand (nt. 3923-3992 HUMLAMB2B, Genebank; figures 3.7.3 and 3.7.4). This large protection is divided in two subregions by an unprotected site (nt. 3964).

Interestingly, *in vivo* DMS-footprinting analysis of the upper strand identified only few protected nucleotides in the same region. On the other hand, *in vivo* DNase I digestion pattern displayed periodic hypersensitive nucleotides with an average spacing of 11-12 nt., a pattern of cleavage typically found when DNA wraps around a protein core.

A computer search of known protein recognition sequences did not reveal any significant homology to the 70 bp footprint discovered here. It is worth noting, that the leftmost part of the footprint overlaps with a previously identified putative binding site for topoisomerase II (nt. 3921-3929).

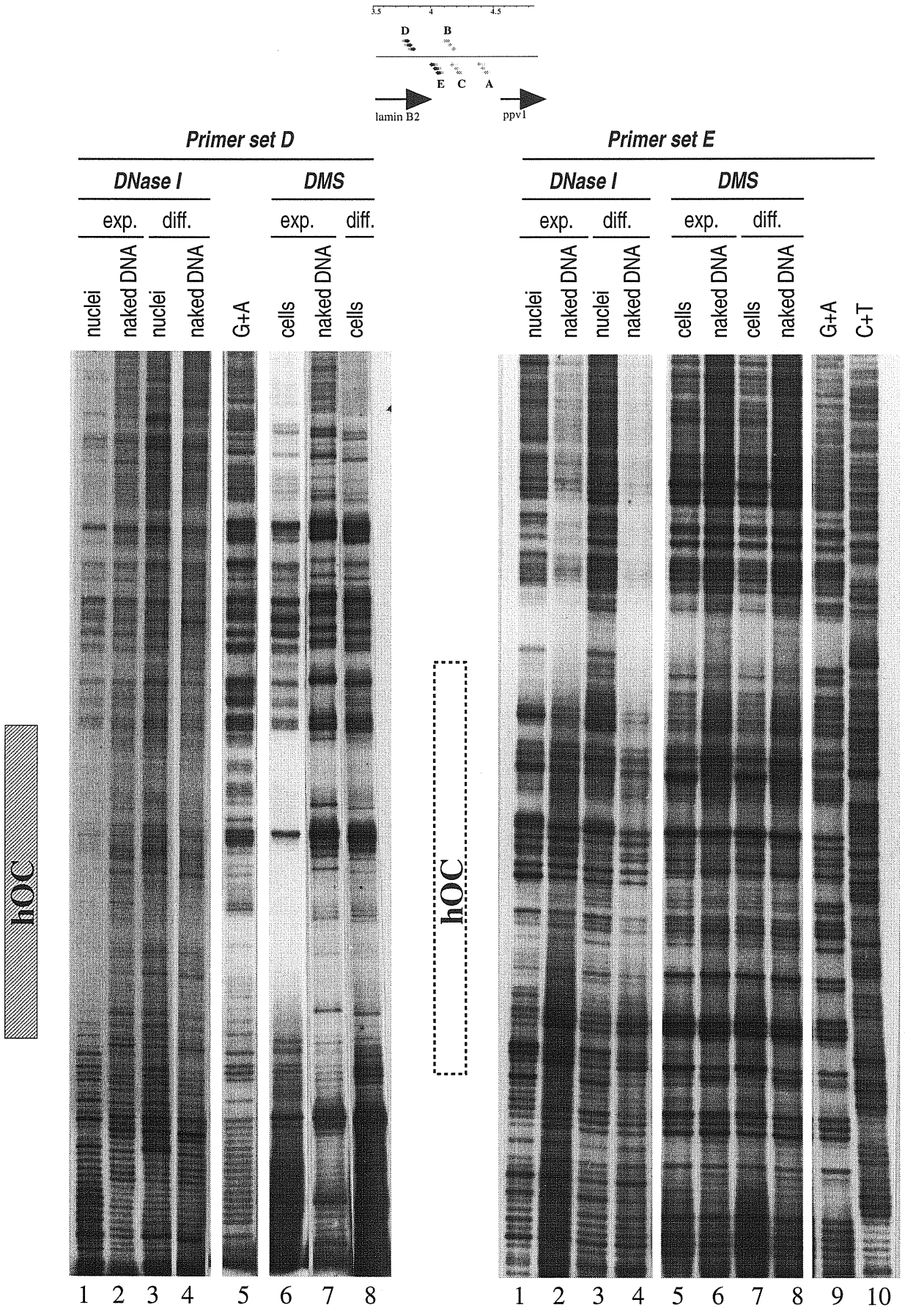
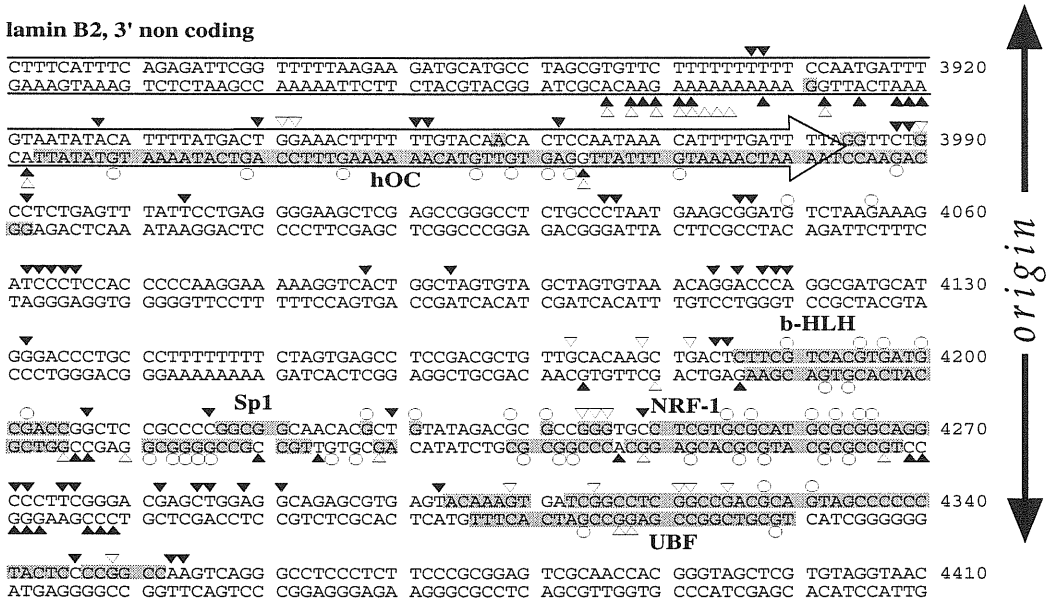
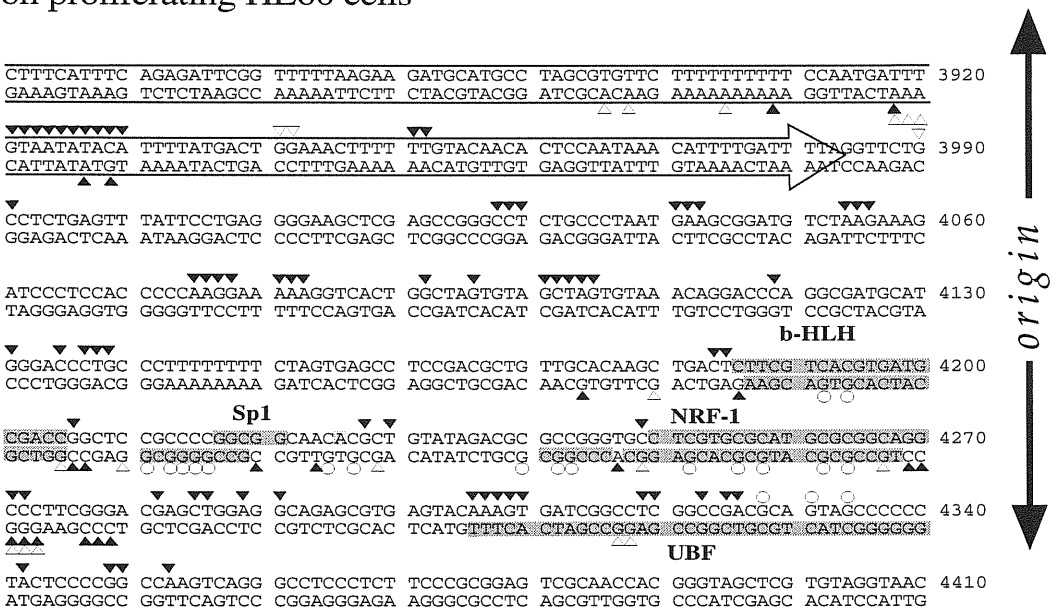


Figure 3.7.3 Genomic fotprinting using the primer sets D and E. DNaseI and DMS *in vivo* footprinting in exponentially growing and differentiated HL60 cells. Primer set D: lanes 2, 4 and 7, naked DNA; lanes 1, 3, 6 and 8, *in vivo* footprinting patterns; lane 5, G+A sequence. Primer set E: lanes 2, 4, 6 and 8, naked DNA; lanes 1, 3, 5 and 7, *in vivo* footprinting patterns; lanes 9 and 10, sequence reactions. The 70 nt. region protected in exponentially growing cells (hOC site) is indicated by the the boxed regions on the protected (on the left) and on the unprotected (on the right) strands.



Summary of the *in vivo* footprinting experiments on proliferating HL60 cells



Summary of the *in vivo* footprinting experiments on differentiated HL60 cells (Non proliferating)

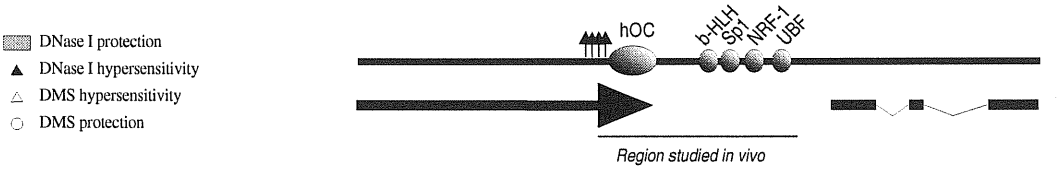


Figure 3.7.4 Genomic footprinting at the lamin B2 origin of replication in HL60 cells. Summary of the protections and of the hypersensitive sites in proliferating and in differentiated cells. Some of the contacts identified *in vivo* appear to correlate with the proliferative state of the cells (hOC region) while others do not show dramatic changes between proliferating and non proliferating cells.

Considering the large size and bipartite structure of the protected sequence, it is likely that more than one protein binds there. It is noteworthy that the binding of the protein(s) to DNA induces a structural deformation in the adjacent AT-rich region which makes it hypersensitive to DNase I-cleavage. This footprint completely disappeared in non-proliferating HL60 cells (figures 3.7.3 and 3.7.4).
These observations suggest that this specific protein-DNA interaction is involved in the process of origin activation and for this reason, in this thesis, we will name the DNA-binding activity responsible for this interaction human Origin Complex (hOC). A summary of the results obtained *in vivo* with primer sets A, B, C, D and E is shown in figure 3.7.4.
The result obtained with the HL60 cells was also confirmed in IMR90 which is a lung fibroblast cell line with a defined life span in which the lamin B2 origin is active (not shown). Furthermore, genomic footprinting studies in synchronized IMR90 (figures 3.7.5) showed that the pattern of protection changes during the cell cycle. The protected region is more extended in G1 (almost 200 nt. on the lower strand; figure 3.7.6) while during the S-phase the pattern is similar to that found in unsynchronized cells. On the contrary, the protections in the ppv1 promoter region do not show dramatic differences passing from the G1 to the S phase (not shown). No protection was identified in metaphase cells both in the B48 region and in the ppv1 promoter.
Another data that strongly suggests a possible involvement of the hOC site in the origin activity is the observation that genomic footprintings performed in IMR90 cells blocked in G0, by serum starvation, didn't show any protection at the hOC site (figure 3.7.5), as was previously observed in differentiated (not replicating) HL60 cells. This means that the activity responsible for these interactions is present only in cycling cells and is able to assemble at the origins only when the replication of DNA is needed.

On the basis of these results, *in vitro* experiments were performed with the aim of identifying the protein factors involved in the described interactions.

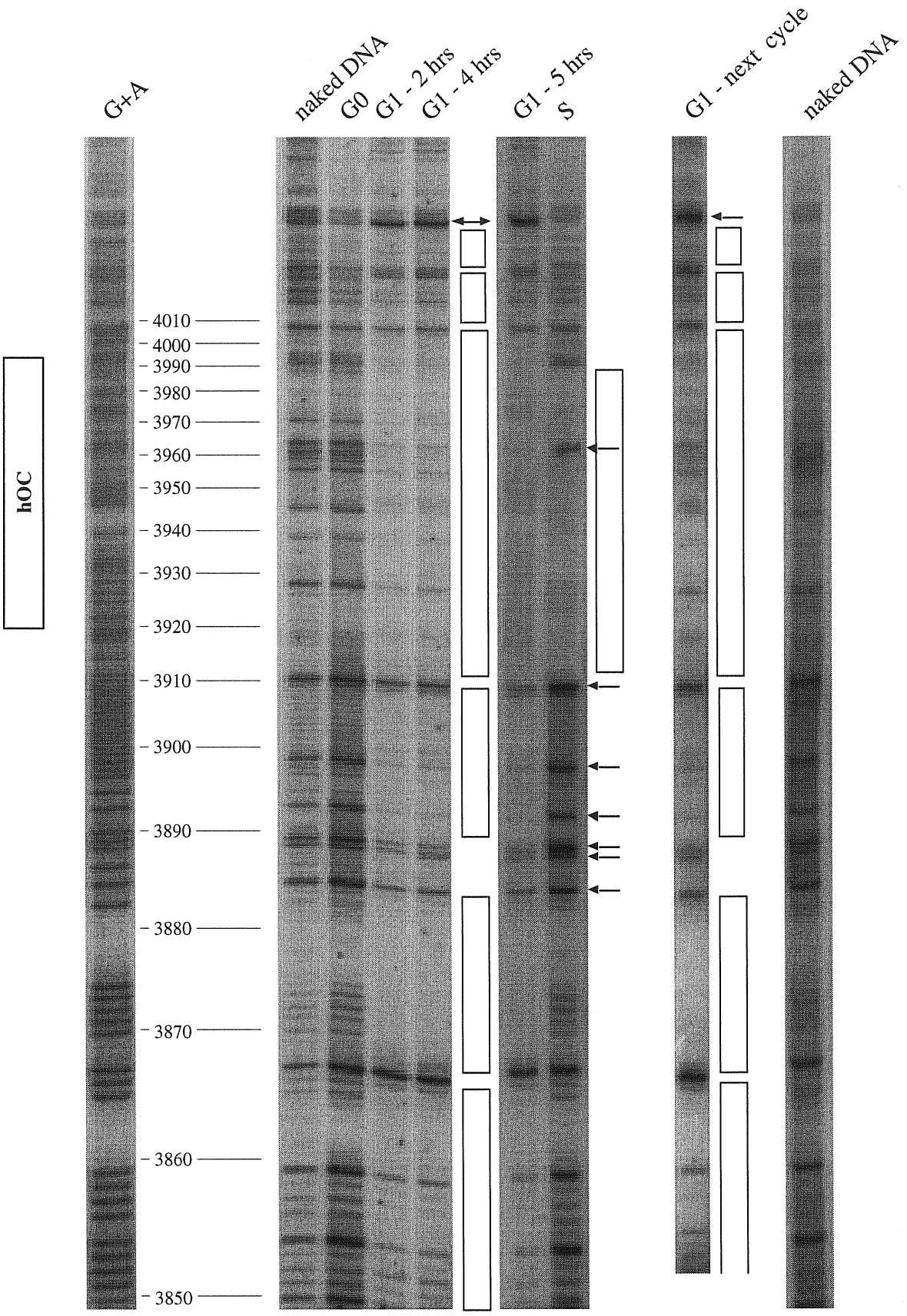


Figure 3.7.5 Genomic footprinting in synchronized IMR90 cells. DMS protections (Boxed regions)

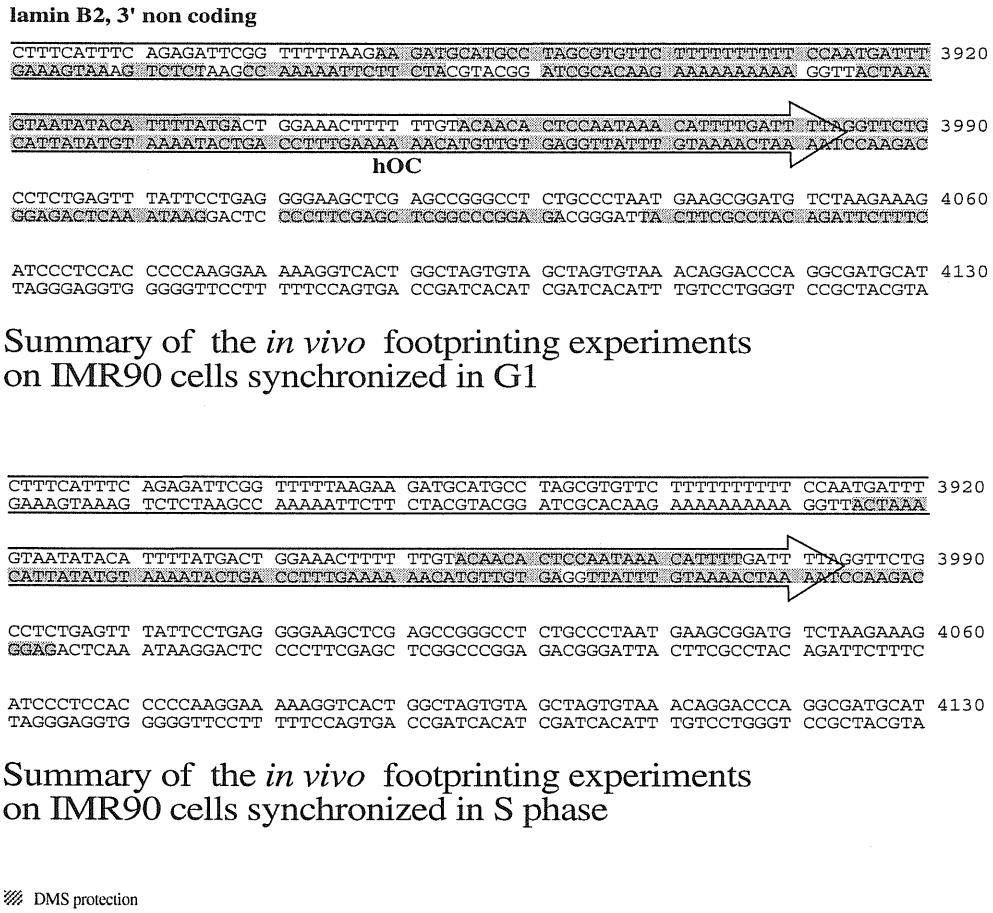


Figure 3.7.6 Genomic footprinting at the lamin B2 origin of replication in synchronized IMR90 cells. Summary of the protections and of the hypersensitive sites in G1 and S phases of the cell cycle.

3.8 *IN VITRO* PROTEIN-DNA INTERACTIONS AT THE PPV1 PROMOTER

Early experiments aimed at the identification of DNA binding proteins for the B48 region are reported in Csordas *et al.* 1993. In this case, gel retardation assays using the 24 bp putative binding site for USF (USF-BS probe) identified a factor able to shift specifically the probe in HeLa nuclear extracts (Dignam). The same probe used in a southwestern assay

revealed a specific complex of roughly 44 KDal together with other unspecific complexes. The purification of the factor by classical chromatographic techniques confirmed that the protein was the ubiquitous transcription factor USF.

Having fine mapped the lamin B2 origin, the *in vitro* DNA-protein interactions at the intergenic region were further investigated using nuclear extracts of different cell lines.

Dignam extracts from HL60 and HeLa cell lines were used in band shift assays with probes representing sequences of the ppv1 promoter. A long probe encompassing the four sites protected *in vivo* (probe 4SR) produced a confused pattern of bands due to the simultaneous presence of several complexes (not shown).

DNaseI *in vitro* footprinting assay with a related probe (probe 4SR-lower) demonstrates that using Dignam extract from HL60 or HeLa cells it is possible to reproduce *in vitro* the same pattern identified *in vivo* for the ppv1 promoter. Three clear protections are present at the putative binding sites for USF, Sp1 and NFR1 while only a faint protection is present in Dignam extracts from HL60 at the putative binding site for UBF (figure 3.8.1).

Western blots with antibodies against USF and NRF1 demonstrate that both proteins are present in the extracts positive in footprinting (not shown).

Gel retardation assays using the 25 bp probe for NRF1 (NRF1-BS) showed two retarded bands. The specific band was able to resist a 100 fold molar excess of an unspecific competitor (the 24 bp binding site for USF) added to the probe prior to incubation with the protein extract while it almost disappeared after addition of only 10 fold molar excess of cold competitor. On the contrary the unspecific band disappeared only when a high excess of both cold specific or unspecific competitors were used (figure 3.8.2 panel A). The addition of increasing amounts of an anti-NRF1 serum (gift of R. C. Scarpulla) to some band-shift reactions showed the super-shift of the specific band leaving the unspecific band unchanged (figure 3.8.2 panel B). This means that NRF1 is present and is able to bind *in vitro* its putative binding site in the ppv1 promoter. Similar experiments were also repeated for USF (not shown).

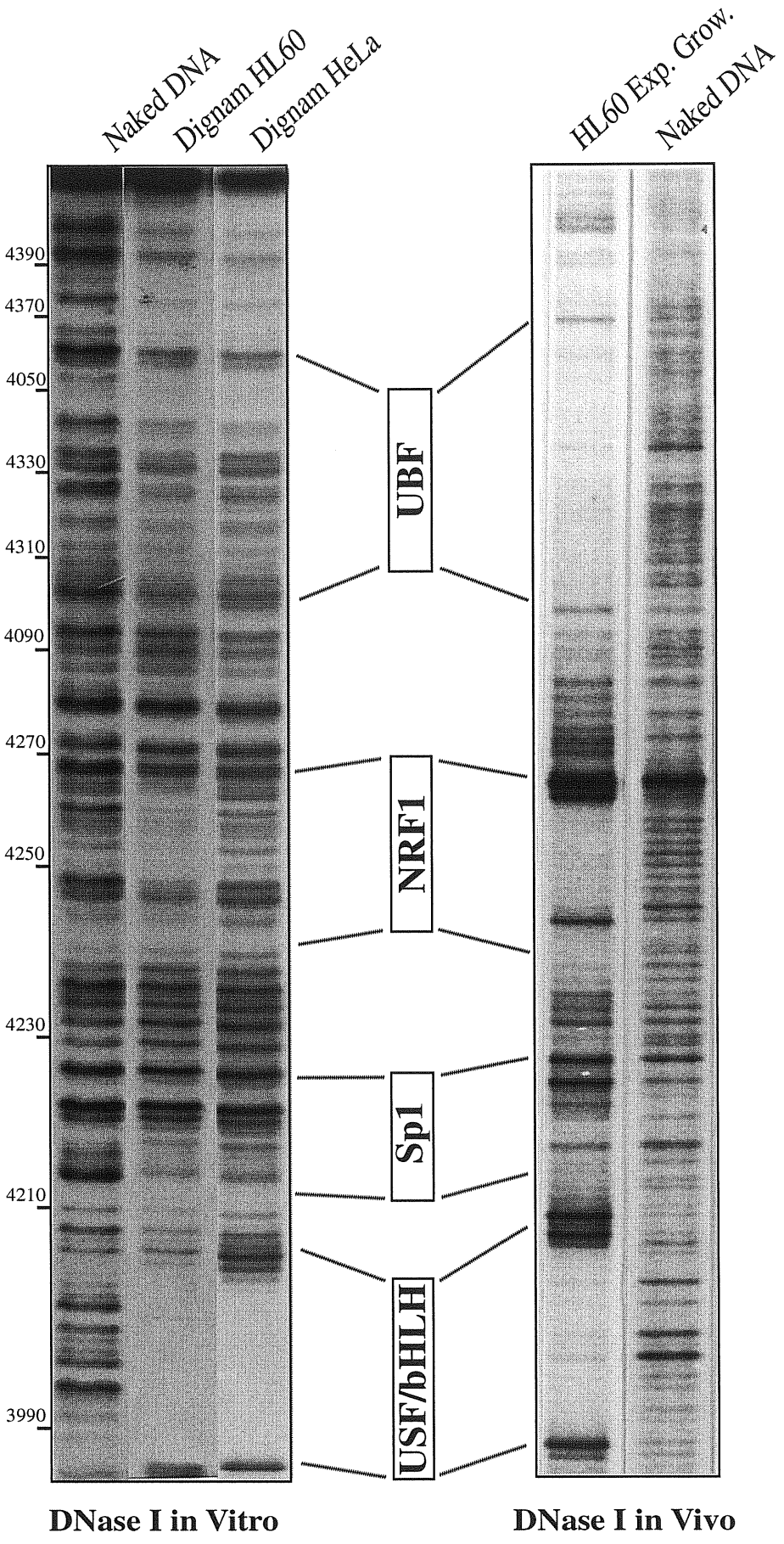


Figure 3.8.1 DNase I *in vitro* and *in vivo* footprinting on the portion of the lamin B2 origin encompassing the ppv1 promoter. The putative binding sites identified by sequence analysis (white boxes) are compared to the protection obtained *in vitro* (on the left) and *in vivo* (on the right).

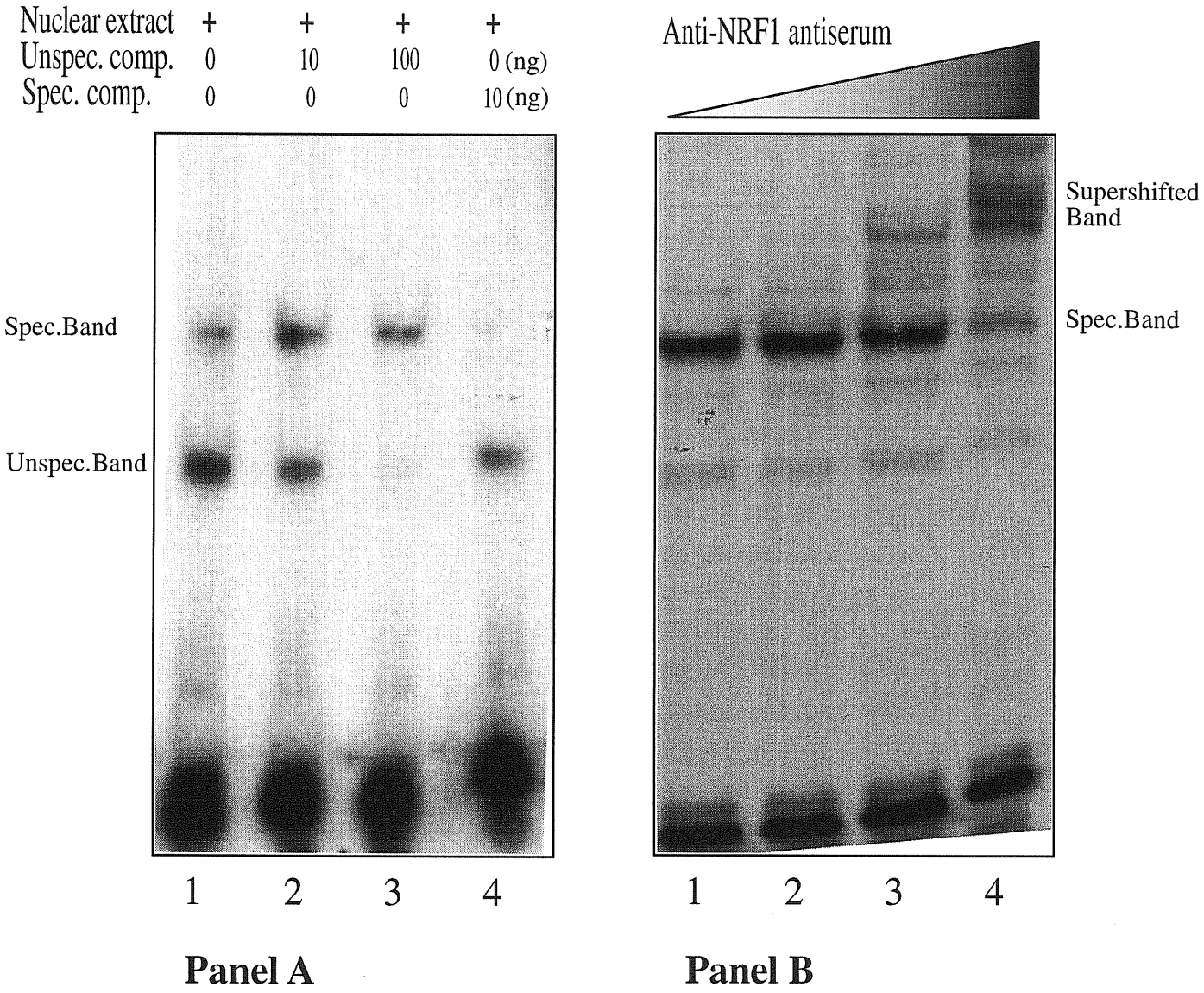


Figure 3.8.2 NRF1 from HL60 nuclear extracts is able to bind to its putative binding site in the lamin B2 gene origin. **Panel A,** band-shift experiment using the NRF1 binding site of the lamin B2 origin as a probe. 10 µg of nuclear extract were used in each reaction with 1 ng of probe. Lane 1, retarded bands; lanes 2-3, unspecific competition with increasing amounts of a 25 bp oligonucleotide; lane 4, specific competition with the cold probe. **Panel B,** supershift experiment obtained incubating the band shift reaction with a polyclonal antiserum anti-NRF1. Lanes 1 to 4, increasing amounts of antiserum.

These observations demonstrated that all the protections detected *in vivo* can be reproduced *in vitro* using crude nuclear extracts (with the exception of the UBF site that is only partially protected) while no extra protections were detected. In at least two cases (USF and NRF1) it was also possible to demonstrate that the factors binding *in vitro* correspond to the factors expected to bind to the bases of the nucleotide sequence.

3.9 *IN VITRO* PROTEIN-DNA INTERACTIONS AT THE hOC SITE IN HL60 EXTRACTS

Experiments aimed to identify the factor/s responsible for the interaction in the region of the lamin B2 origin were also performed. Firstly the 75 bp probe hOCBS-75 was used in a gel retardation assay. The probe exactly spans over the 70 bp protected *in vivo* in unsynchronized cells. Using different extracts (i.e. Dignam extract, 0.4 M extract, 0.8 M extract, 1.2 M extract, S-100 cytoplasmic extract, 1+1 extract) and different binding conditions, no specific retarded bands were detected. On the contrary, gel retardation assays performed using longer probes hOCBS-280 or hOCBS-120 (containing respectively 280 and 120 bp of the origin region) gave extremely complex pattern of bands and in many cases all of the probe was retained in the wells also in presence of high amounts of unspecific competitor (not shown).
In this situation *in vitro* DNaseI footprinting resulted more convenient as an assay using the probe hOCBS-lower to study the interactions at the strand protected *in vivo* in unsynchronized cells. In this case also, several different protein extracts and binding conditions were tested. A DNA-binding activity (Act-1) spanning the region protected *in vivo* was found in Dignam[Na] extracts from HL60 cells after an incubation for 20 min. at RT in buffer F. Addition of ATP to the binding reaction resulted in an enhancement of the binding activity and it should be considered that the relevant experiments were performed in the presence of ATP, if not differently specified.

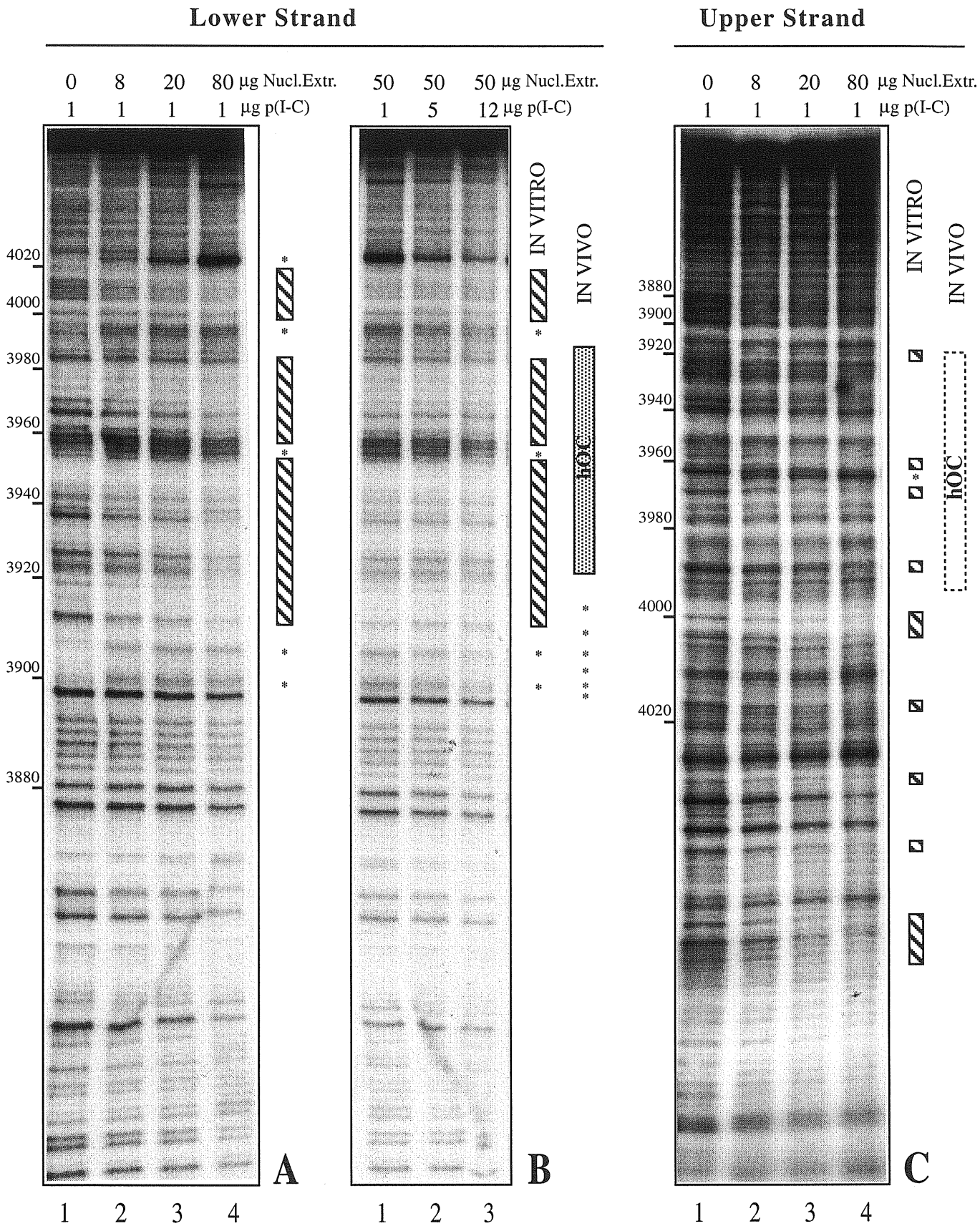


Figure 3.9.1 *In vitro* DNase I footprinting assays performed using Dignam extracts from HL60 cells. **Panel A**, lower strand: lane 1, naked DNA; lanes 2-4, increasing amounts of crude nuclear extract. **Panel B**, lower strand: lanes 1-3, increasing amounts of unspecific competitor. **Panel C**, upper strand: lane 1, naked DNA; lanes 2-4, increasing amounts of crude nuclear extract. DNaseI protections (boxed regions) and hypersensitive sites (stars) are indicated.

The result of some footprinting reactions performed in presence of increasing amounts of Dignam extract is reported in figure 3.9.1 panel A. The higher the amount of extract in the reaction, the stronger is the protection of the probe particularly in the region corresponding to the area protected *in vivo.*

The specificity of the DNA-protein interaction was tested by a competition assay. Increasing amounts of the unspecific competitor poly d(I-C) were added to similar footprinting reactions (figure 3.9.1 panel B). In this case, the excess of poly d(I-C) can compete with the factors that bind outside the region of 70 bp (in particular the unprotected band at the map position 4030) but not with the factors that bind the same one. This result indicates that Act-1 binds DNA in a sequence specific manner.

Footprinting reactions performed, in the same conditions, with a probe for the other strand of the same region (hOCBS-upper) detected only small protections (figure 3.9.1 panel C) accordingly with the result obtained *in vivo.* Furthermore, an analogous study performed with a Dignam extract obtained from differentiated (non replicating) HL60 cells did not detect any protection inside the origin region.

This data strongly suggests that the DNA-binding activity found in the HL60 nuclear extracts could be responsible for the binding activity found *in vivo.*

3.10 THE hOC-BINDING "Act-1" IS ALSO PRESENT IN PROTEIN EXTRACT FROM OTHER CELL LINES

Nuclear extracts from other cell lines were tested to clarify whether the activity found in HL60 cells was peculiar to this cell line or was a more general one. Dignam extracts from cell lines related to HL60 promyelocytic leukemia cells, as U937 and K562 cells (hystiocytic lymphoma and myeloid leukemia lines respectively), or from cells with different hystological derivation, as HeLa and IMR90 cells (epithelial and lung fibroblast lines respectively), were tested by *in vitro* footprinting assay.

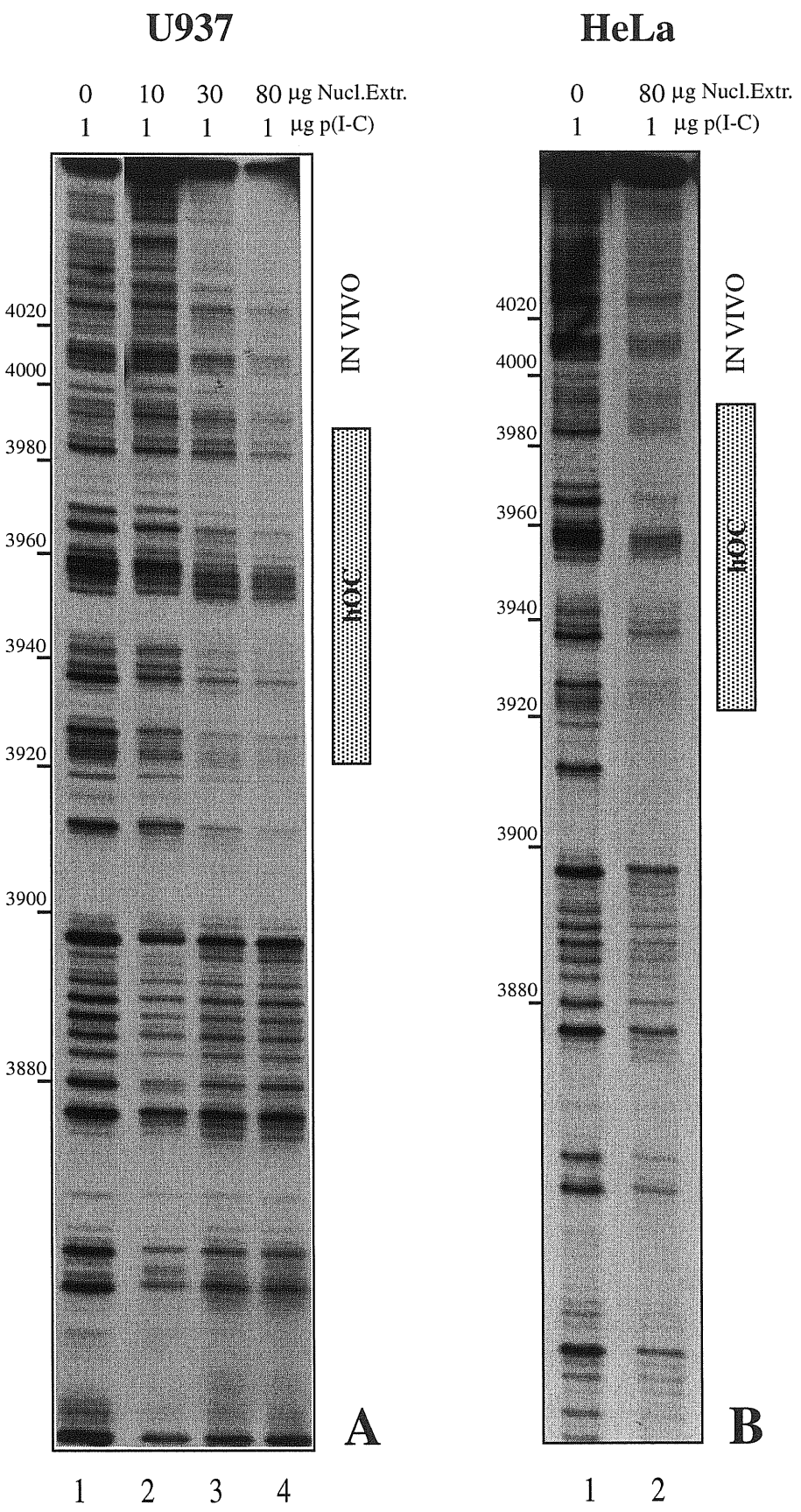


Figure 3.10.1 DNaseI *in vitro* footprinting using protein extracts from different cell lines. **Panel A**, Dignam extract from U937 cells. Lane 1, naked DNA; lanes 2-4, increasing amounts of extract. **Panel B**, protein extract from HeLa cells. Lane 1, naked DNA; lanes 2, increasing amounts of extract.

In all cases, using Dignam[Na] extracts, a DNA-binding activity was detected in the 70 bp region with only minor differences, in the protection pattern, between the different cell lines (figure 3.10.1 for some examples). The protection is a bit larger compared to that found in unsynchronized cell by DNase I genomic footprinting. On the other hand, protein extracts were prepared from asynchronous cells in which most of the cells are in G1. The DMS genomic-footprinting on G1 cells detected a protection of roughly 200 nt.. Therefore, it is reasonable to find, in protein extracts, an activity that protect more than 70 nt. in the hOC site region.
The finding that the same DNA-binding activity is present in different cell lines also suggests that this activity could be constitutive in actively replicating cells.

3.11 OTHER hOC-BINDING ACTIVITIES

Testing the protein fraction recovered by extracting with 1 M NaCl and 1% NP40 the nuclear pellet obtained at the end of the Dignam[Na] procedure (1+1 Extract) a second activity able to bind the hOC-biding site was detected (Act-2)
I this case the protection is on both strands and it is limited to roughly 30 bp, corresponding to the left portion of the 70 bp protected *in vivo* in unsynchronized cells (figure 3.11.1). The same region has been previously described as topoisomerase II consensus sequence (Falaschi *et al.* 1988).
The same activity has been detected in all cell lines tested up to now (HL60, HeLa, U937, K562, IMR90) suggesting that also in this case a general factor could be responsible for this protection.

3.12 ONE-HYBRID SCREENING FOR hOC SITE-BINDING PROTEINS

The head to tail trimer of the hOC binding site (nt. 3921-3993 HUMLAMB2B) was used as a bait in a one-hybrid screening to identify the c-DNA of proteins able to interact with the human lamin B2 gene origin. The plasmid pJL638 was utilized to screen a cDNA library of HeLa cells.

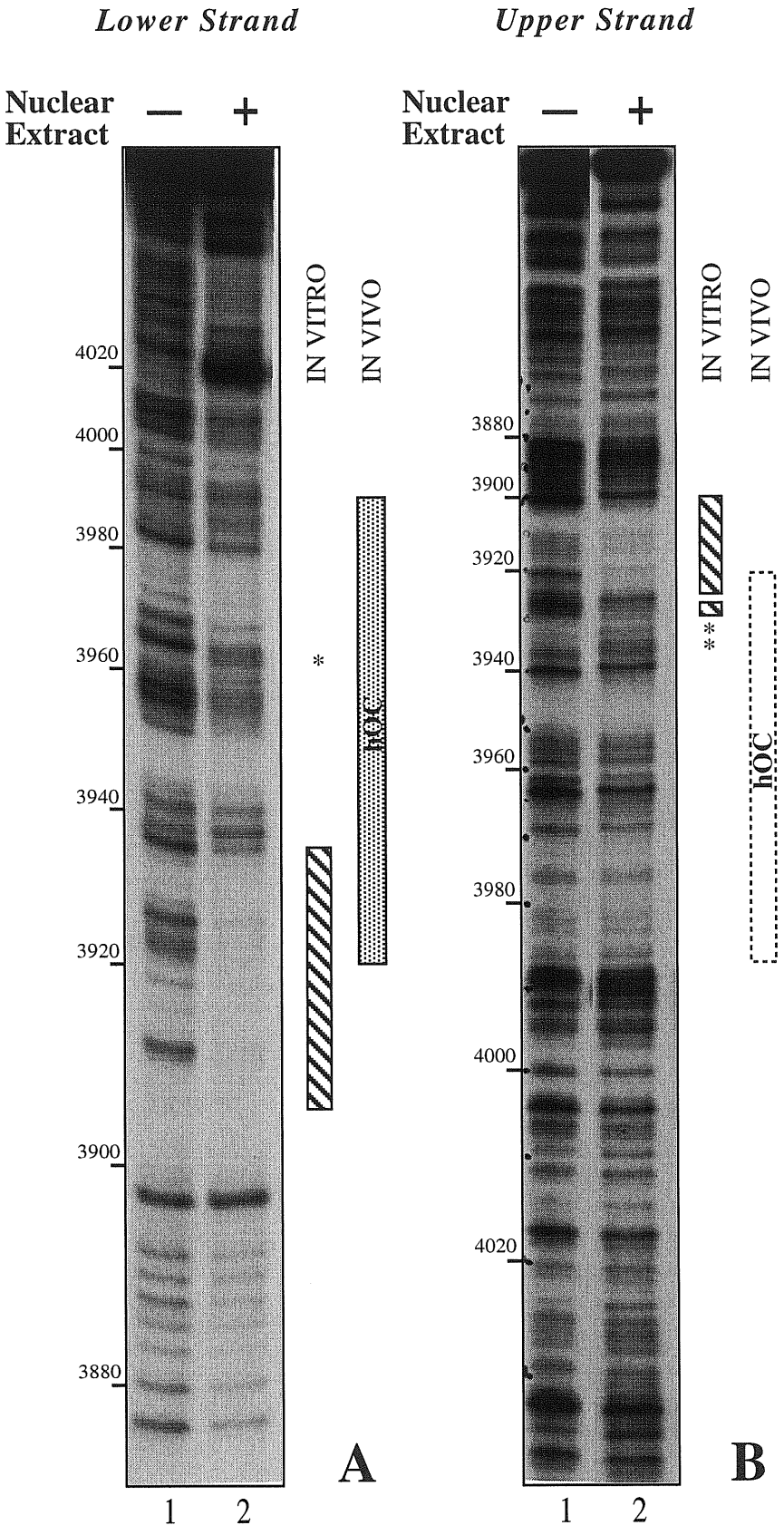


Figure 3.11.1 *In vitro* DNase I footprinting assays. Protein extracts obtained with high salt and detergents from HL60 cells, show the presence of a very strong specific DNA-binding activity that protects only a part of the hOC region on both strands. **Panel A**, lower strand. Lane 1, naked DNA; lane 2, 6 μg of protein extract. **Panel B**, upper strand. Lane 1, naked DNA; lane 2, 6 μg of protein extract. DNase I protections (boxed regions) and hypersensitive sites (stars) are indicated.

The screening was performed as follows. Plasmids were extracted from the clones that produced a blue colony in the first plating. These plasmids were used to transform other yeast cells containing pJL638/ori+. Only the clones that were able to produce blue colonies also in the second plating were considered as positive.
Another control was done plating the positive clones after a period of growth without selection for the plasmid. The clones producing blue colonies in this third plating were discarded.
All the positive clones (Table 3.12.1 column K699/ori+) were inserted in K699 yeast cells in the presence of a pJL638 without the hOC binding site (Table 3.12.1 column K699/ori-) in order to verify that the cDNAs were able to activate the Lac Z reporter only in presence of the lamin B2 origin. The result of the screening is shown in table 3.12.1; 73 clones were identified as positive using the described criteria.

Clone	Number of	K699/ori+	K699/ori-	Identified
3	5	blue	white	Hox C-10
134	67	blue	white	Hox C-13
477	1	blue	white	Hox A-13
627A	1	white	white	
667	2	white	white	
704	1	blue	blue	

Table 3.12.1 Results of the yeast one-hybrid screening for human proteins binding to the 70-bp protected region in the lamin B2 origin of replication. Results from the screening of 7.4 x 10 transformants.

67 out of these 73 was representing the same cDNA, a second cDNA was present in five clones and a third cDNA was present as a single clone. All these three cDNA were identified on the base of a database research for DNA sequence homology as proteins of the human HOX family.
For one of these proteins (HOX C13) the full length clone was obtained while for the other two (HOX C10 and HOX A13) only the partial cDNAs were isolated.
HOX C13 was used to produce a GST-fusion protein cloning its cDNA in the expression vector pGEX-2T (GST-HOX C13) while HOX C10 was produced as a histidine-tagged protein cloning the cDNA in pTrcHis-A (His-HOX C10). The fusion products of these two proteins were used to characterize the *in vitro* binding to the hOC site.

3.13 *IN VITRO* INTERACTIONS OF THE HOX PROTEINS WITH THE hOC SITE

The DNA binding activity of GST-HOX C13 and His-HOX C10 was studied *in vitro* by band-shift and DNase I footprinting assays.
Figure 3.13.1 shows the result of a gel retardation assay performed using 1 ng of a 75 bp probe encompassing the hOC site (hOCBS-75) and 10 ng of recombinant GST-HOX C13. Two retarded bands are present in the uncompeted reaction. Both the bands result as being specific have been competed by an increase in the amounts of cold probe but not by increasing amounts of a double-strand oligonucleotide of the same length. The result can be explained with the presence of two binding sites for HOX C13 inside the lamin B2 gene origin, alternatively the slower complex could be due to a dimerization of the fusion protein as a consequence of protein-protein interactions between the GST domains.
DNase I footprinting reactions were performed using the same conditions adopted for the gel retardation assay. Figure 3.13.2 panel A and panel B show the protection pattern obtained for the two strands of the hOC site (probes hOCBS-upper and hOCBS-lower). Within the 70 bp region, both strands exhibit two separate cluster of protections, surrounded by hypersensitive sites and separated by an unprotected region. This pattern strengthens the hypothesis of presence of two

different binding sites for HOX C13 in the hOC site region (figure 3.13.2 panel C).
In the same conditions, no protections were detected using probes for other regions of the genome (as for example the probe 4SR-lower encompassing the ppv1 promoter) also in presence of high amounts of recombinant protein (not shown). This result confirms that HOX C13 binds the DNA in a sequence specific manner.

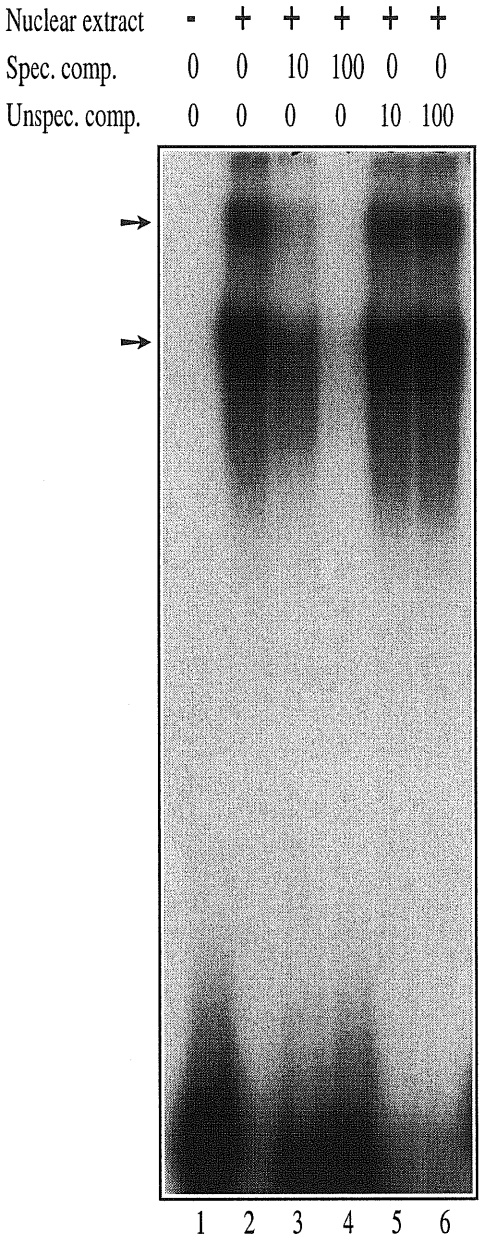


Figure 3.13.1 Band-shift assay with GST-HOX C13 fusion-protein (25 ng) and hOCBS-75 as probe (1 ng). Lane 1, free probe; lane 2, uncompeted control; lanes 3-4, increasing amounts of specific competitor; lanes 5-6, increasing amounts of unspecific competitor.

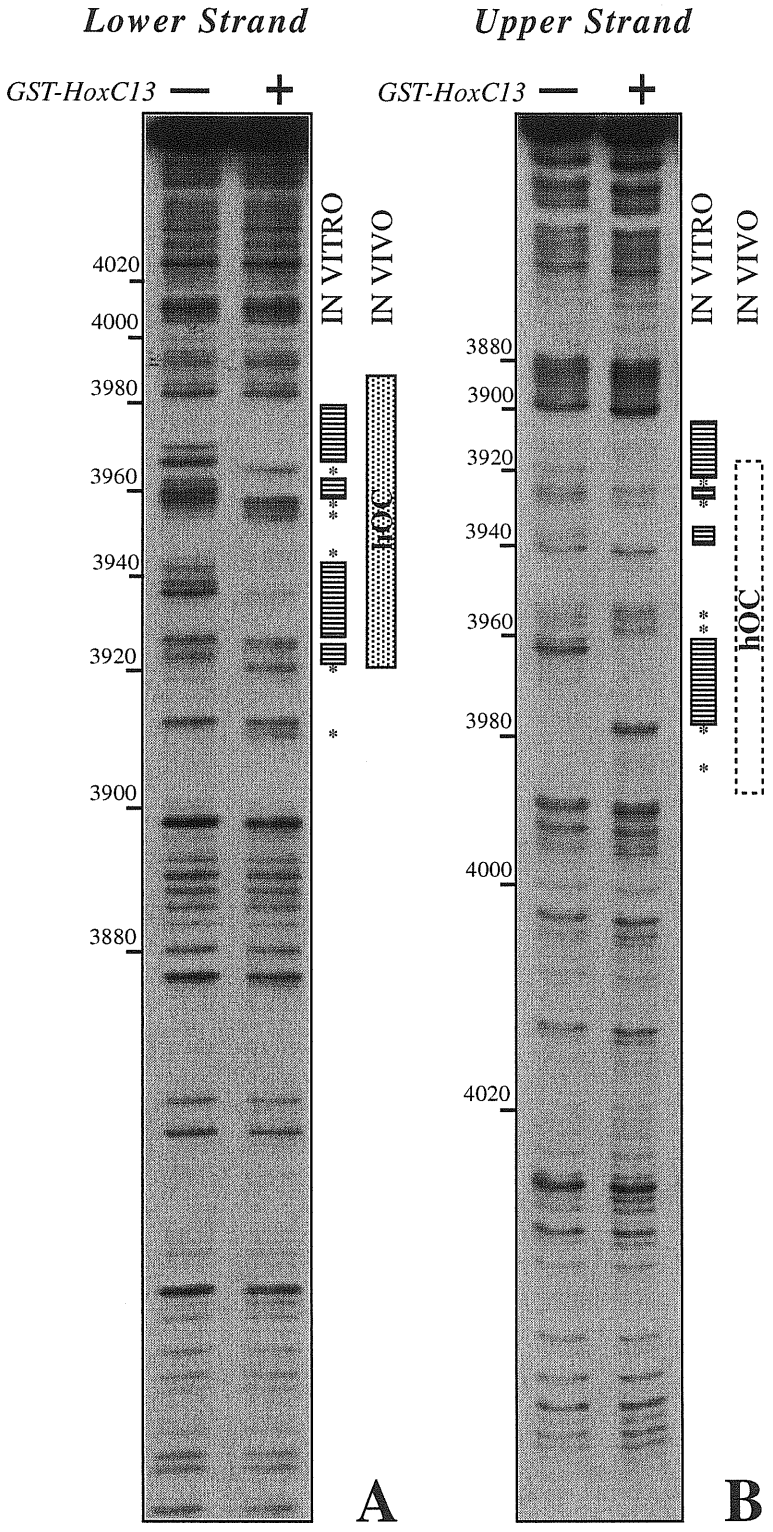
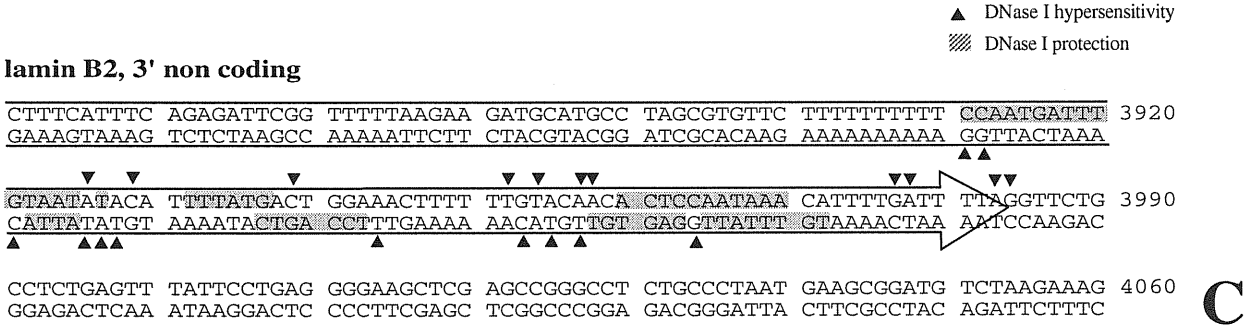


Figure 3.13.2 *In vitro* footprinting experiments with GST-HOX C13. **Panel A,** *in vitro* DNaseI footprinting experiment on the hOC DNA binding site (lower strand). Lane 1, naked DNA; lane 2, 10 ng of GST-134. DNase I protections (boxed regions) and hypersensitive sites (stars) are indicated. **Panel B,** *in vitro* DNaseI footprinting experiments on the hOC DNA binding site (upper strand). Lane 1, naked DNA; lane 2, 10 ng of GST-134. **Panel C**, summary of the DNA-protein interactions produced by GST-HOX C13 at the hOC site.



The footprinting experiments were repeated with a probe having the same sequence of hOCBS-lower except for a mutation obtained by a permutation of 40 nucleotides encompassing one of the two GST-HOX C13 binding sites (hOCmut-lower). In this case, while the interaction at the retained site was almost unmodified, no protections were detected at the mutated site (figure 3.13.3 panel A). This data strongly supports the idea that HOX C13 do not require the whole 70 bp sequence to bind specifically to the hOC site and that the lamin B2 origin of replication contains at least two binding sites for this protein.

In all cases, the binding of HOX C13 seemed to be strongly influenced by the presence of divalent cations. Figure 3.13.3 panel B shows the result of DNase I *in vitro* footprinting reactions performed with the addition of different amounts of Mg^{++} and Zn^{++} cations to the binding reactions. When divalent cations are added (20 mM $MgCl_2$ and 0.5 mM $ZnCl_2$), DNase I hypersensitive sites surround the protected sites (as shown in the experiments previously described). When Zn^{++} is not added to the reactions and the Mg^{++} concentration is only 6 mM, no DNaseI hypersensitive sites were detected. Moreover, less protein is needed to produce a full protection of the probe when divalent cations are added.

Gel retardation assays were performed also for His-HOX C10. In this case, only a specific retarded band was identified (figure 3.13.4) implying the presence of a single binding site for this protein. A large excess of recombinant protein is required to shift the same amount of DNA probe compared to what was found for HOX C13. This probably explains why a clear protection was not identified by DNase I *in vitro* footprinting using the same conditions used for GST-HOX C13.

His-HOX C10 was produced from a partial cDNA that contains only a small part of the protein (roughly the homeodomain and the C-terminal portion of the protein). The lack of the N-terminal could impair the stability of the DNA-protein complexes. For these reasons, further experiments with the full length protein are needed to characterize the interactions of this protein with the DNA.

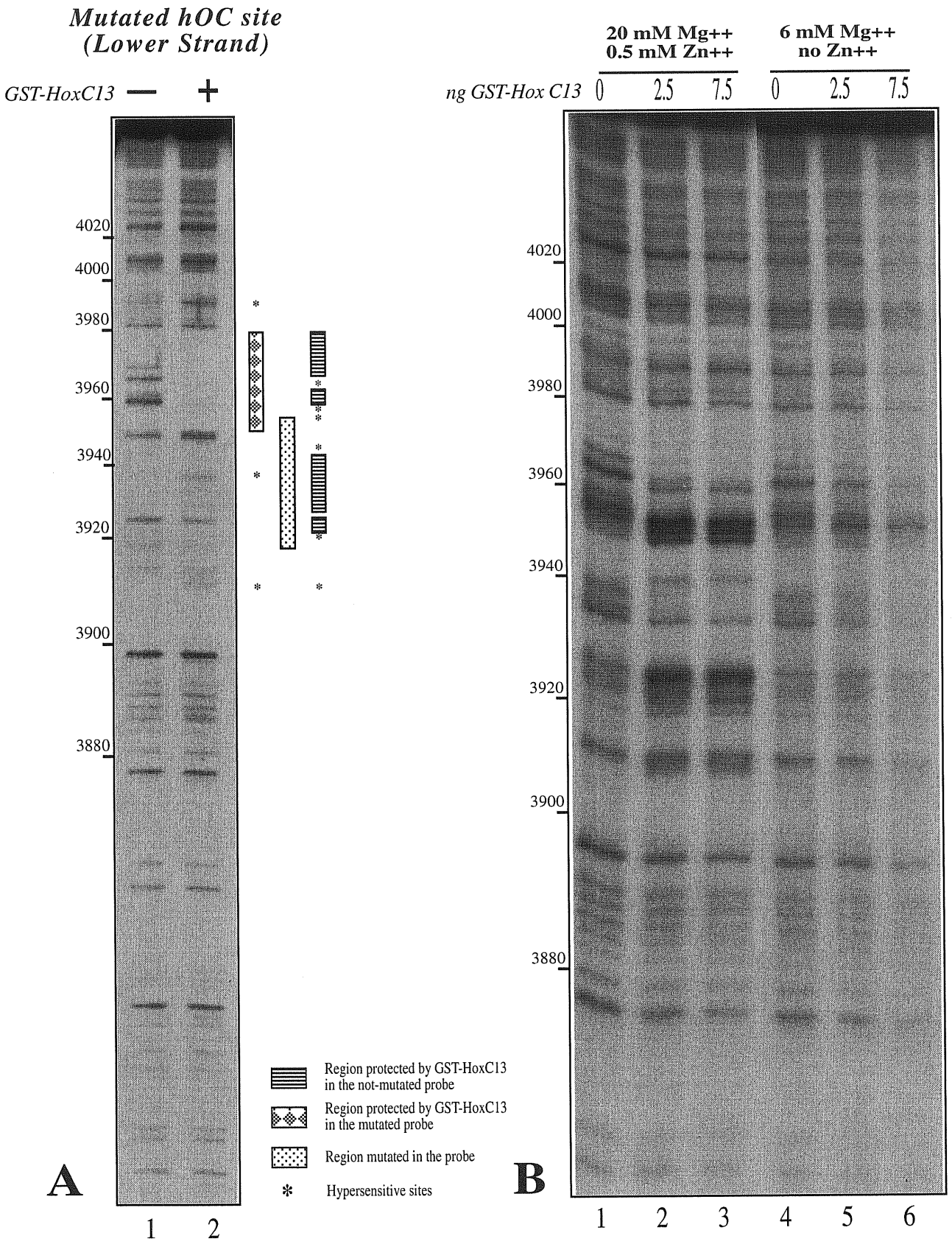


Figure 3.13.3 Panel A, *in vitro* footprinting using a hOC binding region mutated by a permutation of 40 nucleotides centered on one of the two groups of protections. This mutation does not affect the binding of GST-HOX C13 to the other putative binding site. Lane 1, naked DNA; lane 2, 10 ng of GST-HOX C13. **Panel B** Effect of divalent cations on GST-HOX C13 binding to the hOC site. Lanes 1 and 4, naked DNA; lanes 2 and 3, increasing amounts of GST-HOX C13 in the presence of Zn++ and Mg++; lanes 5 and 6, increasing amounts of GST-HOX C13 in the absence of Zn++ and with low concentrations of Mg++.

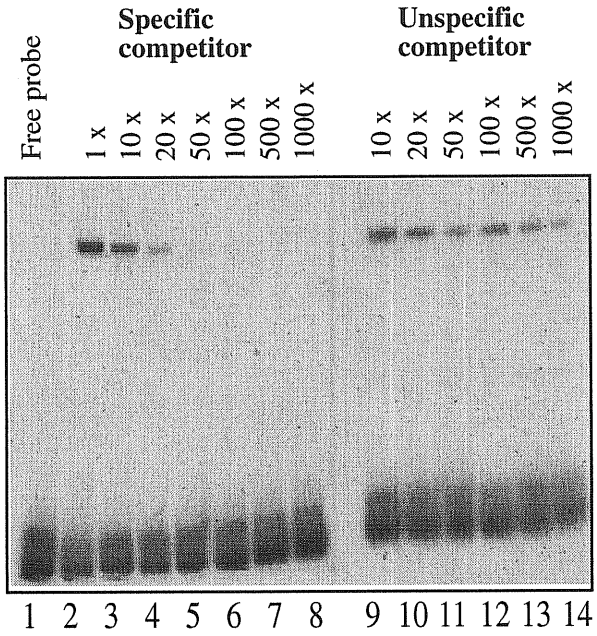


Figure 3.13.4 Gel retardation assay using His-HOX C10 fusion protein. Lane 1, free probe (hOCBS-75); lanes 2 to 8, increasing amounts of specific competitor; lanes 9 to 14, increasing amounts of unspecific competitor.

3.14 PURIFICATION OF SOME hOC-BINDING ACTIVITIES

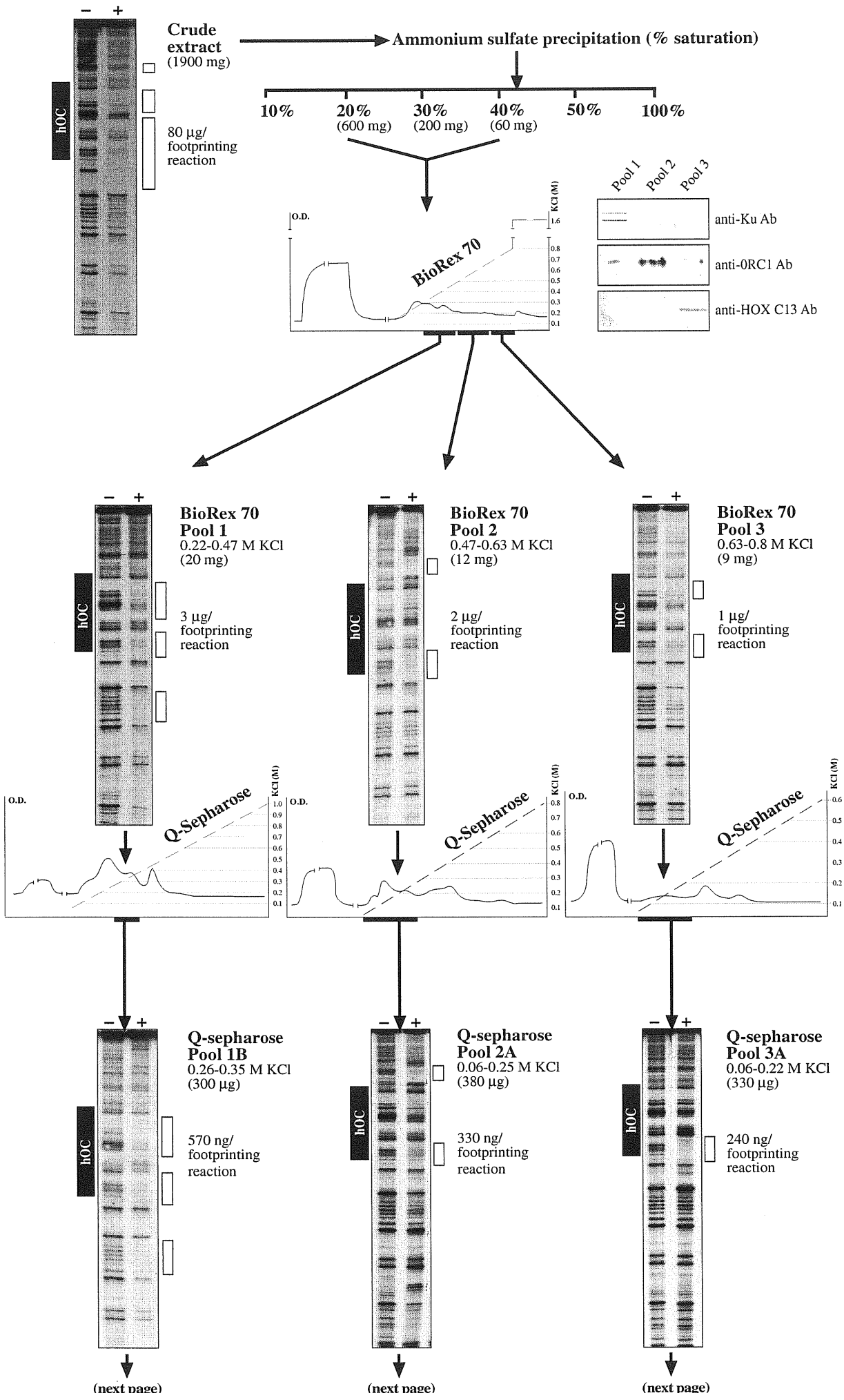
Several attempts were made to purify the factor/s corresponding to the hOC-binding activity found in Dignam[Na] extracts.

Initiating with a small amount (10-20 g) of exponentially growing fresh cells (HL60 of HeLa), different columns were used to fractionate the extract (table 3.14.1 for a list) but in all cases no activity was recovered from the columns or in the different flow through. The result could be explained by the instability of the protein/s responsible for the activity or by the disgregation of a protein complex. In any case complementation assay failed to give any indication of the reconstitution of the activity.

Table 3.14.1

Heparine Sepharose
BioRex 70
Mono-S
Mono-Q
DEAE Sepharose
Hydroxyapatite
Gel filtration

In order to avoid problems arising from the manipulation of small amounts of protein extract, HeLa cells from frozen pellets stored at -80 ^{0}C were chosen as base material for the protein purification.
The extraction procedure was optimized to extract both the activities ("activity 1" and "activity 2") in the same pool. No activity was recovered in the Dignam extract when NaCl was substituted with KCl (Dignam[K] extract) in the extraction procedure while "activity 1" was recovered in the 1+1 Extract (section 2.7) partially mixed with "activity 2".
From 200 g (wet weight) of frozen HeLa cells, 800 ml of 1+1 extract (1900 mg) was obtained. The extract was centrifuged at 4 ^{0}C for 16 hours at 36,000 g to eliminate the genomic DNA. After dialysis in buffer 1, the extract was differentially precipitated using increasing amounts of a saturate $(NH_4)_2SO_4$ solution, as described in figure 3.14.1. Protein pellets were resuspended, dialyzed in buffer 1 and the active fractions were pooled together.
The active pool was loaded on a BioRex 70 column (weak cation exchange) and fractionated using a linear KCl gradient (0.1-0.8 M) followed by a high salt wash with 1.6 M KCl; 10 ml fractions were recovered. Using this procedure, ~95% of the loaded proteins were removed in the wash whereas the hOC-binding activities were eluted between 0.22 and 0.8 M KCl.


Crude extract (1900 mg)
Ammonium sulfate precipitation (% saturation)
10%
20% (600 mg)
30% (200 mg)
40% (60 mg)
50%
100%
80 µg/ footprinting reaction
hOC
BioRex 70
Pool 1
Pool 2
Pool 3
anti-Ku Ab
anti-0RC1 Ab
anti-HOX C13 Ab
BioRex 70 Pool 1
0.22-0.47 M KCl
(20 mg)
3 µg/ footprinting reaction
BioRex 70 Pool 2
0.47-0.63 M KCl
(12 mg)
2 µg/ footprinting reaction
BioRex 70 Pool 3
0.63-0.8 M KCl
(9 mg)
1 µg/ footprinting reaction
Q-Sepharose
O.D.
KCl (M)
Q-sepharose Pool 1B
0.26-0.35 M KCl
(300 µg)
570 ng/ footprinting reaction
Q-sepharose Pool 2A
0.06-0.25 M KCl
(380 µg)
330 ng/ footprinting reaction
Q-sepharose Pool 3A
0.06-0.22 M KCl
(330 µg)
240 ng/ footprinting reaction
(next page)
(next page)
(next page)

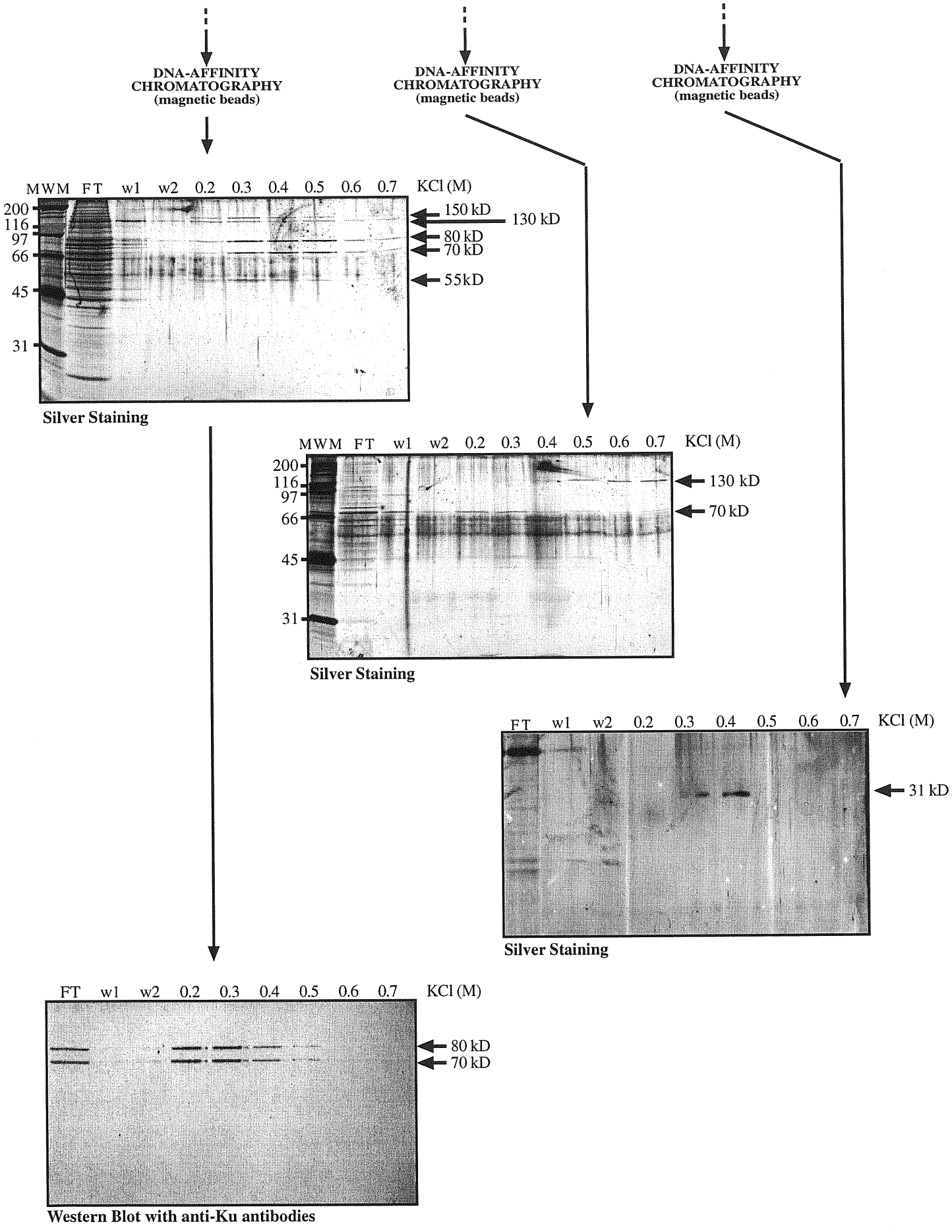


Figure 3.14.1 Purification flow chart of some hOC-site binding activities. Section 3.14 for description.

From the footprinting analysis of the different fractions, three different activities were identified and collected in three different pools (BioRex 70, pools 1-3 figure 3.14.1). A peculiarity of these footprinting activities is that all show an incomplete protection pattern and a decreased stability in competition assays compared to Act-1 and Act-2. This data suggests that Act-1 and Act-2 protection patterns could be the result of the binding of several protein factors to form a stable DNA-protein complex. The reduced stability of the DNA-protein complexes, formed by the partially purified activities, would also explain the failure of the previous trials to purify Act-1 and Act-2 starting from a lower amount of extract. The different protein composition of the pools was demonstrated by a western assay using three different antibodies (anti-Ku antigen, anti-ORC 1 and anti-HOX C13). Each antibody recognized proteins from a single pool only.

BioRex pools 1, 2 and 3 were further fractionated (separately) using a Q-Sepharose column (weak anion exchange). The bound proteins were eluted with linear gradients of KCl (0.06-1.0 M KCl for pool 1, 0.06-0.8 M KCl for pool 2, 0.06-0.6 M KCl for pool 3) collecting 5 ml fractions. All the DNA-binding activities were eluted in the range 0.06-0.35 M KCl and maintained, in DNase I footprinting assay, the protection pattern found in the previous purification step.

The footprinting assay does not allow a precise quantitation of the DNA-binding activities. Nevertheless it is possible to observe, during the purification procedure, a reduction in the amount of protein preparation needed to fully protect 1 ng of DNA probe (from 80 µg required in the crude extract to ~300 ng after the Q-Sepharose column). This means that the concentration of the proteins responsible for the activities, increased almost 300 times during this purification procedure.

The three active pools were further fractionated by an affinity step using magnetic beads coupled to 280 bp long DNA region containing hOC site (3795-4074 HUMLAMB2B). After two washes in binding buffer (figure 3.14.1, w1 and w2) the bound proteins were eluted using buffers with increasing concentration of KCl.

Five different polypeptide were purified from the first pool. Analysis on

a 12% SDS-Page revealed molecular weights of approximately 55, 70, 80, 130 and 150 kD, while the smeared bands visible in the silver stained gels (55-65 kD) were artifacts. Moreover, two proteins of 70 and 130 kD from the second pool and an approximately 31 kD protein from the third pool were also recovered.

Two of the proteins recovered from the first pool were identified by western blot as the subunits of the Ku antigen. This heterodimer binds the DNA non specifically regardless of its sequence with a preference for the free extremities and was probably purified due to its high concentration in the nuclear extract.

The purified Ku heterodimer is not able to produce a specific footprint using hOCBS-lower as a probe (not shown), therefore, the protection pattern seen with pool 1 is probably due to the other proteins of the pool.

Trials to identify the other purified proteins are currently in progress.

DISCUSSION

4.1 COMPETITIVE PCR FOR QUANTITATION OF NUCLEIC ACIDS IN TRACES

The factors that affect any PCR amplification can be roughly grouped in four categories: I) quality of the DNA sample used for the amplification, II) choice of the primers, III) parameters used for the amplifications, IV) tube-to-tube variations. All these factors contribute in making most of the quantitative-PCR techniques extremely challenging, decreasing the reproducibility and as a consequence the reliability of any quantitation.

The principle of competitive PCR relies on the assumption that the competitor and target templates compete for amplification and are equally subjected to all the factors affecting amplification. Therefore, competitor and target must share most of their nucleotide sequence and in particular the primer recognition sites. Additionally, the amplification products obtained from the two molecular species must be similar in size. According to these principles, the ratio between the competitor and the template amplification products is linearly related to the number of competitor molecules initially added to the reaction. Since the number of competitor molecules is known, the concentration of each investigated genomic segment in the sample can be calculated easily from the interpolation of the regression line.

The superiority of the competitive-PCR technique compared to other semi-quantitative PCR strategies is summarized in figure 4.1.1.

In practice, the quantitations obtained in this way have the following advantages:

- are not affected by differences in amplification efficiency among primer pairs
- are not affected by the yield of the PCR reaction
- are not affected by the presence of non specific PCR amplification products
- do not require the maintenance of the exponential phase of PCR amplification
- allow quantitation of the absolute number of molecules

	considers sample quality	considers unpredictable variations in amplification efficiency	requires the exponential phase of amplification	requires that both primer pairs have the same kinetics of amplification
Comparison with the amplification of varying dilutions of a reference sample in parallel reactions	*no*	*no*	*yes*	-
Amplification of varying dilutions of the sample to determine the limiting dilution	*no*	*no*	*no*	-
Simultaneous amplification of sample with two different pairs of primers	*yes*	*yes*	*yes*	*yes*
Simultaneous amplification of the sample and of a competitor template with the same pair of primers *(competitive PCR)*	*yes*	*yes*	*no*	-

Table 4.1.1 Advantages of competitive PCR

These qualities make all the procedure easy and highly reproducible and it is for these reasons that we choose competitive PCR to quantitate the presence of different target sequences in the nascent DNA preparations.

4.2 A MAPPING TOOL FOR ORIGINS

The only way to identify DNA origins of replication in higher eukaryotes is by somehow mapping the replication activity. The method we used is based on the quantification of newly synthesized DNA molecules in exponentially growing cells. The principle of the technique is in choosing a group of marker sequences in the region in which we want map an active origin, the relative abundance of a marker in the nascent DNA is directly proportional to its closeness to the origin.

The advantages of the competitive PCR technique which we have developed compared to other mapping strategies previously illustrated, can be summarized as follows:

- allows quantitation of the absolute number of molecules
- does not rely on the study of amplified sequences
- can be applied to intact cells without the need of permeabilization for labeled precursors
- does not require cell synchronization
- does not depend on the use of metabolic inhibitors with possible pleiotropic effects
- is able to identify origins at high resolution.

Firstly, we demonstrated that it could be possible to map an active origin of replication by this method. For early experiments we chose the SV40 origin of replication. An actively replicating plasmid, containing the whole SV40 genome, was transfected in COS cells and nascent DNA was extracted. Competitive PCR on purified nascent DNA demonstrated a clear enrichments for the markers close to the SV40 origin, confirming the validity of the approach.
The same procedure was also used to map the *ori* activity in plasmids containing the SV40 minimal origin (pSV2neo and pSV2neo-inv). These findings demonstrated that, in some constructs, the replication can proceed in a monodirectional way as a consequence of the plasmid features. Moreover they indicated that competitive-PCR mapping can also be used to define the direction of the replication forks from a known DNA origin of replication.

4.3 MAPPING OF THE LAMIN B2 GENE DOMAIN ORIGIN OF REPLICATION

The high sensitivity of this approach allowed us to map an origin of replication contained in a non amplified region of the human genome. The 14 Kb lamin B2 gene domain was chosen for this purpose.
In this case, nascent BrdUrd-labeled DNA obtained from HL60 cells, synchronized in early-S, was used for the quantitations. An active origin of replication was identified in a region between the 3'-terminal

of the transcriptional unit for lamin B2 and the promoter for another downstream positioned transcript close to the target sequence named B48.
The result obtained with synchronized cells was also seen in unsynchronized cells thus confirming that it was not due to an artifactual origin activation induced by aphidicolin. Furthermore, analogous studies performed using differentiated HL60 cells (not proliferating) or unstimulated, quiescent lymphocytes (Kumar et al 1996), gave no evidence for lamin B2 *ori* activity.
These observations attest, once again, the validity of the method developed for origin mapping. On the other hand, the results reported in sections 3.5 and 3.6 indicate that the region, amplified by the primer set B48 is only ten times more represented than external sequences, also when these are far from any known replication origin. This small enrichment was unexpected but it could be well explained if we assume that the Okazaki fragments are ligated not immediately and in series but in patches. In this case, the "patches" could contaminate even the size-selected pool of nascent DNA. This, together with the technical sources of error, would very easily increase the background and reduce the signal/noise ratio to the level we saw.

Another fascinating hypothesis derives from a recent study on another origin mapped with the same procedure. While the B48 marker, that detects the lamin B2 *ori*, shows a 6-10 fold enrichment in its relative abundance compared to other genomic sequences, the SK4 marker, in the AAV insertion site origin, shows a "signal" up to 40-fold the background in activated peripheral blood lymphocytes (Kumar unpublished). These differences in the relative enrichment could reflect differences in the frequency of use of certain origins thus giving to the mechanism of origin activation the typical features of a stochastic event.
On the contrary, these differences could be the result of small differences in the distance of the chosen markers from the origins or could be a consequence of the different length of the two markers that produce a different sensitivity in the quantitation of a nascent DNA pool of short size.

The use of BrdUrd pulse-labelling of newly replicated DNA and as a consequence of the anti-BrdUrd antibodies in the process of nascent DNA preparation was originally introduced with the aim of obtaining a high signal-to-noise ratio. It was later seen that the method can be further simplified and that the use of BrdUrd and immunoaffinity chromatography can be omitted.
This result can be attributed to careful size selection of the nascent DNA strand pool in use as well as to the reduction in the background due to the high molecular weight DNA breakage. In fact, BrdUrd-substituted DNA is prone to breakage under a mild stressing condition.
Thanks to this simplification, the mapping of the origin activity was made easier and faster allowing the mapping of the lamin B2 origin in several human cell lines with different histological derivation. The distribution of DNA segment abundance in samples of nascent DNA clearly suggests that the lamin B2 *ori* is active in all the analyzed cell lines, behaving as a constitutive DNA origin of replication.
The results reported above were obtained in established cell lines that replicate at a continuous rate in the absence of inhibitory treatments. In addition to these findings, it was also recently demonstrated that the same origin is active in primary human cells (peripheral blood lymphocytes activated with phytohemoagglutinin and interleukin-2) (Kumar *et al.* 1996) confirming the physiological role of the lamin B2 origin of replication. Table 4.3.1 shows the cases in which the lamin B2 origin of replication is active or inactive.
The universal use of this origin in all the analyzed cells is not surprising. In many cases, there is a close relationship between transcriptional activity and initiation of DNA replication (DePamphilis, 1993). This phenomenon is probably related to the accessibility of chromosomal domains to proteins involved in both processes. The observation that lamin B2 origin lies in a constitutively expressed gene domain, coding for a house-keeping protein (Biamonti *et al.* 1992b), predicts origin usage by all cells.
Likewise, other regions of the genome with differential transcriptional activity in different cell lines (for example, the human immunoglobulin H-chain locus) show a different origin usage depending on the transcriptional activity of the domain (Hatton *et al.* 1988).

Type of cell line	*Origin Activity*
HL60 exponentially growing	+
HL60 differentiated	—
HeLa exponentially growing	+
U937 exponentially growing	+
SKNBE exponentially growing	+
SKNMC exponentially growing	+
IMR32 exponentially growing	+
IMR90 exponentially growing	+
Resting lymphocytes	—
Activated lymphocytes	+

Table 4.3.1 Activity of the lamin B2 origin of replication in different cell lines and situations. The results show that the origin is active in all the proliferating cells so far tested.

All these results strongly suggest that the transcriptional state of a genomic region can have strong consequences on the origin selection and/or activation events.

This new procedure for origin mapping provides a simple tool for the study of the DNA replication process in single copy domains of unsynchronized growing cells. After the lamin B2 domain, other origins were mapped in various regions of the human and hamster genomes (Kumar unpublished; Pellizon *et al.* 1996).
In all these cases, it was found that the same precise and rather narrowly defined region of the chromosome works as an *ori*. This suggests that even in higher eukaryotes *oris* represent precise areas where the generalized inhibition of replication initiation caused by the standard chromatin structure is locally abolished by the interaction with specific protein factors.

4.4 CHROMATIN STRUCTURE AT THE LAMIN B2 GENE ORIGIN OF REPLICATION

DNase I hypersensitive-site mapping in a genomic region encompassing the human lamin B2 origin of replication revealed that *ori* is located in a highly accessible chromatin region characterized by the presence of two closely spaced DNase I-hypersensitive sites. This finding is consistent with numerous other reports on the association of nuclease-hypersensitive sites with replication origins in the genomes of viruses, single-cell eukaryotes, flies, frogs, birds and mammals (Siebenlist *et al.* 1984; Cereghini and Yaniv 1984; Palen and Cech 1984; Dyson and Rabbitts 1985; Lohr and Torchia 1988; Hamlin and Ma 1990; Brown *et al.* 1991; Berberich and Leffak 1993; Gerbi *et al.* 1993; Razin *et al.* 1994). Chromatin structure seems to play a key role in the regulation of DNA replication. The assembly of DNA into nucleosomes is inhibitory for the initiation of DNA replication (Cheng and Kelly 1989; Simpson 1990; Ishimi 1992; Li *et al.* 1993) and overcoming this inhibitory effect may require specific protein-DNA interactions at the replication origins. Furthermore, chromatin structure and chromosomal position influence the efficiency and timing of activation of replication origins (Forrester *et al.* 1990; Fangman and Brewer 1991). Therefore, the nucleosome organization of nuclear DNA can be viewed as an effective, generalized inhibitory structure for the initiation of DNA replication, while replication origins can be envisaged as chromosomal sites with an altered, "permissive" chromatin structure.

The two DHS sites at the lamin B2 *ori* are constitutive, since they are also observed in the chromatin of terminally differentiated, non-proliferating HL60 cells. Replication origins in yeast (Lohr and Torchia 1988; Brown *et al.* 1991) and *Tetrahymena* (Palen and Cech 1984) are also maintained in an "open" chromatin conformation in growing and resting cells. Our observations on differentiated HL60 cells are therefore consistent with the idea that the replication-favorable organization of the eukaryotic genome is preserved even in non-replicating cells.

The presence of such hypersensitive regions reflects the existence of a special chromatin architecture near the origins, created at least in part as a result of specific interactions between DNA and non-histone proteins (Elgin 1988; Gross and Garrard 1988).

The *in vivo* footprinting experiments demonstrate that, in exponentially growing HL60 cells, several specific DNA-protein interactions take place in an intergenic region of 500 bp overlapping the lamin B2 *ori*. Four of the observed footprints are clustered in a region between 130-380 nt. upstream of the start site of the *ppv1* transcript. Upon conversion to a non-proliferative state, a reduction in the intensity of these footprints was observed, that paralleled the diminished transcriptional activity of the region. Although this change is suggestive of some alteration in the DNA-protein contacts in the region, it is clear that the factors still bind *in vivo* to their recognition sites (figure 4.4.1). An explanation for this behavior is that they are probably involved in the transcriptional regulation of ppv1.
Another footprint maps in a region about 190 bp upstream of this group of protections (hOC site). It is the candidate site of interaction for an "origin complex". No footprints were detected within the intervening DNA sequence but the data from the DNase I-digestion of chromatin suggested that this DNA stretch is involved in interactions *in vivo* with proteins, possibly within a nucleosome particle.
Of particular interest, for the objective of the present study is the large and prominent footprint found in the *ori* region, nearest to the replication initiation site. In this context we took advantage of the property of HL60 myeloid cells to undergo terminal differentiation upon chemical stimulation (Barker *et al.* 1988; Meier *et al.* 1992), a process which is accompanied by a complete arrest of DNA replication. Thus, any specific aspect of the chromatin structure in proliferating cells can be compared to the replicationally "null" state of the differentiated cells. The results reported here indeed show evidence for specific, replication correlated, interactions at the lamin B2 origin (figure 4.4.1).

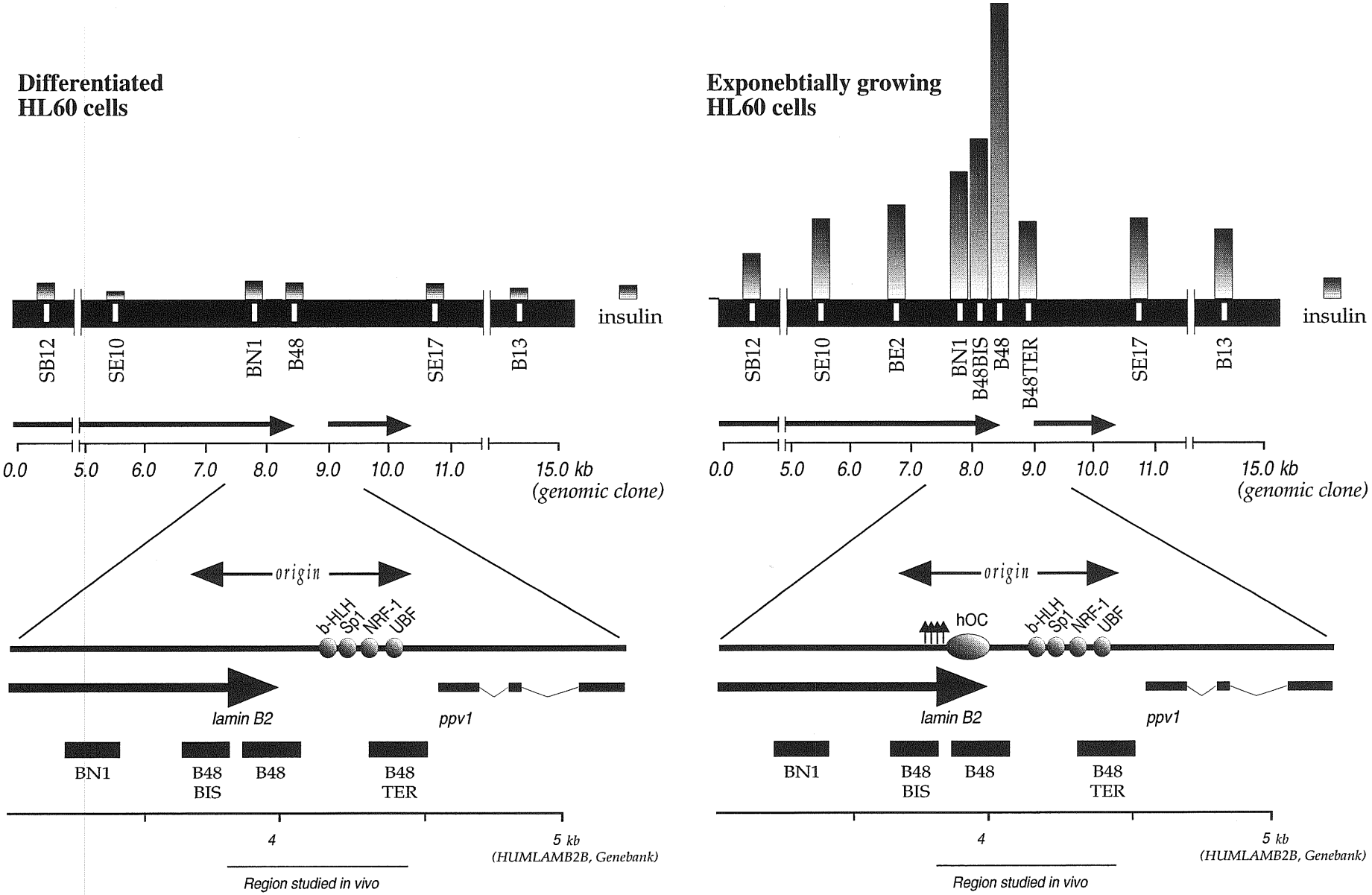


Figure 4.4.1 Comparison of the DNA-protein interactions detected *in vivo* in differentiated and exponentially growing HL60 cells.

The features of the footprint obtained in an asynchronous population of HL60 cells can be summarized as follows: 1) the protection covers 70 nt. on the lower DNA strand only; no protection, apart from few protected single nucleotides, was observed on the upper DNA strand, 2) the footprint is divided in two parts by an "island" of less-protected nucleotides, containing a reactive nucleotide in its centre, 3) periodic hypersensitive nucleotides, spaced at 11-12 bp intervals, were detected in the DNase I-cleavage pattern on the upper strand, a possible indication that the secondary structure of DNA in the region deviates from the B-type, 4) the binding of the protein to this site *in vivo* distorts the adjacent AT-rich region, 5) the protection is complete, which implies that the protein remains firmly bound during most of the cell cycle in all cells, 6) no sequence homology to other protein factor binding sites was found for this footprint, except for a topoisomerase II consensus element, 7) from a functional point of view, the most relevant feature is that the protection of the entire region disappears when HL60 cells cease replication.

Footprinting experiments performed on asynchronous cell cultures provides only a "collective" picture of the protein-DNA contacts in the whole population of cells. The recent results obtained with synchronized IMR90 cells suggest a more complex picture. Two major types of interactions were detected in different phases of the cell cycle. One of these interactions roughly corresponds to the 70 bp protection detected in unsynchronized cells and last for all the S phase (G2 phase was not studied). As in the case of the asynchronous cells, the protected region on the lower strand (non-coding) is divided in two subregions by an unprotected site but contrarily to the data of the asynchronous cells, the upper strand also show a small protected region roughly corresponding on the lower strand to the unprotected site. The second DNA-protein interaction was detected in a region of about 200 bp on both strands. The protection incorporates the 70 bp region and appears at the beginning of the G1 phase (in both cycling cells and cells coming from G0).

Even if these findings do not demonstrate a role of hOC in the control of DNA replication, many of the listed features are consistent with the properties of known (or candidate) *initiator* proteins. The complete

protection observed *in vivo* over the entire 70 bp region during all the cell cycle indicates that a protein complex is bound throughout the whole cell cycle. A similar feature was previously reported for the yeast ORC (Diffley and Cocker 1992; Diffley *et al.* 1994) and for the EBNA-1 protein which activates the latent replication origin, *ori*P, of the Epstein-Barr virus chromosome (Hsieh *et al.* 1993). The large size of the protected area suggests that more than one group of proteins may be responsible for the generation of the footprint. Accordingly, large nucleoprotein complexes have been implicated in the initiation of DNA replication at all replication origins characterized so far. Moreover, the induction of structural deformations of adjacent AT-rich regions of DNA is a hallmark of the interactions of the initiation proteins with DNA at the origins. Likewise, the prevalent protection of one of the strands in S phase is a feature found, *in vitro* or *in vivo*, also in other eukaryotic replication origins, even though its significance and possible functional role remain far from clear (Traut and Fanning 1988; Hofmann and Gasser 1991; Kuno *et al.* 1991; Bergemann and Johnson 1992; Diffley and Cocker 1992; Galli *et al.* 1992; Carmichael *et al.* 1993; Takai *et al.* 1994).

A strong indication for hOC role in replication comes from the genomic footprinting experiments. The situation described for the lamin B2 origin of replications resemble the situation described for the yeast origin of replication with a large pre-replicative complex lasting for all G1 and a smaller post-replicative complex that persists from G1/S through all S (figure 4.4.2).

More relevant data is the observation that genomic footprintings performed in IMR90 cells blocked in G0 by serum starvation didn't show any protection at the hOC site, as previously observed in differentiated (not replicating) HL60 cells. This means that the activity responsible for these interaction is present only in cycling cells ad is able to assemble at the origins only when the replication of DNA is needed.

Should these results be confirmed it will be possible to extrapolate a mechanism which is common for all eukaryotic systems. Origins of replication are "marked" throughout the cell cycle by binding to specific protein factors while other interactions are required to activate the origins during the S phase, as was recently shown for the yeast *ori*s .

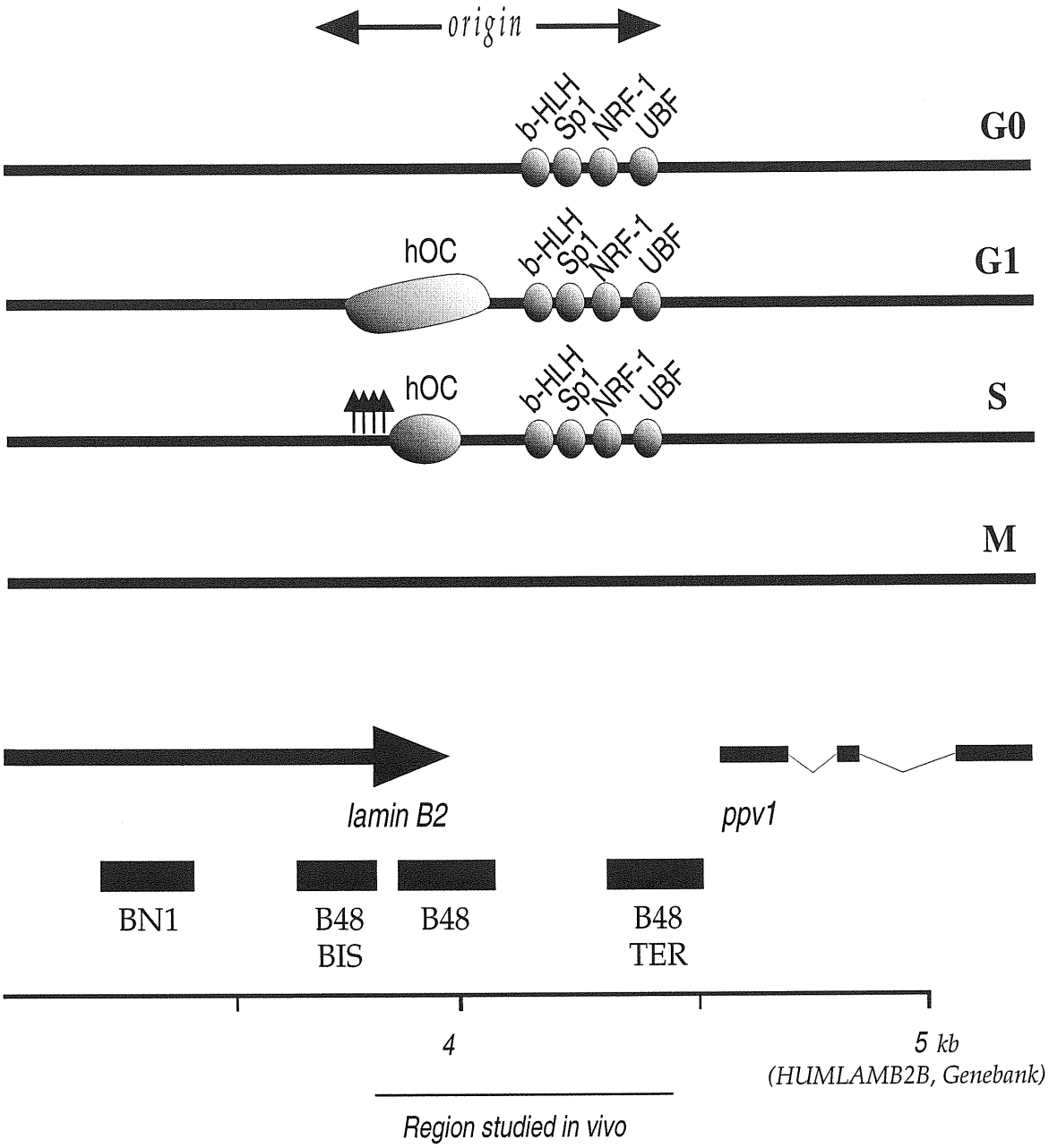


Figure 4.4.2 Variations in the DNA-protein interactions, at the lamin B2 origin of replication, during the various phases of the cell cycle in IMR90 cells.

4.5 FROM CHROMATIN STRUCTURE TO PROTEINS

Indeed, the application of the LMPCR *in vivo* footprinting technique for the high-resolution screening of the protein-DNA contacts at the lamin B2 *ori* revealed that several non-histone proteins bind to DNA in this region. Though the genomic footprinting analysis unambiguously identified these sites it does not present direct evidence regarding the identity of these proteins.

Binding of protein factors probably involved in transcriptional modulation

Based on the sequence homologies with known protein factor recognition sites, four of the footprints detected *in vivo* were designated b-HLH, Sp1, NRF-1 and UBF.

All the interactions detected *in vivo* were reproduced *in vitro* by DNase I footprinting using protein extracts obtained from the same cells used for the genomic footprinting analysis. For some of these interactions (b-HLH/USF and NRF1 sites) it was also demonstrated that nuclear extracts contain these proteins and that these proteins are able to bind their cognate sequences. On the contrary, for Sp1 and UBF, which are constitutive factors and for this reason they are also present in the extracts, the *in vitro* binding was not further investigated.

Due to their positioning in our region, it appears likely that these protein-DNA interactions are involved in the promoter activation of the ppv1 gene. The function of the ppv1 transcript is unknown but previous observations lead us to suppose that this may be a housekeeping gene with a TATA-less promoter. The transcript is present in all the human cell lines tested at a level correlated with proliferation. The spacer region and the 5'-portion of the ppv1 gene are contained within a CpG island which is persistently undermethylated in proliferating and quiescent cells. Furthermore they contain a binding site for Sp1, a cellular transcription factor that binds to the GC-rich promoters of a great number of housekeeping genes.

On the other hand, a role for these interactions in chromatin structure and hence in replication origin activation cannot be ruled out completely. USF, which normally acts as a transcriptional activator, and

in some cases has been reported having an anti-proliferative activity (Luo and Sawadogo 1996), is not the only factor able to bind the b-HLH binding site. Other b-HLH could also be responsible for the *in vivo* binding at this site. One of these proteins could be the Myc-Max complex, which is believed to be involved in DNA replication regulation. Myc is also the main amplified oncogene of the HL60 cells used in most of this study and *in vivo* crossliking experiments will be required to verify the real binding of any candidate protein.
Moreover, for Sp1 and UBF (even if their *in vitro* binding has not been analyzed yet) the ability to relieve the nucleosome-mediated repression on transcription and to function as transcription activators and antirepressors have been reported (Kuhn and Grummt 1992; Kamakaka and Kadonaga 1993). These activities could have a role in maintaining the chromatin in an open configuration that allows the assembly of the replication complex. Interestingly, six Sp1 boxes flank the late side of the SV40 core replication origin and the Sp1 protein has been shown to bind to these sequences *in vitro* (Dynan and Tjian 1983). In addition, *in vivo* footprinting experiments demonstrated that binding of a cellular protein to the SV40 Sp1 sites follows a specific temporal programme and that these protein-DNA interactions influence both the efficiency and timing of the replication of the viral genome (Buchanan and Gralla, 1990).

Binding of protein factors probably involved in origin activation and/or selection

Computer search of known protein recognition sequences did not reveal any significant homology to the 70 bp footprint identified in unsynchronized HL60 cells and in IMR90 cells in S phase. The same analysis performed on the ~200 bp footprint identified in IMR90 cells, synchronized in G1, gave the same result.
The only homology detected, concerns a small portion of the protection detected *in vivo* (3922-3930 nucleotides) that represent a putative topoisomerase II recognition site whose sequence is extremely degenerate.
Considering the large size and bipartite structure of the protected sequence, it is likely that more than one protein binds there. DNase I *in*

vitro footprinting experiments identified an activity that strongly resembles the activity responsible for the interaction detected *in vivo*. The activity named "Act-1" was identified in nuclear extracts from several cell lines in which the origin of the lamin B2 is known to be active (U937, K562, HeLa and IMR90 cells).

The protection obtained *in vitro* extends over all the 70 bp detected in unsynchronized cells leaving a small island of unprotected nucleotides in the same region where the unprotected site is present *in vivo*. Moreover, the protection also extends outside the 70 bp region in both directions forming a pattern intermediate between the G1 and S-phase patterns detected by genomic DMS-footprinting. On the contrary, the protection pattern of the complementary strand shows only small protections in crude extracts while the protection became stronger and wider after some steps of purification. The strength of this activity in crude protein extract is able to resist very high amounts of unspecific competitor and seems to be influenced by the presence of ATP in the binding reaction. Furthermore, Act-1 was identified only in extracts from actively replicating cells and not from non-proliferating cells (differentiated HL60 cells) as expected for an activity involved in DNA replication initiation.

All these findings strongly suggest that Act-1 is the protein factor/complex responsible for the interaction detected *in vivo* . This activity is probably the result of the binding of more than one protein as the fractionation of the protein extract containing Act-1 and -2 led to the identification of 8 different polypeptides.

In this view, the conservation of the same activity in different cell lines, with some small differences in the protection pattern, could be explained with the binding of the same protein complex. This complex should be constitutively present in replicating cells and the small differences could be explained with the binding of some other cell specific factors. Alternatively, the differences in the protection pattern detected in the extracts from different cell lines could simply reflect a different stoichiometry of the various components of the complex in different cell lines.

Two of these 8 proteins are the components of the Ku antigen (a complex that binds the DNA without sequence specificity) and their

identification is probably an artifact. Other components were probably lost during the purification procedure, as HOX C13 that was probably lost after the BioRex 70 column (figure 3.14.1) or the protein responsible for half of the footprinting produced by Act-2 that was lost after ammoniun sulfate precipitation (not shown). Nevertheless there are at least six different polypeptides that could be involved in the protection pattern of Act-1.

An interesting result was the isolation of some HOX proteins by one-hybrid screening. The *in vitro* binding was studied for two of these proteins and we can conclude that the binding is highly sequence-specific for at least one of these proteins. This observation is rather interesting as HOX proteins are mostly known for their function in metazoan development and as being transcriptional activators (Maconochie et al. 1996; Lawrence and Morata 1994; Kenyon 1994; Krumlauf 1994). Recently, a possible role of these proteins in cell proliferation was suggested by several studies (Care *et al.* 1996; Sauvageau *et al.* 1995; Condie and Capecchi 1994; Petrini *et al.* 1992) but up until now, there are no reports on a possible role of these proteins on DNA replication. Nevertheless, some HOX proteins were also identified in a similar one-hybrid screening performed with a fragment of the hamster DHFR origin of replication (C. Houchens and N. Heintz; personal communication). If this binding is confirmed *in vivo* new functions for these proteins involving replication-origin selection could be immagined. In this context, it will be extremely important to identify the proteins that interact with the HOX protein we identified.

4.6 CONCLUSIONS

The results reported in this thesis strongly support the idea that defined replication origins do exist in the genomes of higher eukaryotes. Besides the DNA sequence, other aspects of the nuclear structure and chromatin organization could be crucial for the establishment of a precise initiation process.
With these considerations it emerges that the eukaryotic replication origins might be regulated at several levels. The first level of

regulation could be at the chromatin structure level. An "open" chromatin structure, that allows the binding of the *initiator* factors, could explain the coincidence of replication origins with actively transcribed regions reported in several works. Nuclear structure and genomic regions outside the replication origins could have an important role in influencing this level. A second level of regulation could be the control of the local chromatin structure by the binding of sequence specific factors to facilitate the *initiator* binding. The transcription factors identified in this study could act in this step. The third level could be the binding of the *initiator* factors to mark the origin of replication throughout the cell cycle. At this level, all potentially active origins are expected to compete for initiator binding when the concentration of the latter is limiting. Finally, the activation of the origins should proceed through additional interactions with cell-cycle specific factors. Only at the end of this process, stable initiation complexes and active replicative forks would be assembled at the origins of replication.

In order to understand the details of these specific interactions, further studies together with the isolation of the proteins involved in this process will be required.

ACKNOWLEDGMENTS

The work presented in the thesis has been carried out mostly in the Molecular and Cellular Biology laboratory at the International Centre for Genetic Engineering and Biotechnology (I.C.G.E.B.), Trieste, Italy. Part of the work was also done in collaboration with the Istituto di Genetica Biochimica ed Evoluzionistica (I.G.B.E.-C.N.R.), Pavia, Italy.
During my Ph.D. programme I participated in most of the projects, carried out in the laboratory of Prof. Falaschi, which were aimed at the identification and characterization of the DNA replication origins in higher eukaryotes. To obtain a better organization of the results in my thesis, I also presented data obtained by other researchers in of the two laboratories previously mentioned. The contributors to the various sections of the thesis are indicated in the table reported below.

Development of a competitive PCR assay to quantitate DNA in traces	**P. Norio**
Mapping of the SV40 origin of replication in various plasmids	**P. Norio** S. Diviacco L. Zentilin
Mapping of the lamin B2 origin in synchronized and unsynchronized HL60 cells	S. Diviacco **P. Norio** G. Contreas
Mapping of the lamin B2 origin in several cell lines	S. Kumar S. Diviacco **P. Norio**
In vivo protein-DNA interactions at the lamin B2 origin of replication	D. Dimitrova G. Abdurashidova
In vitro protein-DNA interactions at the lamin B2 origin of replication	**P. Norio**
One-hybrid screening for the hOC site	E. De Stanchina D. Gabellini C. Ghigna
Production of recombinant HOX proteins and production of anti-HOX C13 polyclonal antibodies	E. De Stanchina P. Roncaglia
Characterization of the *in vitro* binding of the HOX C13	**P. Norio**
Identification and characterization of some hOC-site DNA-binding activities	**P. Norio**

The various sections of the project were supervised by:

Prof.	A. Falaschi (I.C.G.E.B.)
Dr.	G. Biamonti (I.G.B.E.)
Dr.	M. Giacca (I.C.G.E.B.)
Dr.	L. Zentilin (I.C.G.E.B.)
Prof.	S. Riva (I.G.B.E.)

I am grateful to Prof. A. Falaschi for having offered me the opportunity to work for a Ph.D. and to the Associazione Italiana per la Ricerca sul Cancro (A.I.R.C.) for the financial support.
My sincere thanks go to Prof. Gulnara Abdurashidova, Dr. Daniela Dimitrova, Dr. Alex Ochem for having provided me with their expertise; to Ms. Barbara Boziglav, Mrs. Maria Elena Lopez, Mr. Mauro Sturnega and Mr. Giancarlo Lunazzi for technical assistance. A special thanks is due to Ms. Ann Crum for her help in compiling the thesis.
Finally, I want to thank all my friends for their patience and moral support during the last years.

REFERENCES

Aladjem MI, Groudine M, Brody LL, Dieken ES, Fournier RE, Wahl GM, Epner EM: *Participation of the human beta-globin locus control region in initiation of DNA replication* Science 270: 815-819 (1995)

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

Anson DS, Choo KH, Rees DJ, Giannelli F, Gould K, Huddleston JA, Brownlee GG: *The gene structure of human anti-haemophilic factor IX* EMBO J. 3: 1053-1060 (1984)

Ariizumi K, Wang Z, Tucker PW: *Immunoglobin heavy chain enhancer is located near or in an initiation zone of chromosomal DNA replication* Proc. Natl. Acad. Sci. USA 90: 3695-3699 (1993)

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA Struhl K: *Current Protocols in Molecular Biology* John Wiley & Sons (1994)

Barker KA, Orkin SH, Newburger PE: *Expression of the X-CGD gene during induced differentiation of myeloid leukimia cell line HL-60* Mol. Cell. Biol. 8: 2804-2810 (1988)

Bell SP, Marc LR, Jamtzen H-M, Tjian R: *Functional cooperativity between transcription factors UBF1 and SL1 mediates human ribosomal RNA synthesis.* Science 241: 1192-1197 (1988)

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

Bell SP, Kobayashi R, Stillman B: *Yeast origin recognition complex functions in transcription silencing and DNA replication* Science 262: 1844-1849 (1993a)

Bell SP, Marahrens Y, Rao H, Stillman B: *The replicon model and eukaryotic chromosomes* Cold Spring Harbor Symp. Quant. Biol. 58: 435-442 (1993b)

Bell SP Mitchell J, Leber J, Kobayashi R, Stillman B: *The multidomain structure of Orc1p reveals similarity to regulators of DNA replication and transcriptional silencing* Cell 83: 563-568 (1995)

Benard M, Lagnel C, Pallotta D, Pierron G: *Mapping of a replication origin within the promoter region of two unlinked, abundantly transcribed actin genes of Physarium polycephalum* Mol. Cell. Biol. 16: 968-976 (1996)

Berberich S and Leffak M: *DNase-sensitive chromatin structure near a chromosomal origin of bidirectional replication of the avian a-globin locus* DNA Cell Biol. 12: 703-714 (1993)

Berberich S, Trivedi A, Daniel DC, Johnson EM, Leffak M: *In vitro replication of plasmids containing human c-myc DNA* J. Mol. Biol. 245: 92-109 (1995)

Bergemann AD and Johnson EM: *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 (1992)

Ma ZW, Bergemann AD, Johnson EM: *Conservation in human and mouse Pur alpha of a motif common to several proteins involved in initiation of DNA replication* Gene 149: 311-314

(1994)

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

Biamonti G, Giacca M, Perini G, Contreas G, Zentilin L, Weighardt F, Guerra M, Della Valle G, Saccone S, Riva S, Falaschi A. *The gene for a 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 (1992a)

Biamonti G, Perini G, Weighardt F, Riva S, Giacca M, Norio P, Zentilin L, Diviacco S, Dimitrova D, Falaschi A: *A human DNA replication origin: localization and transcriptional characterization* Chromosoma 102: S24-S31 (1992b)

Blumenthal AB, Kriegstein HJ, Hogness DS: *The units of DNA replication in Drosophila melanogaster chromosomes* Cold Spring Harbor Symp. Quant. Biol. 38: 205-223 (1973)

Bozzoni I, Baldari CT, Amaldi F, Buongiorno-Nardelli M: *Replication of ribosomal DNA in Xenopus laevis* Eur. J. Biochem. 118: 585-590 (1981)

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

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

Brewer BJ and Fangman WL: *Mapping replication origins in yeast chromosomes* Bioessays 13: 317-322 (1991)

Brewer BJ, Diller JD, Friedman KL, Kolor KM, Raghuraman MK Fangman WL: *The topography of chromosome replication in yeast* Cold Spring Harbor Symp. Quant. Biol. 58: 425-434 (1993)

Briggs MR, Kadonaga JT, Bell SP, Tjian R: *Purification and biochemical characterization of the promoter-specific transcription factor, Sp1* Science 234: 47-52 (1986)

Brown EH, Iqbal MA, Stuart S, Hatton KS, Valinsky J, Schildkraut CL: *Rate of replication of the murine immunoglobulin heavy-chain locus; evidence that the region is part of a single replicon* Mol. Cell. Biol. 7: 450-457 (1987)

Brown JA, Holmes SG, Smith MM: *The chromatin structure of Saccharomyces cerevisiae autonomously replicating sequences changes during the cell division cycle* Mol. Cell. Biol. 11: 5301-5311 (1991)

Buchman AR, Lue NF, Kornberg RD: *Connections between transcriptional activators, silencers, and telomeres as revealed by functional analysis of a yeast DNA-binding protein* Mol.Cell.Biol. 8: 5086-5099 (1988)

Buchanan RL and Gralla JD: *Programmed factor binding to Simian Virus 40 GC-box replication and transcription control sequences* J. Virol. 64: 347-353 (1990)

Burhans WC, Selegue JE, Heintz NH: *Isolation of the origin of replication associated with the amplified Chinese haamster dihydrofolate reductase domain* Proc. Natl. Acad. Sci. USA 83:

7790-7794 (1986)

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

Burhans WC, Vassilev LT, Wu J, Sogo JM, Nallaseth FS, DePamphilis ML: *Emetine allows identification of origins of mammalian DNA replication by imbalanced DNA synthesis, not through conservative nucleosome seggregation* EMBO J. 10: 4351-4360 (1991)

Caddle MS and Calos MP: *Analysis of autonomous replication behaviour in human cells of the dihydrofolate reductase putative chromosomal origin of replication.* Nucleic Acids Res. 20: 5971-5978 (1992)

Cairns J: *Autoradiography of HeLa cell DNA* J. Mol. Biol. 15: 372-373 (1966)

Callan HG: *Replication of DNA in eukaryotic chromosomes* Br. Md. Bull. 29: 192-195 (1973)

Calza RE, Eckardt LA, DelGiudice T, Schildkraut CL: *Changes in gene position are accompanied by a change in time of replication* Cell 36: 689-696 (1984)

Care A, Silvani A, Meccia E, Mattia G, Stoppacciaro A, Parmiani G, Peschle C, Colombo MP: *HOXB7 constitutively activates basic fibroblast growth factor in melanomas* Mol. Cell. Biol. 16: 4842-51 (1996)

Carmichael EP, Roome JM, Wahl AF: *Binding of a sequence-specific single-stranded DNA-binding factor to the Simian Virus 40 core origin inverted repeat domain is cell cycle regulated* Mol. Cell. Biol. 13: 408-420 (1993)

Carpenter PB, Mueller PR, Dunphy WG: *Role for a Xenopus Orc2-related protein in controlling DNA replication* Nature 379: 357-360 (1996)

Carroll SM, DeRose ML, Kolman JL, Nonet GH, Kelly RE, Wahl GM: *Localization of a bidirectional DNA replication origin in the native locus and in episomally amplified murine adenosine deaminase loci* Mol. Cell. Biol. 13: 2971-2981 (1993)

Cereghini S and Yaniv M: *Assembly of transfected DNA into chromatin: structural changes in the origin-promoter-enhancer region upon replication* EMBO J. 3: 1243-1253 (1984)

Challberg MD and Kelly TJ Jr: *Adenovirus DNA replication in vitro* Proc. Natl. Acad. Sci. USA 76: 655-659 (1979)

Chan CS and Tye BK: *Autonomously replicating sequences in Saccharomyces cerevisiae* Proc. Natl. Acad. Sci. USA 77: 6329-6333 (1980)

Cheng L and Kelly TJ: *Transcriptional activator nuclear factor I stimulates the replication of SV40 minichromosomes in vivo and in vitro* Cell 59: 541-551 (1989)

Cocker JH, Piatti S, Santocanale C, Nasmyth K, Diffley JF: *An essential role for the Cdc6 protein in forming the pre-replicative complexes of budding yeast* Nature 379: 180-182 (1996)

Condie BG and Capecchi MR: *Mice with targeted disruptions in the paralogous genes hoxa-3 and hoxd-3 reveal synergistic interactions* Nature 370: 304-307 (1994)

Contreas G, Giacca M, Falaschi A: *Purification of BrdUrd-substituted DNA by immunoaffinity chromatography with anti-BrdUrd antibodies* Biotechniques 12, 824 (1992)

Csordas-Toth E, Marusic L, Ochem A, Patthy A, Pongor S, Giacca M, Falaschi A: *Interactions of USF and Ku antigen with a human DNA region containing a replication origin* Nucl. Acids Res. 21: 3257-3263 (1993)

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

DePamphilis ML: *Eukaryotic DNA replication: anatomy of an origin* Annu. Rev. Biochem. 62: 29-63 (1993)

Deshpande AM and Newlon CS: *The ARS consensus sequence is required for chromosomal origin function in Saccharomyces cerevisiae* Mol. Cell. Biol. 12: 4305-4313 (1992)

Dhar SK, Roy Choudhury N, Mittal V, Bhattacharya A, Bhattacharya S: *Replication initiates at multiple dispersed sites in the ribosomal DNA plasmid of the protozoan parasite Entamoeba histolytica.* Mol. Cell. Biol. 16: 2314-2324 (1996)

Dhar V, Mager D, Iqbal A, Schildkraut CL: *The coordinate replication of the human b-globin gene domain reflects its transcriptional activity and nuclease hypersensitivity* Mol. Cell. Biol. 8: 4958-4965 (1988)

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

Diffley JF and Cocker JH: *Protein-DNA interactions at a yeast replication origin* Nature 357: 169-172 (1992)

Diffley JF, Cocker JH, Dowell SJ, Rowley A: *Two steps in the assembly of complexes at yeast replication origins in vivo* Cell 78: 303-316 (1994)

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

Dijkwel PA, Vaughn JP, Hamlin JL: *Mapping replication initiation sites in mammalian genomes by two-dimensional gel analysis: stabilization and enrcichment of the replication intermediates by isolation on the nuclear matrix* Mol. Cell. Biol. 11: 3850-3859 (1991)

Dijkwel PA and Hamlin JL: *Initiation of DNA replication in the dihydrofolate reductase locus is confined to the early S-period in CHO cells synchronized with plant amino acid mimosine* Mol. Cell. Biol. 12: 3715-3722 (1992)

Diller JD and Raghuraman MK: *Eukaryotic replication origins: control in space and time* Trends Biochem. Sci. 19: 320-325 (1994)

Dimitrova D, Giacca M, Demarchi F, Biamonti G, Riva S, Falaschi A: *In vivo protein-DNA interactions at a human DNA replication origin* Proc. Natl. Acad. Sci. USA 93: 1498-1503 (1996)

Diviacco S, Norio P, Zentilin L, Menzo S, Clementi M, Biamonti G, Riva S, Falaschi A, Giacca M: *A novel procedure for quantitative polymerase chain reaction by coamplification of competitive templates* Gene 122: 313-320 (1992)

Donovan S and Diffley JF: *Replication origins in eukaryotes* Curr. Opinions in Genet. Dev. 6: 203-207 (1996)

Dowell SJ, Romanowski P, Diffley JF: *Interaction of Dbf4, the Cdc7 protein kinase regulatory subunit, with yeast replication origins in vivo* Science 265: 1243-1246 (1994)

Dynan WS and Tjian R: *The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter* Cell 35: 79-87 (1983)

Dyson PJ and Rabbitts TH: *Chromatin structure around the c-myc gene in Burkitt lymphomas with upstream and downstream translocation points* Proc. Natl. Acad. Sci. USA 82: 1984-1988 (1985)

Dutta A, Ruppert JM, Aster JC, Winchester E: *Inhibition of DNA replication factor RPA by p53* Nature 365: 79-82 (1993)

Ehrenhofer-Murray AE, Gossen M, Pak DT, Botchan MR, Rine J: *Separation of origin recognition complex functions by cross-species complementation* Science 270: 1671-1674 (1995)

Elgin SC: *The formation and function of DNase I hypersensitive sites in the process of gene activation* J. Biol. Chem. 263: 19259-19262 (1988)

Evans MJ and Scarpulla RC: *NRF-1: a trans-activator of nuclear encoded respiratory genes in animal cells* Genes & development 4: 1023-1034 (1990)

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

Falaschi A, Giacca M, Zentilin L, Norio P, Dimitrova D, Kumar S, Tuteja R, Biamonti G, Perini G, Weighardt F, Riva S: *Searching for replication origins in mammalian DNA* Gene 135: 125-135 (1993)

Falaschi A and Giacca M: *A quest for a human ori* Genetica 94: 255-266 (1994)

Fangman WL and Brewer BJ: *Activation of replication origins within yeast chromosomes* Annu. Rev. Cell Biol. 7: 375-402 (1991)

Fangman WL and Brewer BJ: *A question of time: replication origins of eukaryotic chromosomes* Cell 71: 363-366 (1992)

Forrester WC, Epner E, Driscoll MC, Enver T, Brice M, Papayannopoulou T, Groudine M: *A deletion of the human b-globin locus activation region causes a major alteration in chromatin structure and replication across the entire b-globin locus* Genes Dev. 4: 1637-1649 (1990)

Foss M, McNally FJ, Laurenson P, Rine J: *Origin recognition complex (ORC) in*

transcriptional silencing and DNA replication in S. cerevisiae Science 262: 1838-44

Fox CA, Loo S, Dillin A, Rine J: *The origin recognition complex has essential functions in transcriptional silencing and chromosomal replication* Genes Dev. 9: 911-924 (1995)

Frappier L and Zannis HM: *Autonomous replication of plasmids bearing monkey DNA origin-enriched sequences* Proc. Natl. Acad. Sci. USA 84: 6668-6672 (1987)

Gale JM, Tobey RA, D'Anna JA: *Localization and DNA sequence of a replication origin in the rhodopsin gene locus of Chinese hamster cells* J. Mol. Biol. 224: 343-358 (1992)

Galli I, Iguchi-Ariga S, Ariga H: *The AT-rich tract of the SV40 ori core: negative synergism and specific recognition by single stranded and duplex DNA binding proteins* Nucl. Acids Res. 20: 3333-3339 (1992)

Gavin KA, Hidaka M, Stillman B: *Conserved initiator proteins in eukaryotes* Science 270: 1667-1671 (1995)

Gencheva M, Anachkova B, Russev G: *Mapping the sites of initiation of DNA replication in rat and human rRNA genes* J. Biol. Chem. 271: 2608-2614 (1996)

Gerbi SA, Liang C, Wu N, DiBartolomeis SM, Bienz TB, Smith HS, Urnov FD: *DNA amplification in DNA puff II/9A of Sciara coprophila* Cold Spring Harbor Symp. Quant. Biol. 58: 487-494 (1993)

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

Giacca M, Zentilin L, Norio P, Diviacco S, Dimitrova D, Contreas G, Biamonti G, Perini G, Weighardt F, Riva S, Falaschi A: *Fine mapping of a replication origin of human DNA.* Proc. Natl. Acad. Sci. USA 91: 7119-7123 (1994)

Gilbert D and Cohen S N: *Autonomous replication in mouse cells: a correction* Cell 56: 143-144 (1989)

Gossen M, Pak DT, Hansen SK, Acharya JK, Botchan MR: *A Drosophila homolog of the yeast origin recognition complex* Science 270: 1674-1677 (1995)

Gross DS and Garrard WT: *Nuclease hypersensitive sites in chromatin* Annu. Rev. Biochem. 57: 159-197 (1988)

Guo ZS, Heine U, DePamphilis ML: *T-antigen binding to site I facilitates initiation of SV40 DNA replication but does not affect bidirectionality* Nucleic Acids Res. 19: 7081-7088 (1991)

Hamlin JL and Ma C: *The mammalian dihydrofolate reductase locus* Biochem. Biophys. Acta 1087: 107-125 (1990)

Hamlin JL and Dijkwel PA: *On the nature of replication origins of higher eukaryotes* Curr. Opin. Genet. Dev. 5: 153-161 (1995)

Hand R: *Eukaryotic DNA: organization of the genome for replication* Cell 15: 317-325 (1978)

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

Hardy CF: *Characterization of an essential Orc2p-associated factor that plays a role* Mol. Cell. Biol. 16: 1832-1841 (1996)

Harland RM and Laskey RA: *Regulated replication of DNA microinjected into eggs of Xenopus laevis* Cell 21: 761-771 (1980)

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

Hay RT and DePamphilis ML: *Initiation of SV40 DNA replication in vivo: location and structure of 5' ends of DNA synthesized in the ori region* Cell 28: 767-779 (1982)

Hay RT, Hendrickson EA, DePamphilis ML: *Sequence specificity for the initiation of RNA-primed simian virus 40 DNA synthesis in vivo* J. Mol. Biol. 175: 131-157 (1984)

Heck MM and Spradling AC: *Multiple replication origins are used during Drosophila chorion gene amplification* J. Cell Biol. 110: 903-914 (1990)

Heintz NH and Hamlin JL: *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 (1982)

Heintz NH and Stillman BW: *Nuclear DNA synthesis in vitro is mediated via stable replication forks assembled in a temporally specific fashion in vivo* Mol. Cell. Biol. 8: 1923-1931 (1988)

Heinzel SS, Krysan PJ, Tran CT, Calos MP: *Autonomous DNA replication in human cells is affected by the size and the source of the DNA* Mol. Cell. Biol. 11: 2263-2272 (1991)

Higuchi R, Krummel B, Saiki RK: *A general method of in vitro preparation and specific mutagenesis of DNA A general method of in vitro preparation and specific mutagenesis of DNA* Nucleic Acids Res. 16: 7351-7367 (1988)

Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR: *Site-directed mutagenesis by overlap extension using the polymerase chain reaction* Gene 77: 51-59 (1989)

Hofmann JF-X and Gasser SM: *Identification and purification of a protein that binds the yeast ARS consensus sequence* Cell 64: 951-960 (1991)

Hsieh DJ, Camiolo SM, Yates JL: *Constitutive binding of EBNA1 protein to the Epstein-Barr virus replication origin, oriP, with distortion of DNA structure during latent infection* EMBO J. 12: 4933-4944 (1993)

Huberman JA: *New views in the biochemistry of the eukaryotic DNA replication revealed by aphidicolin, an unusual inhibitor of DNA polymerase a.* Cell 23: 647-648 (1981)

Huberman JA and Riggs AD: *On the mechanism of DNA replication in mammalian chromosomes* J. Mol. Biol. 32: 327-337 (1968)

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

Hyrien O, Maric C, Mechali M: *Transition in specification of embryonic metazoan DNA replication origins* Science 270: 994-997 (1995)

Iguchi-Ariga SMM, Ogawa N, Ariga H: *Identification of the initiation region of DNA replication in the murine immunoglobulin heavy chain gene and possible function of the octamer motif as a putative DNA replication origin in mammalian cells* Biochem Biophys. Acta 90: 3695-3699 (1993)

Ishimi Y: *Preincubation of T antigen with DNA overcomes repression of SV40 DNA replication by nucleosome assembly* J. Biol. Chem. 267: 10910-10913 (1993)

Jacob F, Brenner S, Cuzin F: *On the regulation of DNA replication in bacteria* Cold Spring Harbor Symp. Quant. Biol. 28: 329-33 (1963)

James CD and Leffak M: *Polarity of DNA replication through the avian alpha-globin genes.* Mol. Cell. Biol. 6: 976-9 (1986)

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

Kalejta RF and Hamlin JL: *The dual effect of mimosine on DNA replication* Exp. cell res. 231: 173-183 (1997)

Kamakaka RT and Kadonaga JT: *Biochemical analysis of the role of chromatin structure in the regulation of transcription by RNA polymerase II* Cold Spring Harbor Symp. Quant. Biol. 58: 205-212 (1993)

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

Kelly RE, DeRose ML, DraperBW, Wahl GM: *Identification of an origin of bidirectional DNA replication in the ubiquitously expressed mammalian CAD gene* Mol. Cell. Biol.15: 4136-4148 (1995)

Kenyon C: *If birds can fly, why can't we? Homeotic genes and evolution* Cell 78: 175-180 (1994)

Kitsberg D, Selig S, Keshet I, Cedar H: *Replication structure of the human b-globin gene domain* Nature 366: 588-590 (1993)

Kornberg A and Baker T: *DNA Replication* - Second edition, Freeman, W. H. and Company, New York (1992)

Krumlauf R: *Hox genes in vertebrate development* Cell 78: 191-201 (1994)

Kuhn A and Grummt I: *Dual role of the nucleolar transcription factor UBF: trans-activator and antirepressor* Proc. Natl. Acad. Sci. USA 89: 7340-7344 (1992)

Kumar S, Giacca M, Norio P, Biamonti G, Riva S, Falaschi A: *Utilization of the same DNA replication origin by human cells of different derivation* Nucleic Acids Research 24: 3289-3294 (1996)

Kuno K, Kuno S, Matsushima K, Murakami S: *Evidence for binding of at least two factors, including T-rich strand-binding factor(s) to the single-stranded ARS1 sequence in Saccharomyces cerevisiae* Mol. Gen. Genet. 230: 45-48 (1991)

Lalande M: *A reversible arrest point in the late G1 phase of the mammalian cell cycle* Exp. Cell Res. 186: 332-339 (1990)

Lawrence PA and Morata G: *Homeobox genes: their function in Drosophila segmentation and pattern formation* Cell 78: 181-189 (1994)

Leatherwood J, Lopez-Girona A, Russell P: *Interaction of Cdc2 and Cdc18 with a fission yeast ORC2-like protein* Nature 379: 360-363 (1996)

Leffak M and James CD: *Opposite replication polarity of the germ line c-myc gene in HeLa cells compared with that of two Burkitt lymphoma cell lines* Mol. Cell. Biol. 9: 586-593 (1989)

Leu TH and Hamlin JL: *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 (1989)

Levine AJ, Kang HS, Billheimer FE: *DNA replication in SV40-infected cells. Analysis of replicating SV40 DNA* J. Mol. Biol. 50: 549-568 (1970)

Li R, Yang L, Fouts E, Botchan MR: *Site-specific DNA-binding proteins important for replication and transcription have multiple activities* Cold Spring Harbor Symp. Quant. Biol. 58: 403-413 (1993)

Li JJ and Herskowitz I: *Isolation of ORC6, a component of the yeast origin recognition complex by a one-hybrid system* Science 262: 1870-1874 (1993)

Li R and Botchan MR: *The acidic transcriptional activation domains of VP16 and p53 bind the cellular replication protein A and stimulate in vitro BPV-1 DNA replication* Cell 73: 1207-21 (1993)

Liang C, Spitzer JD, Smith HS, Gerbi SA: *Replication initiates at a confined region during DNA amplification in Sciara DNA puff II/9A* Genes Dev. 7: 1072-1084 (1993)

Liang C and Gerbi SA: *Analysis of an origin of DNA amplification by a novel three-dimensional gel method.* Mol. Cell. Biol. 14: 1520-1529 (1994)

Liang C, Weinreich M, Stillman B: *ORC and Cdc6p interact and determine the frequency of initiation of DNA replication in the genome* Cell 81: 667-676 (1995)

Linskens MHK and Huberman JA: *Organization of replication of ribosomal DNA in Saccharomyces cerevisiae* Mol. Cell. Biol. 8: 4927-4935 (1988)

Linskens MHK and Huberman JA: *Ambiguities in results obtained with 2D gel replicon mapping techniques* Nucleic Acids Res. 18: 647-652(1990)

Little RD, Platt TH, Schildkraut CL: *Initiation and termination of DNA replication in human rRNA genes* Mol. Cell. Biol. 13: 6600-6613 (1993)

Loo S, Fox CA, Rine J, Kobayashi R, Stillman B, Bell S: *The origin recognition complex in silencing, cell cycle progression, and DNA replication* Mol. Biol. Cell. 6: 741-756 (1995)

Lohr D and Torchia T: *Structure of the chromosomal copy of yeast ARS1* Biochemistry 27: 3961-3965 (1988)

Luo X and Sawadogo M: *Antiproliferative properties of the USF family of helix-loop-helix transcription factors* Proc. Natl. Acad. Sci. USA 93: 1308-1313 (1996)

Ma C, Leu TH, Hamlin JL: *Multiple origins of replication in the dihydrofolate reductase amplicons of a methotrexate-resistant chinese hamster cell line* Mol. Cell. Biol. 10: 1338-1346 (1990)

Maconochiene M, Nonchev S, Morrison A, Krumlauf R: Paralogous HOX genes: function and regulation Annu. Rev. Genet. 30: 529-556 (1996)

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

Mariani BD and Schimke RT: *Gene amplification in a single cell cycle in chinese hamster ovary cells* J. Biol. Chem. 259: 1901-1910 (1984)

Mastrangelo IA, Held PG, Dailey L, Wall JS, Hough PV, Heintz N, Heintz NH: *RIP60 dimers and multiples of dimers assemble link structures at an origin of bidirectional replication in the dihydrofolate reductase amplicon of Chinese hamster ovary cells* J. Mol. Biol. 232: 766-778 (1993)

Marx J: *How DNA replication originates* Science 270; 1585-1587 (1995)

Maxam AM and Gilbert W: *Sequencing end-labeled DNA with base-specific chemical cleavages* In L. Grossman and K. Moldave (ed.), Methods in Enzymology, Academic Press, New York. 65: 499-560 (1980)

McKnight SL, Bustin M, Miller OJ: *Electron microscopic analysis of chromosome metabolism in the Drosophila melanogaster embryo* Cold Spring Harbor Symp. Quant. Biol. 42: 741-754 (1977)

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

Meier RW, Chen T, Mathews S, Niklaus G, Tobler A: *The differentiation pathway of HL60 cells is a model system for studying the specific regulation of some myeloid genes* Cell Growth & Differentiation 3: 663-669 (1992)

Micklem G, Rowley A, Harwood J, Nasmyth K, Diffley JF: *Yeast origin recognition complex is involved in DNA replication and transcriptional silencing* Nature 366: 87-89 (1993)

Miller SA, Dykes DD, Polesky HF: *A simple salting out procedure for extracting DNA from human nucleated cells* Nucl. Acids Res. 16: 1215-1215 (1988)

Mills AD, Blow JJ, White JG, Amos WB, Wilcock D, Laskey RA: *Replication occurs at discrete foci spaced throughout nuclei replicating in vitro* J. Cell. Sci. 94: 471-477 (1989)

Mosca PJ, Dijkwel PA, Hamlin JL: *The plant amino acid mimosine may inhibit initiation at origins of replication in chinese hamster cells* Mol. Cell. Biol. 12: 4375-4383 (1992)

Muzi-Falconi M and Kelly TJ: *Orp1, a member of the Cdc18/Cdc6 family of S-phase regulators, is homologous to a component of the origin recognition complex* Proc. Natl. Acad. Sci. USA 92: 12475-12479 (1995)

Nawotka KA and Huberman JA: *Two-dimensional gel electrophoretic method for mapping DNA replicons* Mol. Cell. Biol. 8: 1408-1413 (1988)

Newlon CS and Theis JF: *The structure and function of yeast autonomously replicating sequences* Curr. Opinions in Genet. Dev. 3: 752-758 (1993)

Palen TE and Cech TR: *Chromatin structure at the replication origins and transcription-initiation regions of the ribosomal RNA genes of Tetrahymena* Cell 36: 933-942 (1984)

Paterno GD, Adra CN, McBurney MW: *X chromosome reactivation in mouse embryonal carcinoma cells* Mol. Cell. Biol. 5: 2705-2712 (1985)

Pelizon C, Diviacco S, Falaschi A, Giacca M: *High resolution mapping of the origin of DNA replication in the hamster dihydrofolate reductase gene domain by competitive PCR* Mol. Cell. Biol. 16: 5358-5364 (1996)

Petrini M, Quaranta MT, Testa U, Samoggia P, Tritarelli E, Care A, Cianetti L, Valtieri M, Barletta C, Peschle C: *Expression of selected human HOX-2 genes in B/T acute lymphoid leukemia and interleukin-2/interleukin-1 beta-stimulated natural killer lymphocytes* Blood 80: 185-193 (1992)

Rao H, Marahrens Y, Stillman B: *Functional conservation of multiple elements in yeast chromosomal replicators* Mol. Cell. Biol. 14; 7643-51 (1994)

Rao H and Stillman B: *The origin recognition complex interacts with a bipartite DNA binding site within yeast replicators* Proc. Natl. Acad. Sci. U S A 92: 2224-2228 (1995)

Razin SV, Kekelidze MG, Lukhanidin EM, Scherrer K, Georgiev GP: *Replication origins are attached to the nuclear skeleton.* Nucleic Acids Res. 14: 8189-8207 (1986)

Razin SV, De Moura Gallo C, Scherrer K: *Characterization of the chromatin structure in the upstream region of the chicken a-globin gene domain* Mol. Gen. Genet. 242: 649-652 (1994)

Romanowski P, Madine MA, Rowles A, Blow JJ, Laskey RA: *The Xenopus origin recognition complex is essential for DNA replication and MCM binding to chromatin* Curr. Biol. 6: 1416-1425 (1996)

Rowles A, Chong JP, Brown L, Howell M, Evan GI, Blow JJ: *Interaction between the origin recognition complex and the replication licensing system in Xenopus* Cell 87: 287-296 (1996)

Rowley A, Cocker LH, Harwood J, Diffley JFX: *Initiation complex assembly at budding yeast replication origins begins with the recognition of a bipartite sequence by limiting amounts of the initiator ORC* EMBO J. 14: 2631-2641 (1995)

Russev G and Vassilev L: *Isolation of a DNA fraction from Ehrlich ascites tumour cells containing the putative origin of replication* J. Mol. Biol. 161: 77-87 (1982)

Saluz HP and Jost JP: *A Laboratory Guide to Genomic Sequencing* Birkhauser Verlag Basel, Basel(1987)

Sambrook J, Fritsch EF, Maniatis T: *Molecular Cloning. A laboratory Manual* II edition, Cold Spring Harbour Laboratory, Cold Spring Harbour, NY. (1989)

Santocanale C and Diffley JF: *ORC- and Cdc6-dependent complexes at active and inactive chromosomal replication origins in Saccharomyces cerevisiae* EMBO J. 15: 6671-6679 (1996)

Sauvageau G, Thorsteinsdottir U, Eaves CJ, Lawrence HJ, Largman C, Lansdorp PM, Humphries RK: *Overexpression of HOXB4 in hematopoietic cells causes the selective expansion of more primitive populations in vitro and in vivo* Genes. Dev. 9 : 1753-1765 (1995)

Selig S, Okumura K, Ward D C, Cedar H: *Delineation of DNA replication time zones by flourescence in situ hybridization* EMBO J. 11: 1217-1225 (1992)

Shinomiya T and Ina S: *DNA replication of histone gene repeats in Drosophila melanogaster tissue culture cells: multiple initiation sites and replication pause sites* Mol. Cell. Biol. 13: 4098-4106 (1993)

Shinomiya T and Ina S: *Mapping an initiation region of DNA replication at a single-copy chromosomal locus in Drosophila melanogaster cells by two-dimensional gel methods and PCR-mediated nascent strand analysis: multiple replication origins in a broad zone* Mol. Cell. Biol. 13: 4098-4106 (1994)

Siebenlist U, Hennighausen L, Battey J, Leder P: *Chromatin structure and protein binding in the putative regulatory region of the c-myc gene in Burkitt lymphoma* Cell 37: 381-391 (1984)

Simpson RT: *Nucleosome positioning can affect the function of a cis-acting DNA element in vivo* Nature 343: 387-389 (1990)

Spack EG, Lewis ED, Paradowski B, Schimke RT, Jones PP: *Temporal order of DNA replication in the H-2 major histocompatibility complex of the mouse* Mol. Cell. Biol. 12: 5174-5188 (1992)

Stinchcomb DT, Struhl K, Davis RW: *Isolation and characterization of a yeast chromosomal replicator* Nature 282: 39-43 (1979)

Stinchomb DT, Thomas M, Kelly J, Selker E, Davis RW: *Isolation and characterization of a yeast chromosomal replicator* Proc. Natl. Acad. Sci. USA 77: 4559-4563 (1980)

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

Sudo K, Ogata M, Sato Y, Iguchi AS, Ariga H: *Cloned origin of DNA replication in c-myc gene can functionand be transmitted in transgenic mice in an episomal state* Nucleic Acids Res. 18: 5425-5432 (1990)

Sweder KS, Rhode PR, Campbell JL: *Purification and characterization of proteins that bind to yeast ARSs* J. Biol. Chem. 263: 17270-17277 (1988)

Taira T, Sanae H, Iguchi-Ariga SMM, Ariga H: *A novel DNA replication origin identified in the human heat shock protein gene promoter* Mol. Cell. Biol. 14: 6386-6397 (1994)

Takai T, Nishita Y, Iguchi-Ariga SMM, Ariga H: *Molecular cloning of MSSP-2, a c-myc gene single-strand binding protein: characterization of binding specificity and DNA replication activity* Nucl. Acids Res. 22: 5576-5581 (1994)

Tasheva ES and Roufa DJ: *A mammalian origin of bidirectional DNA replication within the Chinese hamster RPS14 locus* Mol. Cell. Biol. 14: 5628-5635 (1994)

Traut W and Fanning E: *Sequence-specific interactions between a cellular DNA-binding protein and the simian virus 40 origin of DNA replication* Mol. Cell Biol. 8: 903-911(1988)

Trempe JP, Lindstrom YI, Leffak M: *Opposite replication polarities of transcribed and non-transcribed histone H5 genes* Mol. Cell. Biol. 8: 1657-1663 (1988)

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

Truett M A and Gall JG: *The replication of ribosomal DNA in the macronucleus of Tetrahymena* Chromosoma 64: 295-303 (1977)

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

Vassilev LT, Burhans WC, DePamphilis ML: *Mapping an origin of DNA replication at a single-copy locus in exponentially proliferating mammalian cells* Mol. Cell. Biol. 10: 4685-4689 (1990)

Vassilev LT and DePamphilis ML: *Guide to identification of origins of DNA replication in eukaryotic cell chromosomes* Crit. Rev. Biochem. Molec. Biol. 27: 445-472(1992)

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

Venditti P, Costanzo G, Negri R, Camilloni G: *ABFI contributes to the chromatin organization of Saccharomyces cerevisiae ARS1 B-domain* Biochim. Biophys. Acta 1219: 677-689 (1994)

Virbasius CA, Virbasius JV, Scarpulla RC: *NRF-1, an activator involved in nuclear-mitochondrial interactions, utilizes a new DNA-binding domain conserved in a family of developmental regulators* Genes Dev. 7: 2431-2445 (1993)

Virta-Pearlman VJ, Gunaratne PH, Chinault AC: *Analysis of a replication initiation sequence from the adenosine deaminase region of the mouse genome* Mol. Cell. Biol. 13: 5931-5942 (1993)

Vogt VM and Braun R: *The replication of ribosomal DNA in Physarum polycephalum* Eur. J.

Biochem. 80: 557-566 (1977)

Walter J and Newport JW: *Regulation of replicon size in Xenopus egg extracts* Science 275: 993-995 (1997)

Waltz SE, Trivedi AA, Leffak M: *DNA replication initiates non-randomly at multiple sites near the c-myc gene in HeLa cells* Nucleic Acids Res. 24: 1887-1894 (1996)

Wang TF: *Eukaryotic DNA polymerases* Annu. Rev. Biochem. 60: 513-552(1991)

Yoon Y, Sanchez A, Brun C, Huberman JA: *Mapping of replication initiation sites in human ribosomal DNA by nascent-strand abundance analysis* Mol. Cell. Biol. 15: 2482-2489 (1995)

Zannis-Hadjopoulos M, Persico M, Martin RG: *The remarkable instability of replication loops provides a general method for the isolation of origins of DNA replication* Cell 27: 155-163 (1981)

Zhao Y, Tsutsumi R, Yamaki M, Nagatsuka Y, Ejiri S-I, Tsutsumi K-I: *Initiation zone of DNA replication at the aldolase B locus encompasses transcription promoter region* Nucleic Acids Res. 22: 5385-5390 (1994)

Zou L, Mitchell J, Stillman B: *CDC45, a novel yeast gene that functions with the origin recognition complex and Mcm proteins in initiation of DNA replication* Mol. Cell. Biol. 17: 553-563 (1997)