ISAS - INTERNATIONAL SCHOOL
FOR ADVANCED STUDIES

Analysis of nascent DNA as a tool for the identification of mammalian origins of DNA replication

Thesis submitted for the Degree of

Doctor Philosophiae

Candidate: Cristina Pelizon

Supervisors: Dr. Mauro Giacca

Prof. Arturo Falaschi

Academic Year 1996/1997

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

TRIESTE

Index

SYNOPSIS

In eukaryotic cells, DNA in each chromosome is synthesized by multiple tandemly organized replication units (replicons), which are activated at precisely defined times of the S phase. In each replicon, initiation of DNA replication occurs at a specific site (origin of DNA replication) from which two oppositely moving semiconservative forks originate. The concept that DNA synthesis in metazoan chromosomes initiates at preferred sites is supported by several different studies. However, at present, there is no reliable functional assay for origin activity. Because of this experimental limitation, a variety of physico-chemical or biochemical approaches have been developed for mapping the localization of DNA replication origins directly in the mammalian genome. Most of these studies, however, entail the utilization of potentially artifactual techniques (such as utilization of protein inhibitors, cell synchronization, or cell permeabilization) or produce results which are often difficult to interpret (such as 2D gel analysis).

The work presented in this thesis has been an effort to understand initiation of DNA replication in mammalian cells. For this purpose, I contributed to the development of a novel procedure for mapping origins of DNA replication, based on the quantification of nascent DNA strands by competitive PCR. The research has been initially focused on Chinese hamster cells with the prospect to precisely identify the origin region in the DHFR locus. With this regard, the hamster DHFR origin region is particularly challenging since virtuallly every available mapping approach has been utilized to investigate it. However, despite these intense investigations, contradictions on the nature of DNA replication origins in this region are still unexplained. By using the origin mapping protocol we developed, a high-resolution mapping study on the single copy DHFR gene locus in hamster CHO K1 cells has been performed. The results obtained reinforce the notion that DNA replication starts at a precise site, located ~17 kb downstream of the DHFR gene. In the nascent DNA samples, the fragment containing the start site is enriched over 10-fold with respect to those located at the two boundaries of the analyzed region, or within the 3' end of the DHFR gene itself. The origin region has been trimmed down to a ~800 bp segment.
Further, our experience in nascent DNA analysis, increased by an intense search for technical improvements and comprehension of the mapping technique, has encouraged its extensive application for the identification of other origins in mammalian chromosomes. However, since the systematic analysis of large regions of mammalian chromosomes by PCR quantification of nascent DNA is highly demanding and time consuming, we have attempted to develop a new approach for

the identification of DNA replication origins. For this purpose, I constructed a library of replication origin-enriched sequences from human cells to be used as a probe for hybridization to immobilized cloned fragments from a region of interest. The principle of the method is that, being the library enriched in origin sequences, it will preferentially hybridize to immobilized fragments containing an origin of replication. This approach, when applied to the human lamin B2 region where an origin of replication had already been mapped by quantification of nascent strand abundance by competitive PCR, has allowed to identify in a very simple and rapid way the origin. Library construction and characterization are presented in the second part of the thesis, together with its possible usage to identify new origins of DNA replication in human cells.
Finally, with the purpose to study factors binding to and activating origins, recombinant proteins and antibodies useful in this respect have been prepared. Since work on this topic is still ongoing in the laboratory, preparation of these reagents is presented as an appendix to the thesis.

Part of the results presented in this thesis have been published in the following papers:

Pelizon C., Diviacco, S., Falaschi A. and Giacca M. (1996). 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.

Giacca M., Pelizon C. and Falaschi A. (in press). Mapping replication origins by quantifying relative abundance of nascent DNA strands using competitive polymerase chain reaction. In: Identification and Analysis of Replication Origins in Eukaryotic Cells. Methods: A Companion to Methods in Enzymology, Academic Press.

INTRODUCTION

DNA replication is a process of dramatic interest for its relevance in the transmission of the genetic material from one cell to the next. It is not surprising, therefore, that eukaryotic DNA replication is a highly regulated process, controlled by a range of complex mechanisms.

We know that DNA synthesis in higher eukaryotes normally initiates at specific sites on each chromosome termed origins of DNA replication (oris). We also know that initiation of DNA replication requires the ordered assembly of many proteins at origins in such a way that they fire at precisely defined times of the S phase and the resulting duplication of the genome is restricted to once per cell cycle (Coverly and Laskey, 1994; Harland and Laskey, 1980; Nasmyth, 1996; Romanowski and Madine, 1997). Thus, replication origins ensure complete and efficient duplication of the DNA.

There is a general agreement that the control mechanisms of DNA replication are exerted on the initiation step. Therefore, a key starting point to understand how cells regulate replication is the origin of DNA replication. However, despite massive efforts in the last several years, replication origin localization and specification in mammalian cells are still poorly known.

The replicon model for DNA replication initiation

Most of the important concepts in DNA replication initiation arose from studies of bacteria which led to the formulation of the replicon model for DNA replication (Jacob *et al.*, 1963). According to this model, replication initiates when a positive trans-acting factor called the initiator recognizes and binds to a specific cis-acting element in the DNA called the replicator (Figure 1). This event is followed by DNA unwinding at or near the replicator in order to expose the two single-stranded, complementary DNA templates to the replication machinery. DNA synthesis mostly proceeds

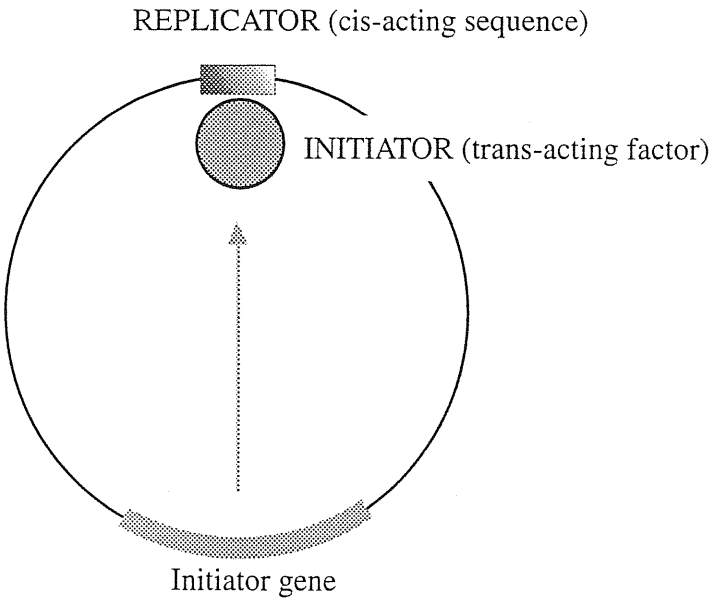


Figure 1. The Replicon model (Jacob and Brenner, 1963).

bidirectionally at the replication forks resulting in bubble structures (for reviews, see: DePamphilis, 1993; DePamphilis, 1996; Falaschi and Giacca, 1994). The unit of DNA replicated as a result of the initiator acting at a specific replicator is called "replicon". In this thesis the term "origin" of DNA replication (ori) will be used to indicate both the DNA sequences at which replication initiates and the cis-acting elements that promote replication.

Analyses at the molecular level of the chromosome of bacteria, bacteriophage, plasmids, animal virus and mitochondria had demonstrated this model of replication to be valid for all of these systems, although acting through a diversity of molecular mechanisms (for reviews, see: Bramhill and Kornberg, 1988; DePamphilis, 1996; Kornberg and Baker, 1992). In particular, the SV40 replication system has been dissected and reconstituted in vitro with pure molecules (Waga and Stillman, 1994).

In comparison with prokaryotes, plasmids and viruses which replicate DNA from a single replicon, the huge amount of DNA that must be replicated within each S phase in eukaryotic cells ($6x10^9$ bp in the case of human cells) requires several hundreds of replicons to ensure the complete and efficient replication of all the genome. Fiber autoradiography studies clearly demonstrated that each linear eukaryotic chromosome is composed of many tandemly organized replicons, irregularly spaced along chromosomal molecules (~50-100 kb intervals) and activated (each once and only once) at different times of the cell cycle (Huberman and Riggs, 1968). According to this picture, each eukaryotic replicon is conceptually similar to the prokaryotic one. Thus, at the level of the replicon, DNA synthesis in eukaryotic cells initiates at a specific site, proceeds bidirectionally at two replication forks, and terminates when forks of neighboring replicons meet.

Validation of the replicon model for DNA replication initiation has also been demonstrated in a simple eukaryote, the yeast *Saccharomyces cerevisiae*, where both functional approaches (Stinchcomb *et al.*, 1979) and in vivo mapping studies (Brewer and Fangman, 1987; Brewer *et al.*, 1988; Huberman *et al.*, 1988; Newlon,

1988) indicated that defined sequences are used as origins of replication in chromosomes (for a review: Newlon, 1996). Since these sequences also allow plasmids to replicate extrachromosomally in *S. cerevisiae*, they were called autonomously replicating sequences (ARSs). Very recently, an important progress towards the understanding of the initiation of DNA replication in eukaryotes has occurred thanks to the discovery of the yeast origin recognition complex (ORC). This six protein complex is thought to be the initiator whose sequence-specific interactions with the ARSs determine origin activation events (Bell *et al.*, 1993; Bell and Stillman, 1992; Diffley and Cocker, 1992).

These findings provided strong support to the notion that the replicon model may help to understand also the initiation of DNA replication in higher eukaryotic chromosomes. However, although potential replicator and initiator proteins are beginning to be characterized in a wide variety of animal cells, many crucial concepts for understanding mammalian chromosomal DNA replication are not clear yet.

Mammalian DNA replication is a process that results from the interplay of specific DNA sequences, chromosome structure and nuclear organization. Therefore, the replicon model should be fitted with a general model for chromatin organization, where sites of replication are immobilized on a structural framework of nuclear matrix, with large loops of DNA bulging out, that might conceivably correspond to replicons (Cook, 1991; Jackson, 1990). According to this model, DNA replication in animal cells results to be confined to a few hundreds of discrete foci, each of which consist of several hundreds of replicative forks (Laskey *et al.*, 1989). In this perspective, chromosome replication involves more than the replication of DNA, since the whole chromatin structure must be replicated, including the assembly of nucleosomes and the chromosome scaffold.

How replication initiates at the DNA level in mammalian cells has still to be understood. Although the replicon model has been generalized and extended also to higher eukaryotes, it has been difficult to demonstrate in most of them since the

actual DNA sequences that specify replication initiation are poorly characterized and the proteins interacting with these sequences remain mostly elusive.

Moreover, some of the results obtained appear controversial as far as the specificity of DNA replication initiation is concerned. Several observations point to the existence of specific replication origins, starting from the work of Amaldi (Amaldi *et al.*, 1973) which showed that the same cells initiate DNA replication in the same chromosomal regions in subsequent cell cycles. Conversely, other data suggest that higher eukaryotic chromosomes do not require specific DNA sequences to initiate replication. This lack of site-specific initiation appears in ribosomal RNA genes (rDNA) during the very early divisions in Xenopus embryos (Hyrien and Mechali, 1993), suggesting that patterns of oris utilization in higher eukaryotes may be imposed by chromosomal context and nuclear structure, rather than specific DNA sequences. This picture seems very likely, since later during development, when the chromatin structure is remodeled to allow gene transcription, initiation sites become more restricted (Hyrien *et al.*, 1995).

The explanation of this controversy on the nature of metazoan replication origins mostly appears to consist in their complexity, a complexity that involves not only specific DNA sequences but also chromosomal and nuclear structures in determining, all together, replication origins organization and usage.

Furthermore, the lack of a genetic assay for origin function has been a great limitation to the research in this field. Therefore, knowledge of replication origins in higher eukaryotes mostly derives from the use of approaches to physically identify oris localization within mammalian chromosomes. However, these studies had produced embarrassing conclusions opening a long debate on the nature of replication origins in animal cells.

Autonomous replication studies

As discussed above, long-standing evidences indicate that DNA replication initiates at specific sites within the chromosomes of higher eukaryotes. Evidences in favour of this conclusion have been obtained from many different approaches and by studying origin regions in several different species. If replication initiates at specific sites, why has it been so difficult to identify specific DNA sequences that function as replication origins in plasmid replication assays in animal cells?

The lack of a specific plasmid-based replication assay (Biamonti *et al.*, 1985) similar to the one that proved successful in studies of yeast cells (Stinchcomb *et al.*, 1979) is a crucial problem which has mainly hampered the identification of DNA replication origins in higher eukaryotes. Attempts to identify putative origin sequences based on autonomous replication of plasmids have been either unsuccessful or unreproducible (Burhans *et al.*, 1986; Burhans *et al.*, 1990; Gilbert and Cohen, 1989) or has proven to be nonspecific (Heinzel *et al.*, 1991; Krysan and Calos, 1991; Krysan *et al.*, 1989). Interestingly, any large (>10 kb) DNA fragment can confer some ARS activity in mammalian cells when inserted into a non-replicating EBV-derived vector (Heinzel *et al.*, 1991). Therefore, the problem here is not that too few sequences replicate, but that all sequences seem to replicate to some extent when reintroduced into mammalian cells. Apart from some criticisms regarding the replication assays used in these studies, the most obvious objection to these results is that plasmid replication may not necessarily occurs like chromosomal replication and may not have the same requirements such as association with nuclear matrix or epigenetic modification of DNA. However, the same conclusion is obtained when any DNA is added to extracts of Xenopus eggs (Blow and Laskey, 1986) or injected into unfertilized eggs of Xenopus (Harland and Laskey, 1980; Coverly and Laskey, 1994): DNA replication is initiated at a single randomly chosen site within any DNA molecule, in a strict once-per-cell-cycle manner (Harland and Laskey, 1980; Hyrien and Mechali, 1992; Mahbubani *et al.*, 1992; Mechali and Kearsey, 1984). It is important to point out that

these studies do not show that replication initiation on endogenous chromosomal DNA is independent of sequence. They show only that exogenous DNA does not require specific DNA sequences for DNA replication or for its coupling to the cell cycle.

In view of the inadequacy of replication assays to recognize mammalian ori sequences for their function, a number of mapping approaches have been developed to identify origins directly on the chromosomes.

Mapping origins of replication in mammalian cells

An origin of bi-directional DNA replication can be operationally defined as the site where: i) nascent DNA strands are initiated (i.e., where short and/or newly synthesized DNA stretches - different from Okazaki fragments - are localized); ii) a transition occurs between continuous DNA synthesis on the leading strand, and discontinuous synthesis on the lagging strand (Okazaki fragments); iii) replication bubbles are localized (Falaschi *et al.*, 1993). These three features have all been utilized in the last years for origin mapping in eukaryotic chromosomes, with different success in terms of resolution, sensitivity, and actual results obtained (for a review, see DePamphilis, 1996).

In general terms, the identification of origins of DNA replication in mammalian cell chromosomes by mapping the actual sites where replication begins is a very demanding task. The overall difficulty of the origin mapping approaches, despite their diversity, derives from the fact that they are expected to detect an extremely limited amount of origin-specific DNA dispersed in a huge amount of unspecific chromosomal DNA. To overcome this limitation, mapping studies have been performed mostly on highly amplified domains, and usually exploiting techniques for the enrichment of newly synthesized DNA such as cell synchronization, cell permeabilization, psoralen cross-linking, and treatment with protein synthesis

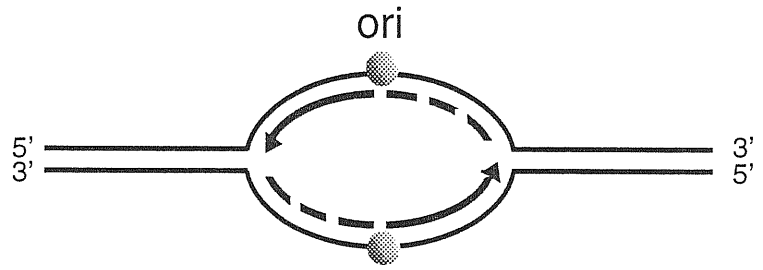
inhibitors (for a review on these methods, see Vassilev and DePamphilis, 1992). Thus, their outcome is potentially biased on one hand by the suspicion that amplified domains may have peculiar features in terms of replication initiation and on the other hand by the possibility that the synchronization protocol may alter the physiological mechanism of DNA replication. In addition, most of the developed protocols for origin mapping are very laborious and tricky to perform. A critical evaluation of these procedures is presented in references (Vassilev and DePamphilis, 1992) and (Falaschi *et al.*, 1993).

Essentially, the mapping methods fall into three categories according to the ori feature they exploit (Figure 2): i) analysis of the distribution of nascent DNA (Anachkova and Hamlin, 1989; Giacca *et al.*, 1994; Vassilev and Johnson, 1990; Vassilev *et al.*, 1990); ii) identification of the region where the transition from discontinuous to continuous DNA synthesis occurs (Burhans *et al.*, 1990; Burhans *et al.*, 1991; Handeli *et al.*, 1989; Kitsberg *et al.*, 1993); iii) analysis of the replication intermediates by two-dimensional (2-D) gel electrophoresis (Little *et al.*, 1993; Vaughn *et al.*, 1990).

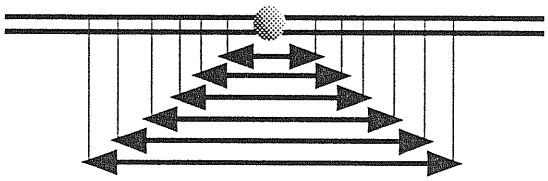
Mapping of nascent DNA strands

When DNA replication initiates, short, nascent DNA fragments are first synthesized from the ori region (see Figure 2, A). This feature has been used in replication timing studies (Brown *et al.*, 1987; Calza *et al.*, 1984; Dhar *et al.*, 1988; Furst *et al.*, 1981; Gale *et al.*, 1992; Selig *et al.*, 1992; Spack *et al.*, 1992) which provided information about the replication properties for wide regions of the genome.

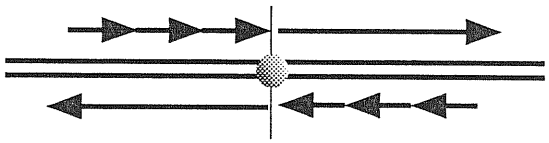
Further replication timing studies on amplified regions provided increased resolution of the ori regions. The most thoroughly investigated one is the dihydrofolate reductase (DHFR) gene domain in CHOC 400 cells, a hamster cell line containing more than 1000 tandem copies of the locus (for a review, see Hamlin and Ma, 1990). In highly synchronized cell populations, nascent DNA fragments synthesized immediately after entry in S phase have been labeled by the incorporation of a



A Mapping the localization of nascent DNA fragments



B Mapping the inversion of polarity of leading or lagging strands



C Mapping the localization of replication intermediates

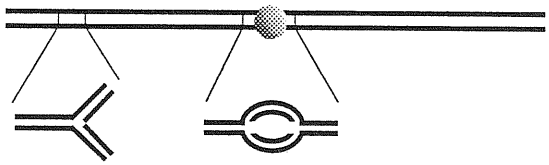


Figure 2. Methods for mapping DNA replication origins in higher eukaryotic cells.

radioactive nucleotide precursor and directly visualized after specific restriction enzyme digestion (Heintz and Hamlin, 1982), or used as probes against cloned DNA fragments of the region (Burhans *et al.*, 1986). These early experiments, and a further refinement obtained by the utilization of a in-gel renaturation procedure to eliminate most of the background due to non amplified genomic sequences (Leu and Hamlin, 1989) suggested the presence of two origins of replication located downstream of the DHFR gene, spaced about 20 kb apart. The limitation of these methods, besides the low resolution capacity, is the poor sensitivity, which allows their application only to highly amplified regions and to highly synchronized cell populations.

Enrichment of nascent DNA can also be obtained by trapping newly synthesized DNA between two neighboring psoralen cross-links flanking the origin, and subsequently extruding it by alkaline denaturation (Anachkova and Hamlin, 1989). While it is possible to apply this method also to non-synchronized cells, its major drawback is the potential alteration of the physiological controls of DNA replication as a consequence of psoralen treatment.

An alternative method for mapping the temporal order of replication of a defined chromosomal region and thus the localization of the ori, is the so-called in vitro run off technique, which consists of allowing isolated nuclei to elongate DNA chains initiated in vivo in the presence of heavy DNA precursors. Since new initiation events should not occur in isolated nuclei, the extent of label incorporation in each molecule should be directly proportional to the distance from the ori. Variations of this method were use to map, although at low resolution, the localization of ori in the avian α-globin (James and Leffak, 1986) and histone H5 (Trempe *et al.*, 1988) genes, upstream of the human c-myc gene (Leffak and James, 1989; McWhinney and Leffak, 1990), and in the amplified DHFR locus (Heintz and Stillman, 1988). An origin has also been identified in the 5' flanking region of the Chinese hamster rhodopsin gene (Gale *et al.*, 1992). However, hybridization signals from single copy mammalian loci are weak and resolution is low.

Although all these methods are not sensitive enough to allow a precise definition of ori regions, the results obtained argue for the presence of fixed initiation sites. The limiting step of all of these approaches is the purification of sufficient amounts of newly synthesized DNA from the bulk of unreplicated DNA.

To analyze samples containing very few molecules of newly synthesized DNA, PCR is the best choice. PCR has been indeed used to quantify the relative abundance of three different regions in the human c-myc and in the hamster DHFR loci (Vassilev and Johnson, 1990; Vassilev *et al.*, 1990) within samples of nascent DNA of different lengths, providing evidence for the localization of DNA replication origins. However, although being extremely sensitive, this approach is not able to provide quantitative results, since its outcome may be affected by many variables influencing conventional PCR (Ferre, 1992; Siebert and Larrick, 1992; see last section of Introduction).

Leading and lagging strand polarity studies

Assuming that DNA replication in mammalian chromosomes is initiated bidirectionally at a specific site with two forks moving in opposite directions, and being DNA synthesis continuous on one strand, but discontinuous on the other, leading- and lagging-strand synthesis must switch template strands at the center of the replication bubble where DNA synthesis begins. Therefore, the identification of template switch points leads to the localization of an origin of bidirectional replication (OBR, see Figure 2, B).

The polarity of labeled Okazaki fragments reveals the expected transition between discontinuous and continuous DNA synthesis, and thus can be used to identify an OBR. This method has led to the precise mapping of the ori-β of the DHFR locus with a resolution of 450 nt (Burhans *et al.*, 1990), and has also been applied to the study of several other regions. It allowed to identify an ori within the amplified adenosine deaminase (ADA) locus in murine cells (Carroll *et al.*, 1993), to map an ori in a ~2 kb region overlapping the ribosomal protein S14 gene in hamster cells (Tasheva and

Roufa, 1994) and an origin in the CAD transcription unit in hamster cells as well (Kelly *et al.*, 1995). However, in order to obtain sufficient sensitivity, cells must be synchronized, permeabilized by the addition of detergents and labeled with a very high amount of ^{32}P-labeled dNTPs. These procedures reduce considerably the rate of DNA synthesis and possibly might also perturb the initiation process in unpredictable ways.

The switch in the polarity of the continuously synthesized leading strand can also be used to localize an ori. This method has been successfully used to map, although at low resolution, ori-β in the hamster DHFR (Burhans *et al.*, 1991; Handeli *et al.*, 1989) and adenine phosphoribosyltransferase-encoding gene (Handeli *et al.*, 1989). The same technique has been successfully applied to the identification of a potential human replicator in the region upstream of the β-globin gene (Kitsberg *et al.*, 1993). However, this technique requires extensive inhibition of protein synthesis by emetine (also causing a preferential inhibition of lagging strand synthesis), which might alter the pattern or mechanism of DNA replication.

It is noteworthy that these two complementary approaches identifying the polarity of leading and lagging strands have provided consisting results when applied to the hamster DHFR locus.

Two-dimensional (2D) gel techniques

Each type of replication intermediate is characterized by a specific 2-D gel electrophoretic pattern (Figure 2, C). In neutral/neutral 2-D gels, DNA fragments containing a single replication fork, a replication bubble or two replication forks give characteristic patterns, referred as fork arcs (represent replication forks), bubble arcs (contain an origin of replication) and double Y (termination sites). Using a different version of the 2-D gel method (neutral/alkaline gels) it is possible to determine the direction of fork travel through a DNA locus.

This method is the technique of choice for mapping origins of DNA replication in low complex genomes and has been successfully employed in a number of studies in the

2 μm circle (Brewer and Fangman, 1987), on viral (Gahn and Schildkraut, 1989; Yang and Botchan, 1990) and yeast replicons (Brewer *et al.*, 1988; Huberman *et al.*, 1988; Linskens and Huberman, 1988). Two dimensional gels have also been used to study DNA replication initiation in metazoan cells. However, hybridization patterns in higher eukaryotes appear more complex resulting in problematic interpretations of the results (Linskens and Huberman, 1990a; Linskens and Huberman, 1990b). In fact, when applied to the DHFR locus, this method gave an unexpected result indicating that initiation of DNA replication occurs at many sites, perhaps even randomly, within a ~50 kb initiation area downstream of the DHFR gene, either in the amplified (Dijkwel and Hamlin, 1995; Dijkwel *et al.*, 1991; Dijkwel *et al.*, 1994; Vaughn *et al.*, 1990) or in the single-copy gene locus (Dijkwel and Hamlin, 1992). The same technique also showed that initiation is delocalized in the human rDNA locus being DNA replication able to initiate at any site within the spacer region between the multiple rDNA genes (Little *et al.*, 1993).

Mapping the location of several different replication origins has revealed a paradox for DNA replication initiation in animal cells (Table 1). According to the results obtained from methods based on the analyses of nascent DNA strands, DNA replication initiates at specific sites on the chromosomes ranging from origin localization in a 0.5 kb to some kb fragments. However, methods analyzing DNA structures by 2-D gels raised doubts about the actual existence of precise replication origins in higher eukaryotes. Detecting larger initiation zones (as long as 3 kb in *S. pombe* (Zhu *et al.*, 1992) and as long as 55 kb in Chinese hamster (Vaughn *et al.*, 1990)) where initiation events may occur, these approaches suggest that in eukaryotes more complex than *S. cerevisiae*, diffused broad areas of the chromosome are activated simultaneously as delocalized origins, in sharp contrast with what happens in prokaryotes (Linskens and Huberman, 1990a).
Since specific origins have been identified by different methods, in independent laboratories, it can be stated that site specific initiation is not an artifact of the

Region studied	Organism	Method	Main conclusion	Reference
Ribosomal protein S14 gene (RSP14)	Hamster	PCR of nascent DNA strands analysis of Okazaki fragments	origin mapped in ~2.5 kb region overlapping RSP14 gene	Tasheva and Roufa, 1994
Rhodopsin gene	Hamster	earliest replicated DNA	origin contained in a 10 kb region overlapping the gene	Gale et al., 1992
APRT locus	Hamster	imbalanced DNA synthesis	origin mapped in a ~6 kb region	Handeli et al., 1989
CAD gene	Hamster	analysis of Okazaki fragments	origin localized in 5 kb region in the CAD transcription unit	Kelly et al., 1995
Aldolase B locus	Rat	earliest replicated DNA	origin mapped in a 1 kb region containing the AldB promoter	Zhao et al., 1994
ADA gene (early S) (late S)	Mouse	analysis of Okazaki fragments PCR of nascent DNA strands	11 kb origin region at 28.5 kb 5' of ADA gene 1-2 kb origin region at 150 kb 3' of ADA gene	Carroll et al., 1993 Virta-Pearlman et al., 1993
Ig heavy chain gene	Mouse	semiquantitative PCR	0.6 kb origin mapped within the enhancer	Ariizumi et al., 1993
HSP70 gene promoter region	Human	PCR of nascent DNA strands	origin region mapped in a ~400 bp fragment	Taira et al., 1994
β-globin gene	Human	imbalanced DNA synthesis	origin mapped in a 2 kb fragment	Kitsberg et al., 1993
rRNA genes	Human	2-D gel electrophoresis analysis of nascent strand abundance	multiple initiation sites in the 31 kb NTS	Little et al., 1993 Yoon et al., 1995
	Human/Rat	analysis of nascent strand abundance	4 kb region	Gencheva et al., 1996
c-myc gene	Human	PCR of nascent DNA strands	origin mapped within a 2.5 kb region upstream of c-myc gene	Vassilev and Johnson, 1990
Lamin B2 locus	Human	quantitation of nascent DNA strands	origin mapped in a 474 bp region in the non transcribed spacer	Giacca et al., 1994
Polymerase α gene	Drosophila	2-D gel electrophoresis PCR of nascent DNA strands	multiple origins identified in a ~10 kb region downstream of pol a gene	Shinomiya and Ina, 1994
Histone gene	Drosophila	2-D gel electrophoresis PCR of nascent DNA strands	multiple initiation sites in the 5 kb histone gene transcription unit	Shinomiya and Ina, 1993
Chorion locus	Drosophila	2-D gel electrophoresis	origin contained in a 1 kb region	Heck and Spradling, 1990
Puff II/9A gene	Sciara coprophila	2-D gel electrophoresis	origin region containing a ~1kb major initiation area	Liang and Gerbi, 1993 Liang and Gerbi, 1994

Table 1. Origin mapping studies in animal cells.

experimental approach used to map them. However, the long-standing discrepancy between this conclusion and the one derived from 2-D gel analysis remains to be understood. Some hypotheses have been put forward in order to reconcile these opposite findings (DePamphilis, 1993; Hamlin, 1992; Linskens and Huberman, 1990a). Possibly, these conflicting data may reflect the complex initiation pattern of replication in mammalian chromosomes suggesting that quite diverse DNA sequence elements can specify origins, together with specific proteins and chromosomal context. In favour of the site specific initiation of higher eukaryotes DNA replication, it should be pointed out that initiation sites are genetically determined as results from maintenance of origin activity after translocation to other chromosomal sites (Handeli *et al.*, 1989; Orr-Weaver, 1991). All together, these data indicate that mammalian origins of DNA replication are most likely localized to discrete regions of the chromosomes, although the specific DNA sequences involved remain largely speculative (DePamphilis, 1996).

The Chinese hamster DHFR replication origin

During the past years, several laboratories have studied replication of the dihydrofolate reductase (DHFR) gene domain of Chinese hamster ovary (CHO) cells in an effort to identify the sequences required for origin specificity. Since the CHO DHFR has been investigated with virtually every available mapping approach, it provides a good perspective of the search for mammalian replication origins. However, conflicting results have been obtained for this region with respect to the localization of the replication origin (Figure 3).

Most of the interest for this region began with the establishment of the CHOC 400 cell line. This cell line is a derivative of the Chinese hamster ovary (CHO)K1 cell line in which the DHFR locus is amplified over 1,000-fold (Milbrandt *et al.*, 1981) greatly facilitating mapping studies. The ~240 kb amplified region in CHOC 400 cells contains

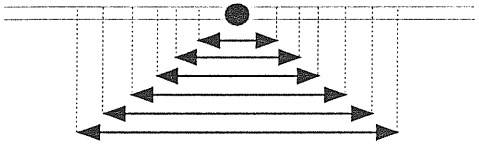
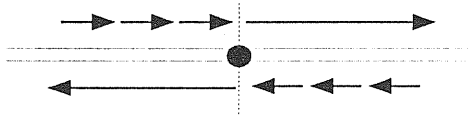
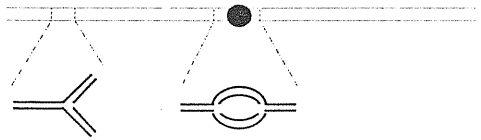
Approach	Method	Cell line and treatment	Main conclusion	Reference
Mapping the localization of nascent DNA fragments 	earliest labeled fragment	CHOC 400 G1/S	ori-β and ori-γ mapped with ~2 kb resolution	Leu and Hamlin, 1989
	trapping of nascent DNA strands by crosslinks	CHOC 400, G1/S psoralen crosslinking UV irradiation	ori-β and ori–γ mapped with ~1 kb resolution	Anachkova and Hamlin, 1989
	PCR of nascent DNA strands	CHOK1	ori-β mapped within 2 kb	Vassilev et al., 1990
Mapping the inversion of polarity of leading or lagging strands 	imbalanced DNA synthesis (leading strand)	CHOC 400, CHOK1 emetine	ori-β mapped within a 14 kb region ~17 kb downstream of the DHFR gene	Handeli, et al., 1989 Burhans, et al., 1991
	analysis of Okazaki fragments (lagging strand)	CHOC 400, CHOK1 G1/S cell permeabilization labeling	ori–β mapped within 450 bp	Burhans, et al., 1990
Analysis of replication intermediates 	2D-gel electrophoresis	CHOC 400, G1/S nuclear matrix	origin dispersed in a large initiation zone (≥50 kb)	Vaughn, et al., 1990 Dijkwel, et al., 1991
		CHOK1,G1/S nuclear matrix		Dijkwell et al., 1995

Figure 3. Methods for mapping the DHFR origin of DNA replication and main conclusions obtained.

the DHFR gene itself, the downstream gene 2BE2121 and the ~70 kb intergenic region between them in which initiation sites have been identified (see Figure 4). Initial labeling studies suggested that each DHFR amplicon contains at least one replication origin that is activated early in the S phase (Milbrandt *et al.*, 1981). Pulse-labeling studies on synchronized CHOC 400 cells further identified a series of restriction fragments that began replication early in the S phase (early-labeled restriction fragments, ELFs) indicating the presence of a DNA replication initiation region downstream of the DHFR gene (Heintz and Hamlin, 1982). Subsequent higher resolution methods have been used in an effort to more precisely map initiation events in the intergenic ELF region. Hybridization of DNA that had been pulse-labeled with radioactive thymidine during various intervals of the CHOC 400 S phase to immobilized plasmids allowed to map an origin of DNA replication (termed ori-β) to an approximately 5 kb fragment positioned ~ 17 kb downstream of the DHFR gene (Burhans *et al.*, 1986b); this area was later narrowed down to an approximately 2 kb DNA subfragment and a second initiation site, termed ori-γ, was identified 20 kb apart (Leu and Hamlin, 1989). Finally, results from studies utilizing a cross-linking strategy suggested that an initiation site may be contained within an approximately 500 bp subfragment of the previously defined earliest labeled fragment (Anachkova and Hamlin, 1989).

A significant step ahead has been the study of the DNA replication pattern of the single copy gene locus in the parental CHO K1 cells through the analysis of polarity of leading strand synthesis (Handeli *et al.*, 1989) and of Okazaki fragments (Burhans *et al.*, 1990). While the former study gave low resolution results grossly overlapping with those obtained with the amplified gene domain, the latter indicated that ori-β DNA replication initiated within a 450 bp segment of the earliest labeled fragment but differing from the one previously identified by cross-linking of nascent DNA. The same region was also identified by a semi-quantitative study using PCR amplification of nascent DNA to measure its length (Vassilev *et al.*, 1990).

Taken together, these observations suggest the existence of a precise initiation site at ori-β which may correspond to (or be close to) a true replicator. Importantly, this locus is incapable of supporting ARS activity (Caddle and Calos, 1992), but still behaves as a replicator when introduced into ectopic chromosomal sites (Handeli *et al.*, 1989). However, this picture is not supported by the analysis of the replication intermediates in this region by the two dimensional (2-D) gel electrophoresis approach. The neutral/neutral 2-D gel technique detected complete bubble arcs and complete single fork arcs in every fragment within the 55 kb intergenic region downstream of the DHFR gene (Dijkwel and Hamlin, 1992; Dijkwel *et al.*, 1991; Vaughn *et al.*, 1990). Moreover, the neutral/alkaline gel method detected replication forks moving in both direction at all locations within the same region (Dijkwel and Hamlin, 1995). The interpretation of these data is that initiation can occur at any site in a 55 kb initiation zone.

The DHFR origin region clearly exemplifies the discrepancy between the results obtained by the analysis of nascent DNA length or polarity (origin circumscribed within a small - 0.5 to 5 kb- genomic region) and those obtained by 2-D gel mapping technique (origin dispersed within a ~55 kb broad initiation zone). For a summary of these results, see Figure 4. However, it should be pointed out that most of the studies performed so far entail the utilization of potentially artifactual techniques (such as treatment with protein inhibitors, cell synchronization, cell permeabilization) or produce results which are often difficult to interpret (such as 2-D gel patterns of replication intermediates migration).

Resolution of this apparent paradox will require the understanding of the relationship between 2-D gel patterns and nascent DNA detected by the other studies.

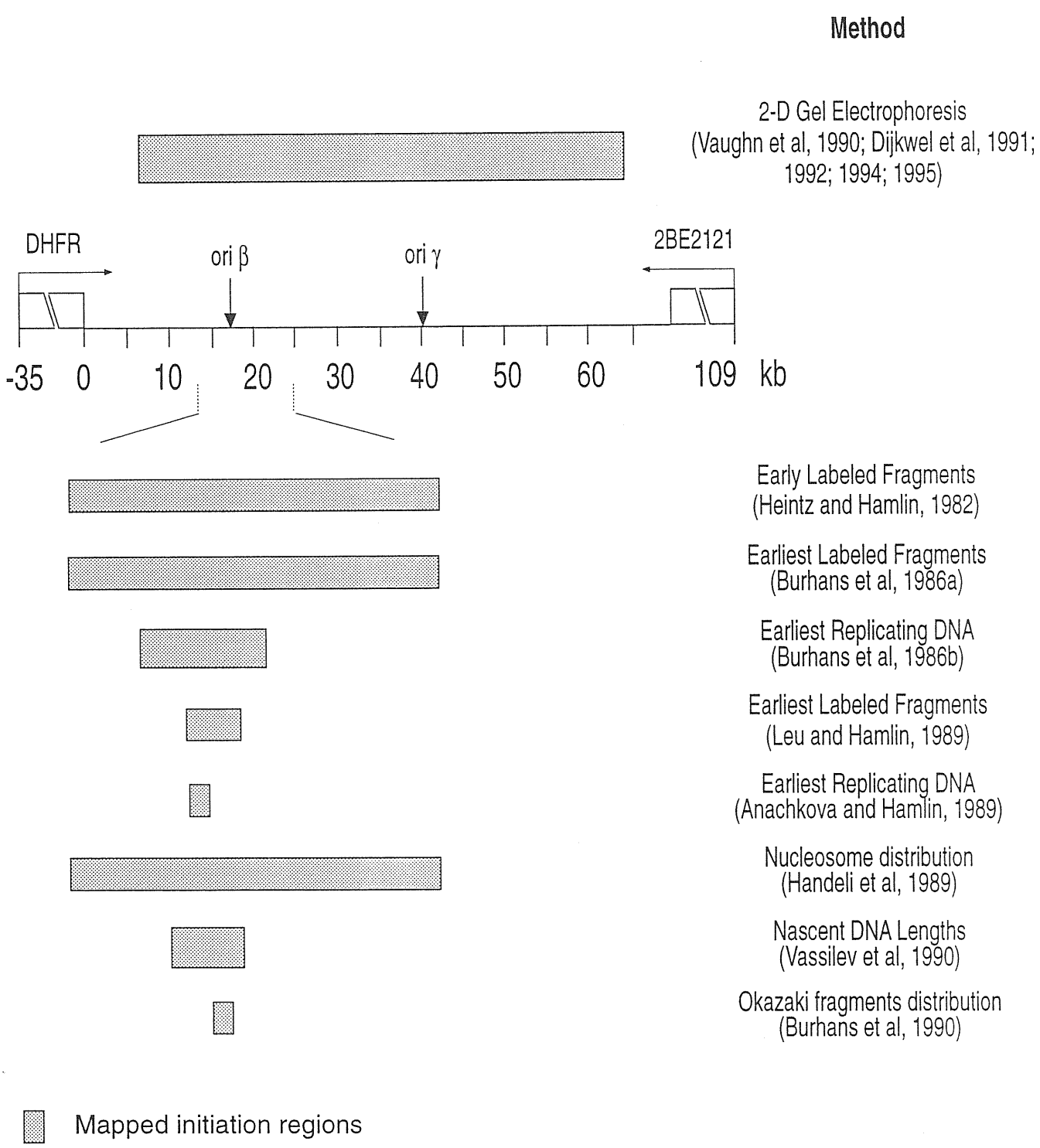


Figure 4. Initiation of DNA replication in the hamster DHFR locus of CHO K1 and CHOC 400 cells. The DHFR and 2BE2121 genes are indicated. The limits of the ori-β region are compared among initiation sites identified by a variety of methods under various experimental conditions. Methods used to map ori-β DNA replication initiation sites are listed and described in the corresponding references.

Nascent DNA strand abundance assay

As already mentioned, several methods have been developed for mapping DNA replication origins in mammalian chromosomes. However, the isolation of sufficient amounts of nascent DNA from the bulk of nonreplicated DNA is the limiting step of most of them. Although procedures for enrichment of newly synthesized DNA have been widely used, a very limited amount of newly synthesized DNA can be obtained, often very close to the detection limit of any hybridization procedure even if starting from a high amount of synchronized cells.

To overcome this limitation, we developed a protocol for origin mapping in mammalian cells based on PCR amplification of nascent DNA, that could be performed in physiological conditions using the simplest and least artifact-prone procedure (Giacca *et al.*, 1994; Kumar *et al.*, 1996). In particular, the developed procedure i) allows the analysis of single copy genomic regions; ii) can be performed on cultured and primary cells in the absence of any chemical treatment; iii) does not require cell synchronization; iv) allows origin mapping at high resolution [few hundred base pairs (bp)]. This permits the study of the cis- and trans-acting elements required for origin function.

The method is based on the notion that an origin of bidirectional DNA replication is the site where short stretches of newly replicated DNA start to be synthesized and are then elongated, progressively covering adjacent sequences on the DNA away from the initiation site. Therefore, nascent DNA issuing from a replication origin consists of a population of short, single stranded DNA molecules, centered over the origin itself. Consequently, the relative abundance of defined markers within a given genomic area in samples of short nascent DNA fragments is inversely correlated with their distance from the origin, giving a distribution of values that peak at the replication start site. As outlined in Figure 5 and detailed in its legend, this protocol entails the isolation of short [~1000 nucleotide (nt)], stretches of newly synthesized 5'-bromodeoxyuridine (BrdU)-labeled DNA and the measurement, within this sample,

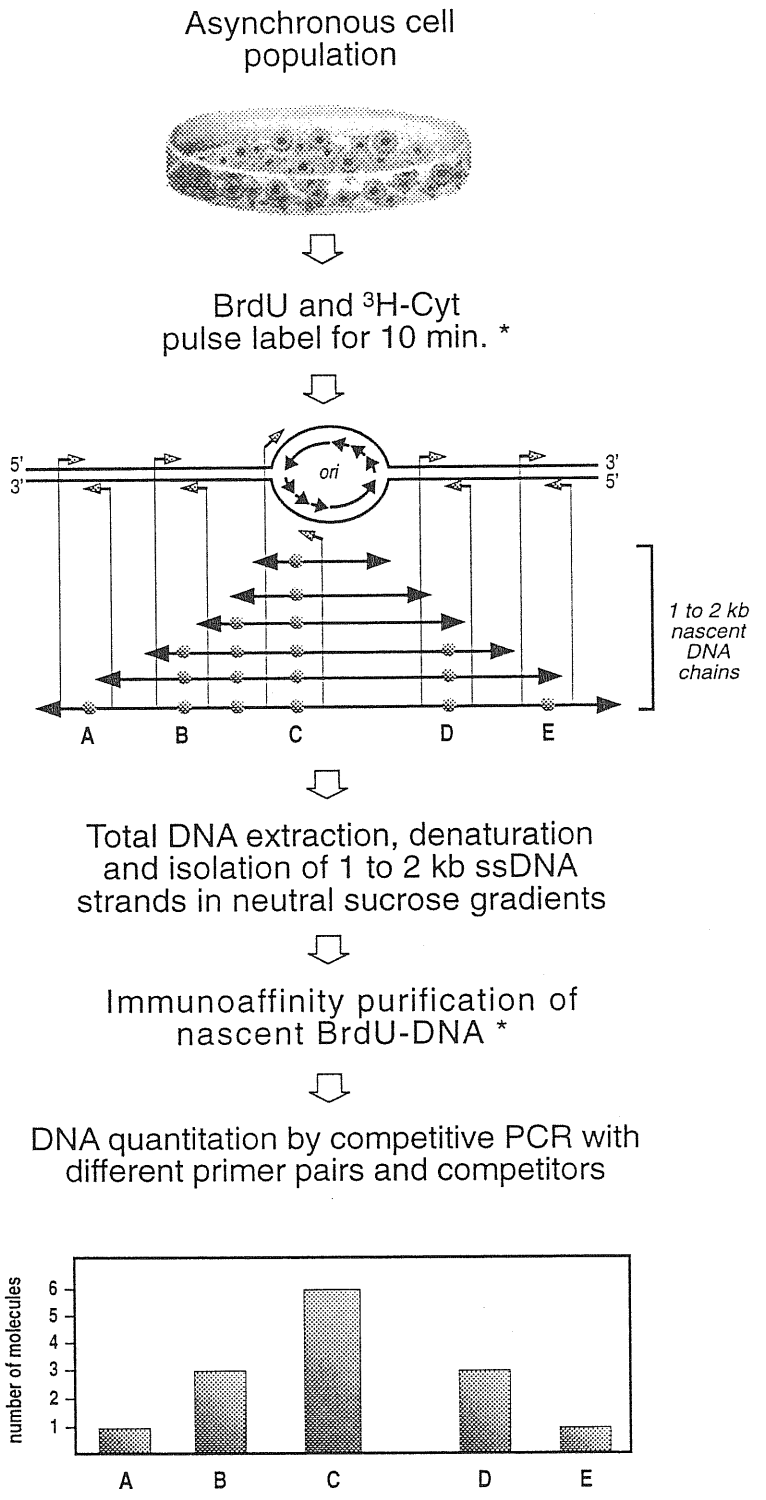


Figure 5. Mapping the localization of a DNA replication origin by quantitation of the abundance of neighboring markers within samples of nascent DNA.
Short fragments of nascent DNA, synthesized by exponentially growing cells from a bidirectional origin of replication (ori) after a short pulse with bromodeoxyuridine (BrdU) and [^{3}H]-deoxycytidine - the latter as a tracer -, are isolated from the bulk of parental DNA by size-fractionation on sucrose gradients and further selectively purified by immunoaffinity chromatography with anti-BrdU antibodies. Within this population of nascent DNA fragments, the number of molecules containing selected target DNA sequences (A through E) scattered within a genomic region are precisely quantified by competitive polymerase chain reaction. The pair of primers (small arrows on top) amplifying from the highest number of molecules in the sample is the closest to the origin (marker C in the Figure).
For a further simplification of the procedure, the labeling step and the subsequent affinity chromatography step can be omitted; these steps are indicated by a star.

of the abundance of different DNA segments scattered along the genomic region of interest. Precise quantification is obtained by a quantitative PCR procedure based on competitive PCR. This assay allows quantification of a few tens of molecules of target DNA by the simultaneous co-amplification of a quantified competitor DNA added to the same reaction, nevertheless, being able to detect differences of 2-fold between the relative abundance of two target DNA regions (Sestini *et al.*, 1994). Consequently, when applied to closely spaced genomic segments, it allows the mapping of origins at high resolution.

Using this assay, a replication origin has been mapped to an ~0.45 kb locus in the human lamin B2 gene region (Giacca *et al.*, 1994.). Lamin B2 origin activity is strictly dependent on the proliferative state of the cells; no activity was detected in terminally differentiated myeloid cells (Giacca *et al.*, 1994) or in resting primary lymphocytes (Kumar *et al.*, 1996). Moreover, the human lamin B2 replication origin exhibits specific protein-DNA interactions (Dimitrova *et al.*, 1996).

In the originally described procedure (Giacca *et al.*, 1994), nascent DNA was first purified from total newly synthesized DNA according to its size and then further enriched from the bulk unreplicated DNA by affinity chromatography. The use of BrdU-labeling of nascent DNA in origin mapping experiments was originally introduced with the aim to obtain a high signal-to noise ratio. However, since the origin mapping procedure by nascent strand abundance detection relies in principle only on the selection of nascent DNA stretches, a simplification of this procedure was subsequently attempted by omitting BrdU-labeling and subsequent purification of BrdU-substituted DNA fragments.

We recently demonstrated that in spite of this omission neither the detectability of DNA segments by competitive PCR nor the relative abundance of the markers changed in any appreciable way (Kumar *et al.*, 1996). This observation clearly indicates that in non synchronized cells little if any significant fragmentation of DNA occurs during the nascent DNA extraction and purification procedures. The omission of the BrdU-labeling step greatly simplifies the overall protocol, avoids the

introduction of a possible bias in fragment selection on the anti-BrdU affinity column, and reduces the background resulting from artifactual DNA fragmentation by photo-damage of BrdU-DNA. Still, BrdU labeling can be conceived when origin mapping is performed in synchronized cells, that may suffer DNA strand breaks due to the synchronization treatment itself.
The simplified procedure (see Figure 5) now consists of only DNA extraction, size fractionation of short nascent DNA strands (~1 kb) and quantitation of target DNA abundance.

In this thesis, the competitive PCR-based procedure for mapping origins of DNA replication in mammalian cells is presented together with the results obtained from a high-resolution study on the single-copy hamster DHFR locus by using the same procedure. These results are discussed in view of the confusing conclusions derived from previous studies on the same region, however employing different mapping techniques.
Furthermore, it is described how a library of origin-enriched sequences has been constructed and utilized to develop a potential method for the identification of new origins of DNA replication in human cells.
Finally, with the purpose to study protein factors binding to and activating origins, reagents useful in this respect have been prepared. Since work on this topic is still ongoing in the laboratory, preparation of these reagents is presented as an appendix to the thesis.

MATERIALS AND METHODS

Cell culture and DNA labeling

Chinese Hamster Ovary (CHO) K1 cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum, 2 mM L-glutamine and 50 μg of gentamicin per ml.

Human myelocitic leukemia HL-60 cells were grown in RPMI 1640 medium supplemented with 10% fetal calf serum, 2 mM L-glutamine and 50 μg of gentamicin per ml.

For nascent DNA labeling, $1x10^8$ to $2x10^8$ CHO K1 cells, maintained at a high replication rate by splitting the culture 12-24 hrs in advance, were pulse-labeled for 10 to 15 min in a CO_2 incubator by the addition of 1 μM (final concentration) [^{3}H]-deoxycytidine (21.5 Ci/mmol, Amersham) and 100 μM (final concentration) unlabeled 5'-bromodeoxyuridine (BrdU; Boehringer Mannheim GmbH, Mannheim, Germany). While labeled deoxycytidine was used as a tracer, incorporation of BrdU into stretches of newly synthesized DNA was employed to allow further purification of nascent DNA. After labeling, cells were collected by trypsinization, washed once in cold PBS and once in cold RBS buffer (Tris 10 mM pH 7.4, NaCl 10 mM, $MgCl_2$ 3 mM) and resuspended in RBS buffer at ~$2.5x10^7$ cells/ml. For the experiments with the HL-60 cell line, 10^8 HL-60 cells were collected without any labeling and washed as described.

Nuclei were prepared by the addition of one volume of RBS buffer plus nonionic detergent octylphenoxy polyethoxy ethanol (NP-40; Calbiochem, La Jolla CA) at 0.4%. The suspension of nuclei was handled carefully and pipetted very gently to avoid lysis. After 10 min in ice, nuclei were pelleted by centrifugation (10 min 2000 rpm at 4°C in a swing bucket rotor), washed in RBS and resuspended in the same buffer at a concentration of $5x10^7$ nuclei/ml. One volume of lysis buffer (Tris 20 mM pH 8.0, EDTA 20 mM) was then added, containing 2% sodium n-dodecyl sulfate

(SDS; Calbiochem) and 500 µg/ml proteinase K (Sigma, St. Louis MI) and the suspension was incubated overnight at 56°C.
Total genomic DNA was then extracted by the standard phenol-chloroform-isoamylic alcohol procedure (Sambrook *et al.*, 1989).

Isolation and purification of newly synthesized DNA

Total extracted genomic DNA was denatured either by alkali (CHO K1 DNA) or heat treatment (HL-60 DNA). In the former case, the DNA sample is chilled on ice and NaOH is added from a 10 N stock solution to a final concentration of 0.5 N. After 10 min on ice, the pH of the solution is neutralized by the addition of an equal volume of HCl 10 N. Then Tris pH 8.0 is added from a 1 M stock solution to obtain a final concentration of 250 mM. The solution is then dialyzed against at least three liters of TNE buffer (Tris 10 mM pH 8.0, NaCl 100 mM, EDTA 1 mM; pH 8.0) at 4°C overnight. Alternatively, DNA is denatured by 10 min incubation in boiling water, followed by chilling in ice.
Denatured DNA was then loaded on 5% to 30% (w/v) linear neutral sucrose gradients in TNE (300 µg DNA in up to 500 µl/35 ml gradient) and centrifuged for 16 to 20 hrs at 20°C in a Beckman SW28 rotor at 26000 rpm. In parallel, a reference tube with a double stranded size marker DNA (containing different sized DNA fragments in the range of 500 to 5000 bp obtained by appropriate restriction enzyme digestion of plasmid DNAs) was also run on a identical gradient. In our experience, the use of alkaline gradients results in significant degradation of BrdU-substituted DNA. Gradients were fractionated from the top to avoid contamination with large molecular weight DNA that pellets. Fractions of 1 ml each were collected; those from the gradient with the markers were run on a 1% agarose gel. Using the separation pattern of the marker DNA on the gel as a sedimentation velocity reference, the fractions containing single stranded DNA with an average size of ~1000 nt were

chosen as representative of nascent DNA issuing from all origins of DNA replication of the genome (the cells being non synchronized). The precise isolation of nascent DNA by size is crucial in order to obtain samples that are enriched in origin-specific DNA: ~1000 to 2000 nt is the optimal size to avoid contamination of nascent DNA by Okazaki fragments deriving from all the genome (which are shorter) or by randomly broken genomic DNA (which is likely to be larger). In addition, this average length of nascent DNA is the optimal for PCR amplification, the size of the chosen amplification products being 150 to 300 bp, and gives a good resolution among closely spaced (<500 bp) genomic segments. The chosen fractions were pooled and dialyzed against TE buffer (Tris 10 mM pH 8.0, EDTA 1 mM).

Further purification of BrdU-substituted CHO DNA was obtained by immunoaffinity chromatography using anti-BrdU antibody coupled to Sepharose beads, as described in reference (Contreas *et al.*, 1992). Then, DNA was again dialyzed against TE buffer. Greater than 95% of BrdU-DNA is usually recovered by this procedure, eliminating any trace of parental, unlabeled DNA that may have contaminated the preparation. In our experience, this procedure is superior to immunoprecipitation of BrdU-DNA.

After dialysis, HL-60 unlabeled, single stranded DNA was directly used for the construction of a library of sequences deriving from replication origins. To this purpose, nascent DNA was treated with NaOH 0.5N (from a 10N stock) for 1 hr at 37°C to hydrolyze the RNA primer at the 5' end of newly replicated DNA, first dialyzed against Tris 50 mM pH 8.0, EDTA 1 mM and then against Tris 10 mM pH 8.0, EDTA 1 mM, both dialyses prolonged over night.

Primers for PCR amplification

Seven sets of four primers each were constructed in the hamster DHFR region according to the nucleotide sequence determined. The localization of these primer

sets is reported in Figure 6 and their actual nucleotide sequence is given in Table 2. Each primer set consists of two oligonucleotides used for PCR amplification (external primers, the first two ones of each set in Table 2) and two oligonucleotides used only for competitor construction (internal primers). The latter two primers consist of two common 5' tails of 20 nt (tail1: 5'-ACCTGCAGGGATCCGTCGAC-3'; tail2: 5'-GTCGACGGATCCCTGCAGGT-3') linked to specific sequences complementary to the respective genomic targets at their 3' ends.

The localization of primer sets used for PCR amplification of lamin B2 region are reported in Figure 7. Primer sets and competitors were already described (Giacca *et al.*, 1994). Briefly, two primer sets (B48 and B13) were used for competitive PCR quantification of nascent DNA. The former one (B48) detects the most abundant fragment in this region in nascent DNA samples; therefore, it is very likely to coincide with a start site for DNA replication. The latest (B13) detects a reference, low abundant DNA target in the same region, ~4.5 kb far from the initiation site.

Construction and quantification of competitors for competitive PCR

Competitors for competitive PCR experiments were constituted by the amplification products obtained by the two external primers with the addition of a 20 bp unrelated sequence in the middle, in order to allow identification of genomic target and competitor PCR products by polyacrylamide gel electrophoresis. The method for competitor construction is outlined in the Results section and described in detail elsewhere (Diviacco *et al.*, 1992; Grassi *et al.*, 1994).

Each preparation of competitor was carefully quantified. Quantification of the absolute competitor concentration was obtained by PCR amplification of the competitor itself in the presence of a radiolabeled precursor dNTP, followed by evaluation of the incorporated radioactivity. The actual competitor concentration was

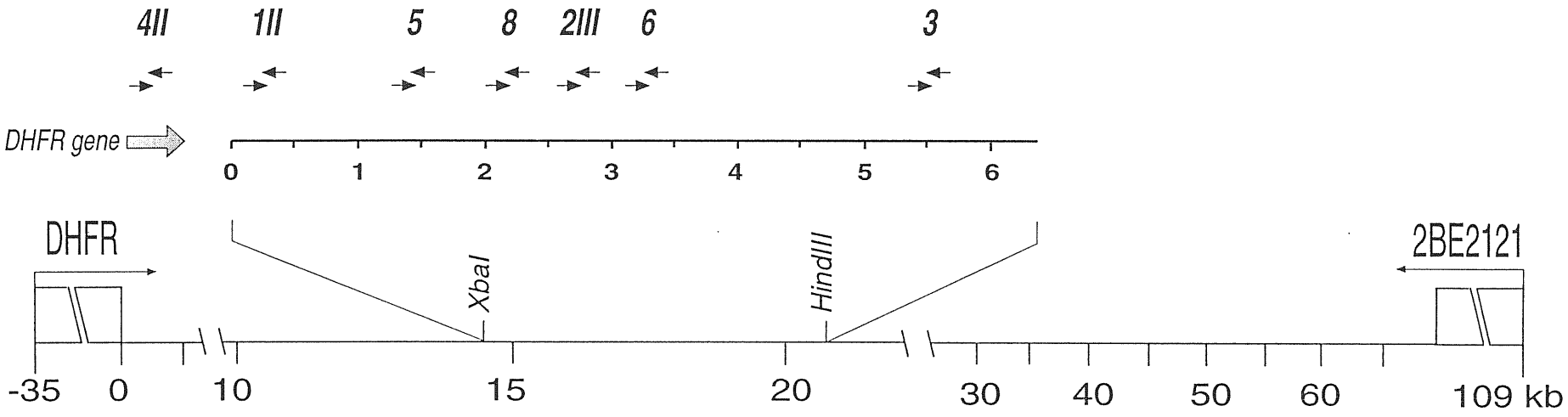


Figure 6. Schematic representation of the hamster DHFR gene and of the genomic area located ~15 kb downstream of it, where other mapping studies indicated the location of an origin of bidirectional DNA replication. In this region, seven primer sets were constructed. Six of them were selected within the intergenic region, and one primer pair was selected within the 3' end of the gene itself. The locations of these primers are indicated by converging arrows, and their nucleotide sequences are presented in Table 2. For each primer pair, a competitor DNA fragment for quantitative PCR was constructed and quantified as described in Materials and Methods.

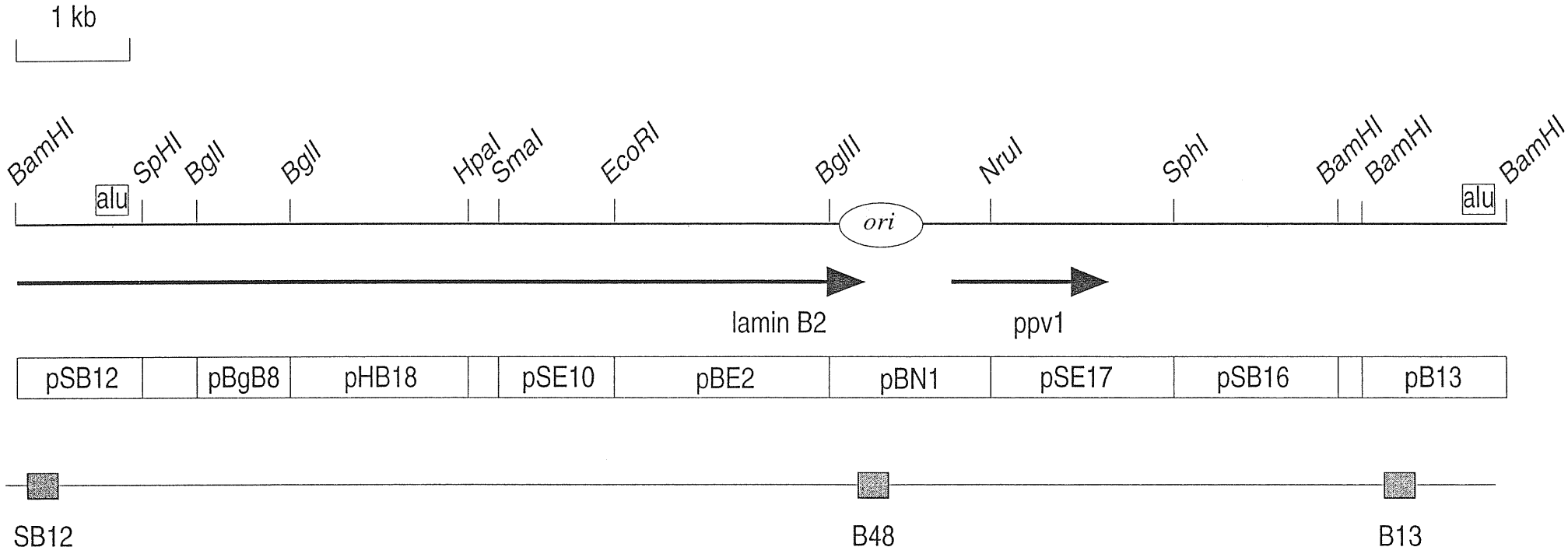


Figure 7. Schematic representation of the human lamin B2 locus. The restriction sites define the indicated subclones. The origin region is indicated and lies in the pBN1 subclone in the centre of the lamin B2 region. SB12, B48 and B13 are the markers used for the competitive PCR experiments on nascent DNA.

calculated according to the specific activity of the labeled dNTP and the nucleotide composition of the competitor sequence. However, careful determination of the molar concentration of competitor still does not formally guarantee that the target genomic DNA and the competitor fragment PCR amplify with the same efficiency. To overcome this limitation, the competitor concentrations were also evaluated by quantification of a sample of total genomic DNA obtained from the same cell line in which origin mapping had to be performed. This was obtained by the co-amplification of scalar amounts (10-fold dilutions, followed by progressively closer dilutions in the range of equivalence) of competitor against a fixed amount of the total DNA sample. Consequently, the number of copies of each target DNA sequence in the nascent DNA sample was expressed as a ratio to the number of copies of the same target in total DNA. This procedure corrects for differences in relative abundance of nascent DNA that might arise simply as a result of differences in the ability of amplifying different target sequences. Careful quantification of competitor DNA was repeated each time one mapping of the same origin was repeated.

Competitive PCR experiments

Quantification of the abundance of the DNA segments amplified with the chosen primer sets was obtained in samples of nascent DNA by competitive PCR. The experiments were carried out by mixing a fixed quantity of BrdU DNA (5 µl of each preparation) with scalar quantities of competitor DNA fragments. Rough quantitations were obtained by tenfold dilutions; precise quantitations were then obtained by using competitor dilutions in the equivalence range.

Conditions for PCR amplification were as follows: denaturation at 94°C for 30 sec; annealing at the temperatures reported in Table 2 and Table 3 for each primer set for 30 sec; extension at 72°C for 30 sec. Fifty cycles of amplification were performed for each amplification experiment. PCR reactions are assembled in a final volume of 50

Primer set	Primer name	Location (nt)	Sequence	Size (bp) of PCR product	Annealing temp (°C)
1II	1DXII	377-358	5'-AAAAGCCACTCAAGCGCCAA-3'	225	54
	1SX	153-172	5'-GGAAGGATTGTAGGAGCCAA-3'		
	1PERDXII	275-294	5'-tail1+CAACCTAGGAACTCTGCATG-3'		
	1PERSXII	274-255	5'-tail2+GCCTCATGCAGTTCCCTAAT-3'		
2III	2DXIII	2879-2860	5'-CCTTCATGCTGACATTTGTC-3'	278	64
	2SXIII	2602-2621	5'-GTCCCTGCCTCAAAACACAA-3'		
	2PERDXIII	2679-2698	5'-tail1+CTCAGTGAGTCCACTTCTTT-3'		
	2PERSXIII	2678-2659	5'-tail2+AAGGAAGGAAAGAAAGGCCC-3'		
3	3DX	5652-5633	5'-GCTGGGATAAGTTGAAATCC-3'	259	54
	3SX	5394-5416	5'-GGACACTAAGTCTAGGTACTACA-3'		
	3PERDX	5524-5543	5'-tail1+AGGACTCAGCTCTTACTAAC-3'		
	3PERSX	5523-5504	5'-tail2+TAGGAAACTGAGATGCCAGG-3'		
4II	4DX	1898-1877	5'-TATGGGCTACCTCCTTAGGAGC-3'	213	60
	4SXII	1686-1705	5'-CCAGTGATATGCATAGCACC-3'		
	4PERDX	1821-1840	5'-tail1+CCATCCAAGAGCAATGGCAA-3'		
	4PERSX	1820-1801	5'-tail2+TCACCACAGCTAGATGGTAA-3'		
5	5DX	1546-1527	5'-GAGCTAGGAGGATCCATTCT-3'	210	60
	5SX	1337-1356	5'-GAGACGAGGGATTTACTCTA-3'		
	5PERDX	1438-1457	5'-tail1+TGGAATGCTCTCTCTAGCTT-3'		
	5PERSX	1437-1418	5'-tail2+TGAACACAGACACTATGGTC-3'		
6	6DX	3210-3191	5'-AACCTCTGAACTGTAAGCTG-3'	169	62
	6SXbis	3041-3061	5'-GAACTGGCTTCCCAAGAAAT-3'		
	6PERDX	3133-3151	5'-tail1+TGCTGTGAAGAGACACCATG-3'		
	6PERSX	3132-3113	5'-tail2+ATAGAAACCCCAGCTAAGAC-3'		
8	8DX	2205-2186	5'-GTCCTCGGTATTAGTTCTCC-3'	204	56
	8SXII	2002-2021	5'-CTCTCTCATAGTTCTCAGGC-3'		
	8PERDX	2081-2100	5'-tail1+TCCATGGCAGTCTTCACACT-3'		
	8PERSX	2080-2061	5'-tail2+CATTCATCAAGCTGGAAAGC-3'		

Table 2. Primer sequences and PCR parameters for amplification of the DHFR region.The first two primers of each primer set were used for amplification from CHO genomic DNA; their localization is reported in Figure 6. The last two primers of each set were used for competitor construction. The sequences of tail1 and tail2 are reported in Materials and Methods. Nucleotide numbering is referred to sequence *cgdhfrori* in GenBank (although the actual nucleotide sequences are according to the new sequencing data obtained for CHO K1 cells) for all the primer sets with the exception of primer set 4II whose nucleotide numbering is referred to sequence *crudhfraa* in GenBank.

Primer set	Primer name	Location (nt)	Sequence	Size (bp) of PCR product	Annealing temp (°C)
SB12	SB12SX2	156-175	5'-CACACCTTGGCCTGGTAGGT-3'	109	60
	SB12DX	264-245	5'-GAGTACGACTTCAAGATGGC-3'		
	SB12PERSX	212-231	5'-tail1+GTCGTGCTGGCTCCGCAGCT-3'		
	SB12PERDX	264-245	5'-tail2+AGCAAGTGCGGCTCTACAAG-3'		
B48	B48SXL	4104-4077	5'-TAGCTACACTAGCCAGTGACCTTTTTCC-3'	168	60
	B48DXII	3937-3956	5'-GACTGGAAACTTTTTTGTAC-3'		
	B48PERSX	3999-3980	5'-tail1+CTCAGTGAGTCCACTTCTTT-3'		
	B48PERSX	4000-4019	5'-tail2+AAGGAAGGAAAGAAAGGCCC-3'		
B13	B13SX	54-73	5'-CCTCAGAACCCAGCTGTGGA-3'	164	60
	B13DX	217-198	5'-GCCAGCTGGGTGGTGATAGA-3'		
	B13PERSX	153-172	5'-tail1+CTCCAGAGGCGTGTTTTCCT-3'		
	B13PERDX	152-133	5'-tail2+ACAGCAGGTGGGTATGGGAC-3'		

Table 3. Lamin B2 region primer sequences and PCR parameters. The first two primers of each primer set were used for amplification from human genomic DNA; their localization is reported in Figure 7. The last two primers of each set were used for competitor construction. The sequences of tail1 and tail2 are reported in Materials and Methods. Lamin B2 primers utilization and competitors construction have already been described in (Giacca et al, 1994).

µl, containing both primers 1 µM, the four dNTPs 200 µM each, 1 U of *Taq* DNA polymerase (Amplitaq, Perkin Elmer, Emmeryville CA), 5 µl of the nascent DNA sample, and 5 µl of appropriate scalar dilutions of competitor DNA.

DNA sequencing

A segment of the genomic CHO K1 cell DNA in the DHFR genomic region (nucleotides 1500-3210 of sequence *cgdhfrori* in GenBank) was sequenced by the dideoxy-chain termination method (T7 Sequencing™ Kit-Pharmacia Biotech). PCR primers chosen in this region were constructed according to the sequence obtained, which, in several instances, was diverging from the published one (Caddle *et al.*, 1990) . The updated nucleotide sequence, obtained in collaboration with E. Fanning, A. Schmidt and M. DePamphilis, is available in the EMBL nucleotide sequence data base (accession number X94372)

Construction of a library of origin-enriched sequences

Two fractions of HL-60 ~1 kb single stranded nascent DNA were pooled together, treated with 0.5 N NaOH to remove the RNA primer at the 5' end of newly synthesized DNA, dialyzed, precipitated and then resuspended in 10 µl of water. Since nascent DNA consists of a random population of fragments deriving from all the origin regions, specific tails have to be added for its amplification and cloning. To this purpose, an oligo (dG_{15-35}) tail was added to the 3' end of ssDNA using terminal deoxynucleotidyl transferase (TdT, Pharmacia Biotech); conditions for tailing were those suggested by the manufacturer: 1 µM dGTP, 24 units of TdT, 20' of incubation at 37°C. This tail allows the synthesis of the complementary strand by annealing oligo Sal/Cla-dC14 (5'-

AAGGATCCGTCGACATCGATAACGACCCCCCCCCCCCCC-3'), containing a stretch of 14 cytosine nucleotides and a 5' tail bearing Bam HI, Sal I and Cla I restriction sites, and extending it with *Taq* DNA polymerase (Figure 8). This reaction was performed according to the conditions suggested by the manufacturer, with the exception that the amount of the oligo was 0.5 pmoles per reaction and the final volume was 100 µl. The dG-tailed DNA sample was denatured at 95°C for 5', then oligoSal/Cla-dC14 annealed at 58°C for 10' and extended at 72° for 10' in a termocycler. Only one cycle was performed. Usually, 1/10 of the single strand, dG-tailed nascent DNA template was used in parallel to check the second strand synthesis reaction either by [α-^{32}P]dCTP (Amersham, UK; 3000 Ci/mmole; 10 mCi/ml) incorporation or by extension of the hot Sal/Cla-dC14 primer (see Figure 14 and part II of Results). Further, dsDNA was precipitated, phosphorylated (T4 polynucleotide kinase, Pharmacia, Biotech) and ligated (T4 DNA ligase, Biolabs) to a double strand asymmetric linker as already described in (Dimitrova *et al.*, 1996). However, since *Taq* DNA polymerase is capable of adding a single nontemplate-directed dATP to a blunt end ds DNA fragment causing problems for blunt-ended ligation, we experienced that the use of a mix of two ds linkers, one of witch containing a protruding T, greatly increases the efficiency of ligation. The effective ligation of the linker was also checked on an aliquot of the sample by annealing and extension of oligo LLONG (5'-GCGGTGACCCGGGAGATCTGAATTC-3') end-labeled with [γ-^{32}P]ATP (Amersham, UK; 3000 Ci/mmole; 10 mCi/ml). This oligo is complementary to the shortest strand of the linker. The nascent DNA sample, provided in this way with specific tails, was purified from the excess of linker (Microcon, Amicon), and amplified using oligo LLONG and Sal/Cla (5'-ATCCGTCGACATCGATAACG-3'), the last one annealing to internal sequences of the Sal/Cla-dC14 tail. Conditions for PCR were the following: denaturation at 94°C for 1 min, annealing at 58°C for 2 min and extension at 72°C for 4 min; 15 cycles were performed. All the oligo used for library construction and amplification are reported in Table 4. These PCR products were purified and cloned in a pUC18 plasmid, by

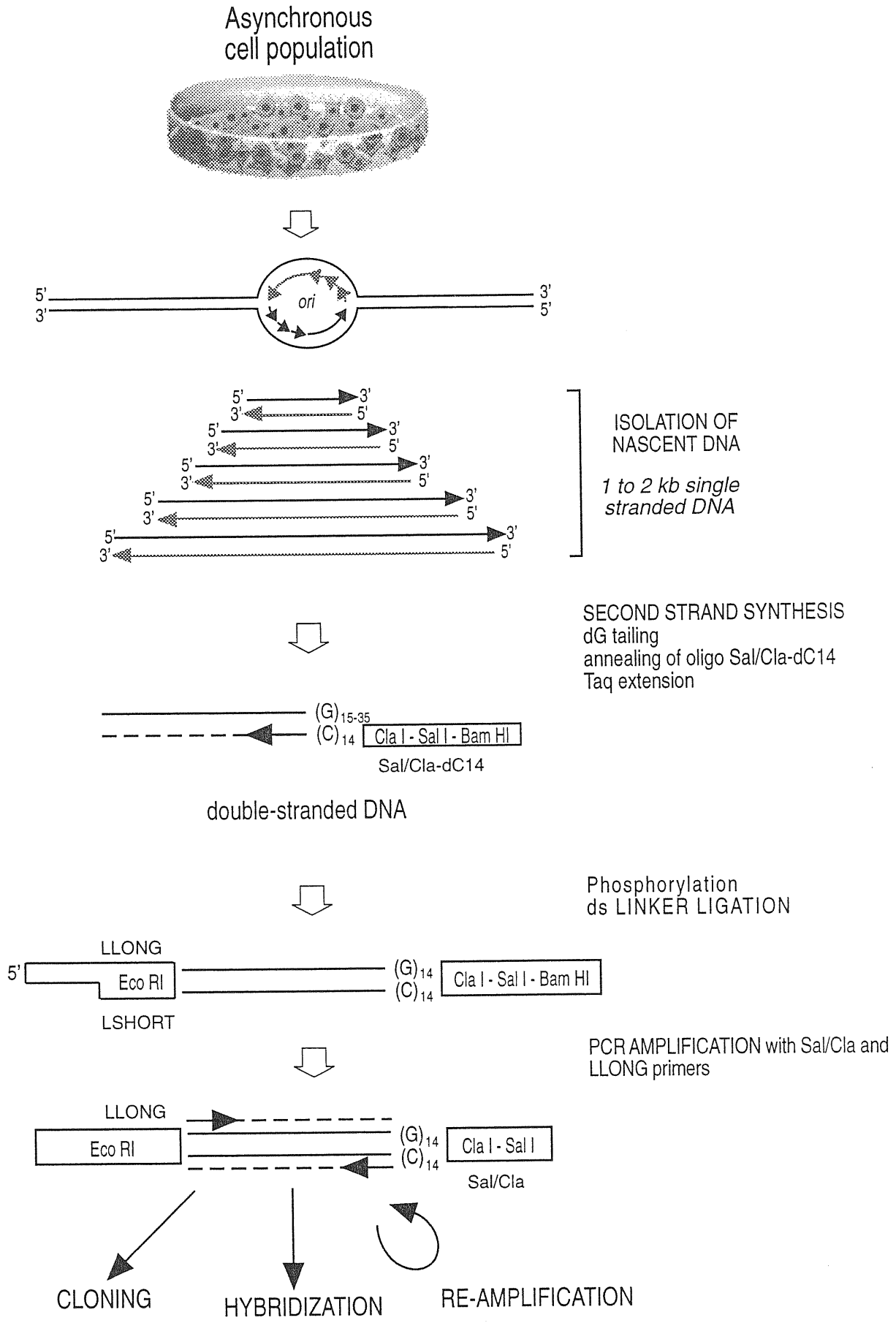


Figure 8. Flow chart of ori-library construction.

A

Primer name	Primer length (bp)	Sequence (5'→3')
Sal/Cla-dC14	39	AAGGATCCGTCGACATCGATAACGACCCCCCCCCCCCCC
Sal/Cla	20	ATCCGTCGACATCGATAACG
LSHORT	11	GAATTCAGATC
LLONG	25	GCGGTGACCCGGGAGATCTGAATTC
LLONG+T	26	GCGGTGACCCGGGAGATCTGAATTCT

B

a) 5'-**GCGGTGACCCGGGGAGATCTGAATTC**-3'
3'-**CTAGACTTAAG**-5'
Eco RI

b) 5'-**GCGGTGACCCGGGGAGATCTGAATTCT**-3'
3'-**CTAGACTTAAG**-5'
Eco RI

Structure of the Eco RI ds linkers

Table 4. Primers for library construction. Primer sequences are given in panel A. Structure of asymmetric linkers is reported in panel B. Primers utilization is described in Material and Methods and illustrated in Figure 8. Briefly, oligo Sal/Cla-dC14 was used for second strand synthesis by annealing to single stranded dG-tailed nascent DNA by means of its dC tail. Oligo LSHORT was annealed either to LLONG or LLONG+T to obtain the asymmetric ds linkers. Oligo LLONG and Sal/Cla, matching to the specific tails added to nascent DNA, were used for library PCR amplification.

means of Sal I and EcoR I restriction sites. Transformation of DH5α E.coli cells was obtain by electroporation.

Alternately, the nascent DNA library was used for hybridization experiments. Aliquots of the library were spotted on nylon membranes (Stratagene) and probed with three fragments of the lamin B2 region; conversely, the library was used as radioactive probe in hybridization to immobilized plasmids from the lamin B2 region as well. A small aliquot of the first amplification product was diluted 50 to1000 folds and amplified again with the same oligos LLONG and Sal/Cla for 30 PCR cycles. This product was also cloned and used in hybridization experiments.

Therefore, the ori-library can be handled either as a cloned product and therefore amplified in bacteria or as a PCR product amplifiable with oligos LLONG and Sal/Cla.

Hybridization experiments

Nascent DNA PCR products were purified form oligos and salts (Microcon, Amicon) before any use. For dot-blot hybridization, aliquots of the library (1/20 and 1/200 of the PCR reaction) were loaded on nylon filters (Stratagene) together with scalar amounts of total chromosomal DNA as recommended by the manufacturer using a manifold dot-blotter (Bio-Rad). Seven serial dilutions (from 1 μg to 5 ng) of total DNA were used as a reference. Positive (the same as the probe) and negative controls (pUC18 plasmid) were also included. Filters were probed with three lamin B2 subclones: one in the center (pBN1) and two at the both sides (pSB16 and pBgB8) of the lamin B2 region, spanning a distance of ~8.500 nt. These plasmids were chosen as representative of fragments containing respectively the origin region (pBN1) and sequences far from it (pBgB8 and pSB16). Each filter was hybridized with the specific probe and dot blots quantified; the intensity of hybridization to nascent DNA relative to total DNA is a measure of the abundance of each fragment in nascent DNA. It is expected that the fragment containing an origin of DNA replication is the one

enriched in a library of origin sequences, while the other two fragments far from the initiation site, are less represented.

Conversely, nascent DNA was then used as radioactive probe in hybridization to cloned genomic DNA fragments. The rational of the experiment is that a nascent DNA library preparation (that is a collection of fragments deriving from ori regions) will preferentially hybridize to cloned fragments containing origin sequences. Therefore, using plasmids or cosmids containing fragments which span a region of interest, an origin of replication will be identified by the cloned fragment which gives the stronger hybridization signal (paying most attention to avoid cloned repetitive sequences or Alu repeats). Briefly, plasmids containing fragments from the lamin B2 locus (pBN1, pSB16 and pBgB8) were digested, run on a 1% agarose gel and transferred on nylon membranes (Sambrook *et al.*, 1989). Usually, 1 μg of each digestion was immobilized on filters; 1/10 to 1/20 of the PCR product representing the library was labeled with [α-^{32}P]dCTP (3000 Ci/mmole) by the random primer technique as specified by the supplier (Pharmacia, Biotech) and used at $2x10^6$ cpm/ml. As a control of hybridization efficiency among different sequences, total genomic DNA from human cells (200 ng) was sonicated to obtain 500-1000 bp fragments, labeled as described and used in hybridization on plasmids. It is expected that the three fragments of the lamin B2 locus being equally represented in genomic DNA, give the same intensity of hybridization. However, since the amount of each sequence in a total DNA probe is under the detectability threshold, we could hardly see hybridizations (data not shown). Most importantly, neither of the cloned sequences showed preferential hybridization to total DNA.

Hybridization conditions were as follows: 6X SSC, 5X Denhardt's solution, 0.5% SDS, 100 μg of herring sperm DNA per ml at 68°C for 14-18h; washing was done in 1X SSC-0.1% SDS at 68°C for 20 min, followed by a second washing in 0.5X SSC-0.1% SDS in the same conditions.

The plasmid pUC18 was used as a probe to check possible differences in transfer efficiency among the three samples.

Analysis of the library of nascent DNA strands

The PCR product representing the library of nascent DNA sequences was analyzed by competitive PCR before use in hybridization experiments. Most important to determine was its representativity of origin sequences. To this purpose, markers B48 and B13 in the lamin B2 region were quantified in the nascent DNA preparation before amplification, but also after tailing and the first amplification, and after the second re-amplification of the library.

If any bias or selection have been introduced during the construction of the library, these two fragments have to be represented after PCR as in the original preparation, being B48 fragment 10-15 times more abundant than B13 (Giacca *et al.*, 1994; Kumar *et al.*, 1996). Therefore, the relative abundance of the two markers, in addition of being an indication of lamin B2 origin activity, has been used as a assay to test the content of the library.

PCR amplification were performed as already described. A fixed amount of library DNA (diluted 1/50 to 1/1000) was coamplified with scalar amounts of competitor DNA in a 50 µl standard reaction. PCR products were resolved on 8% polyacrylamide gels and the two bands (competitor and genomic DNA) were quantified.

RESULTS

PART I

Mapping DNA replication origins by quantifying the relative abundance of nascent DNA strands using competitive PCR

During my pre-doctoral work at the Molecular and Cell Biology of ICGEB, I contributed to the development of a novel procedure for mapping origins of DNA replication in mammalian cell chromosomes. This procedure is based on determining the relative abundance of nascent DNA strands throughout a specific genomic region. More in detail, the method entails the purification of short strands of nascent DNA derived from recently activated origins and the quantification, within this sample, of the relative abundance of different adjacent DNA segments. It is expected that the abundance of defined markers within the origin region is greatest at the site where DNA replication begins.

Precise quantification of the abundance of different segments in the region of interest is obtained by means of competitive PCR. This technique is based on the co-amplification of the target DNA with known amounts of a competitor DNA fragment sharing the same primer recognition sites. Since both templates compete for amplification, any variable affecting the reaction has the same effect on both species. As a result, the two templates are amplified at the same rate and the amount of target DNA can be calculated from the ratio between the final amplification products. This ratio precisely reflects the ratio between the input amounts of the two molecules; the number of competitor molecules added being known, the number of target molecules in the sample can be precisely measured.

Outline of the procedure

For each genomic DNA target to be quantified, a primer set is selected and a competitor DNA fragment is constructed. When any indication exists on the location of the origin, equally spaced primer sets scattered along the region of interest and

lying 1-2 kb apart are initially chosen. According to the results obtained by PCR quantification, additional primers are then selected to refine the mapping and narrow down the initiation site. As a control, a primer set detecting a reference DNA target far from the region under investigation is usually chosen to evaluate the background (possibly resulting from contamination from non specifically broken DNA) in the same nascent DNA sample.

A primer set consists of two oligonucleotides used for PCR amplification (external primers) and two oligonucleotides used for competitor construction (internal primers; Figure 9, a). The first two primers are chosen to amplify segments of 150 to 300 nt. Amplification products of this length are well resolved by 8% polyacrylamide gel electrophoresis and are suitable for high-resolution studies on ~1 kb nascent DNA, as discussed above. The rules to follow for the choice of the external primers are not different from those to follow for the set up of any PCR experiment (primer GC/AT ratio should be ~1, primer should be devoid of intra- and inter-complementary regions, genomic target regions that are too GC- or AT-rich should be avoided, target DNA sequences should be present in single copy in the genome). While the use of computer programs for the optimization of primer sequences and the prediction of optimal annealing temperatures are helpful (e.g. Amplify, by Bill Engels, University of Wisconsin, 1992), as a matter of fact the performance of any primer pair must be experimentally evaluated. Similarly, the PCR cycling profiles have to be individually optimized. This is usually accomplished by amplification at progressively increasing annealing temperatures and for different extension times. As a general rule, all PCR amplifications were performed at the highest annealing temperature allowed by the chosen primer pair (preferably >60°C) and for the shortest time required by each cycle step (≤30 sec). We usually prefer primer pairs not giving rise to the formation of aspecific PCR products.

Competitors were constructed by an application of the recombinant PCR methodology, using the internal primers of each primer set. The sequences of these

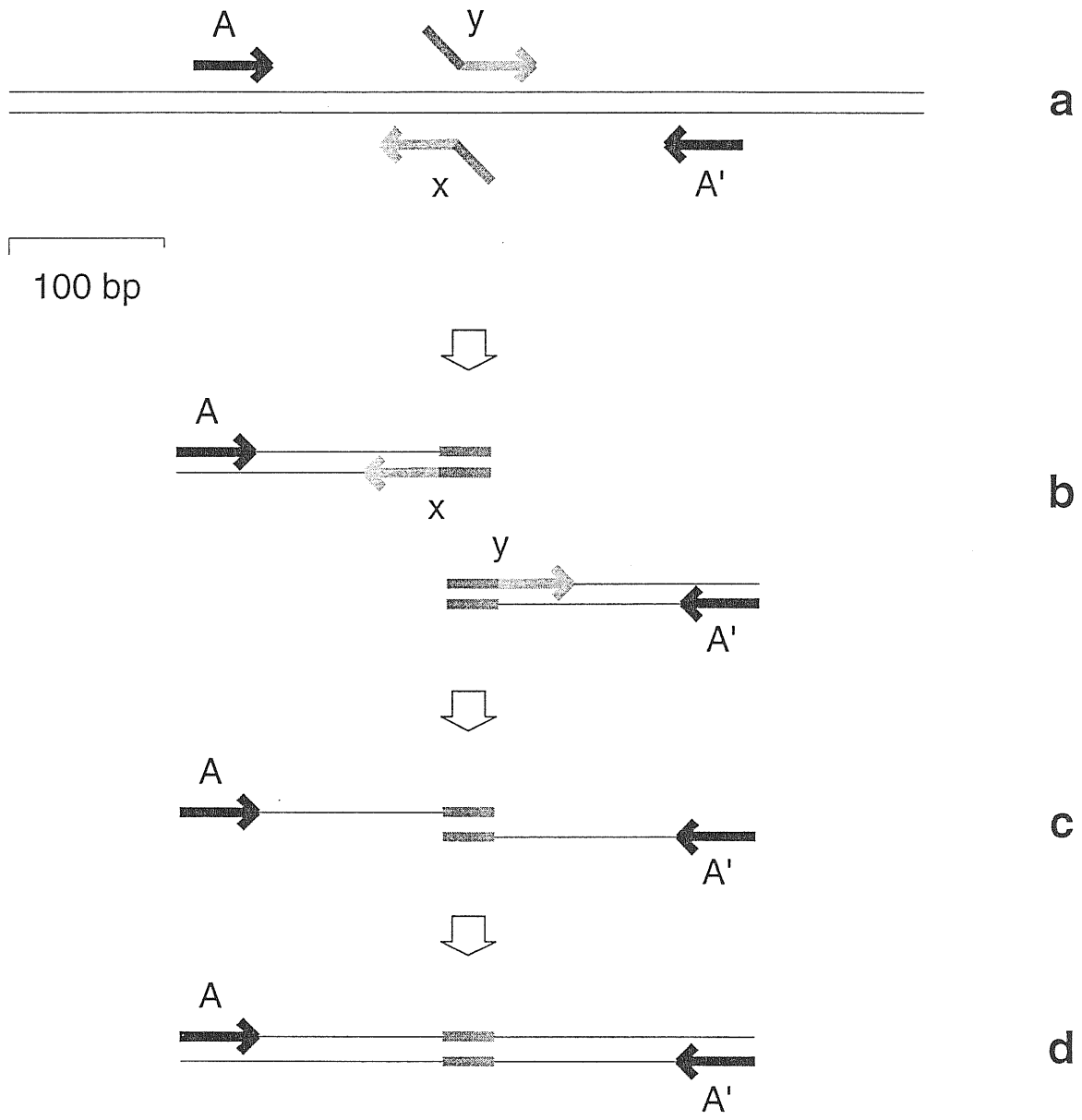


Figure 9. Schematic representation of competitor construction by recombinant PCR. For each segment to be quantified, a set of primers is synthesized, composed of two external primers (A and A') and two internal primers (x and y). The nucleotide sequences of the 3' ends of primers x and y are identical to contiguous sequences on opposite strands of the target DNA. Additionally, these primers also contain an ~20 nt tails at their 5' ends. These tails are complementary one each other [e.g. tail 1: 5'-GTCGACGGATCCCTGCAGGT-3' and tail 2: 5'-ACCTGCAGGGATCCGTCGAC-3']. With these primers, two separate amplifications are carried out using pairs A-x and A'-y to obtain hemi-competitor products. These two amplification products are eluted from a 8% polyacrylamide gel, denatured, mixed, annealed by virtue of the complementarity of the 5' tails of primers x and y and amplified using external primers A and A'. With the exception of the ~20 bp insertion, the competitor DNA fragment contains exactly the same sequence as genomic DNA.

primers correspond to contiguous sequences on opposite strands within the amplified PCR fragment. In addition, the two internal primers bear 5' tails containing two ~20 nt sequences which are unrelated to the target DNA and complementary to each other. The competitor fragments are directly obtained from the amplification products of genomic DNA, using a two step procedure. First, each of the external primer is used in combination with the internal primer on the opposite strand in two separate PCR amplifications (Figure 9, b). Subsequently, the two amplification products are recovered, mixed, denatured, reannealed by virtue of the complementarity of the two tails of the internal primers (Figure 9, sc), and re-amplified with the external primers (Figure 9, d). The description of this procedure is detailed in reference (Diviacco *et al.*, 1992). With the exception of the ~20 bp insertion corresponding to the 5' tails of the two internal primers - which allows the identification of the two PCR products by polyacrylamide gel electrophoresis -, the competitor DNA fragment contains the same sequence as the target genomic DNA.

Competitive PCR experiments for the quantification of the abundance of different DNA segments in the nascent DNA preparations were performed by mixing scalar dilutions of competitor DNA to a fixed amount of nascent DNA followed by amplification with the appropriate primer pair. Since the amplified segments are usually short (150 to 300 bp), 15 to 30 seconds for each PCR step are sufficient. This limited amount of time prevents inactivation of the *Taq* DNA polymerase and the performance of several PCR cycles. As a consequence, even if starting from very few molecules (less than 100) of nascent DNA, the final PCR products can be directly visualized after staining of 8% polyacrylamide gels by ethidium bromide. It should be reminded that the principles of competitive PCR ensure that the maintenance of the exponential phase of amplification is not required.

The procedure for competitive PCR is outlined in Figure 10. A fixed amount of nascent DNA is mixed with scalar amounts of competitor and amplified with the specific primers. At the beginning, 10-fold scalar dilutions of competitor are used, to

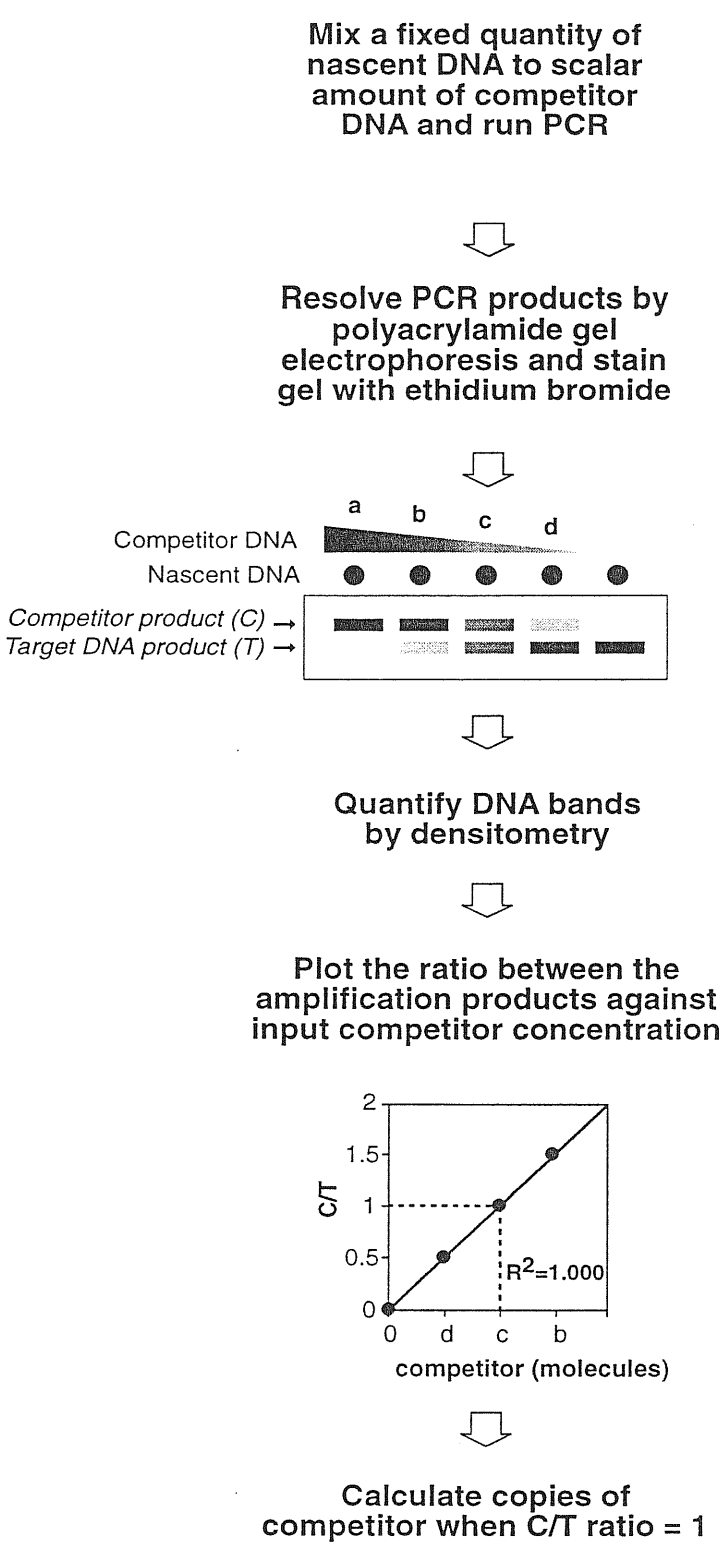


Figure 10. Competitive PCR. A typical competitive PCR experiment is performed by the addition of scalar amounts of competitor DNA in known concentration to a fixed amount of the DNA to be quantified. The amplification products are resolved by gel electrophoresis, stained with ethidium bromide, and the peak areas corresponding to the bands of the amplification products of the competitor and target DNAs are quantified by densitometric scanning. According to the principles of competitive PCR, the ratio between the amplification products for the two species (competitor/target, C/T) is linearly related to the input amount of the competitor molecules added to the reaction. The experimental points are fitted by a linear function (R^2: correlation coefficient), and, according to the equation describing the line, the amount of competitor corresponding to a 1:1 amplification product ratio is calculated. This is the same amount of target genomic molecules present in the nascent DNA sample.

roughly estimate the target concentration; precise quantifications are then obtained by using competitor dilutions in the equivalence range. The PCR products are then resolved by gel electrophoresis, and the intensity of each band (genomic and competitor) is determined by densitometric scanning. According to the principle of competitive PCR (Diviacco *et al.*, 1992; Siebert and Larrick, 1992), the ratio between the two molecular species remains unchanged during the amplification process. Since the amount of input competitor is known, and the final ratio between the PCR products is measured, the initial amount of target DNA can be calculated in a simple way.

By the analysis of adjacent DNA segments scattered along a region of interest, the origin region is expected to be identified by the peak of a gaussian distribution of segment abundance in the nascent DNA sample. On the contrary, a flat distribution of marker abundance means that no origin activity is associated with the analyzed region (Giacca *et al.*, 1994; Kumar *et al.*, 1996). The gaussian distribution of segment abundance peaking at the replication origin - instead of the theoretically expected sharp peak - is likely to be due to the heterogeneous length of the nascent DNA fragments obtained by sucrose gradient purification. The collection of nascent DNA fragments with average size of 1000 nt has a very wide distribution, producing a smear in the range of ~300 to ~2000 nt. Additionally, it is likely that the nascent DNA sample might also be contaminated by randomly broken total genomic DNA fragments or, more probably, by clusters of ligated Okazaki fragments. This most likely explains the reason why a background of target DNA molecules is always present even far from the origin as well as in non proliferating cells. Consequently, origin activity has to be defined not by the absolute abundance of a single DNA segment in nascent DNA samples, but by the analysis of the relative abundances of different target DNAs scattered in the region of interest.

Mapping the DHFR origin region in single copy CHO K1 cells by analysis of nascent DNA strands abundance by competitive PCR

The mapping procedure, described in the previous section, was applied to the single-copy DHFR gene domain in order to identify the localization of an origin of DNA replication. For this purpose, exponentially growing CHO K1 cells were pulse-labeled for 15 minutes with ^{3}H-deoxycytidine and bromodeoxyuridine. During this period of time, stretches of newly synthesized DNA deriving from all portions of the genome (and thus being at different distances from origins of DNA replication) were labeled. Subsequently, cells were collected and total genomic DNA was extracted and denatured by alkali treatment. Within this sample, short stretches of single-stranded newly synthesized DNA (i.e. nascent DNA which is the closest to the origin) were size-selected by sedimentation through neutral 5-30% sucrose gradients. Fractions of average size of ~1000 nt were collected as representative of nascent DNA emanating from all origins of DNA replication of the genome. In our experience, this size is optimal in order to avoid contamination by Okazaki fragments (which are shorter) or contamination by randomly broken genomic DNA (which is likely to be larger). Additionally, this average length of nascent DNA is optimal for PCR amplification, the size of the chosen amplification products being of 150-300 bp, and gives a satisfactory resolution among closely spaced genomic segments, contrary to the results obtained with longer fragments.

Further purification of newly synthesized, BrdU-substituted DNA was obtained by batch chromatography with an affinity resin bearing anti-BrdU antibodies, according to an already published procedure (Contreas *et al.*, 1992). This final purification step is used to completely eliminate any trace of parental, unlabeled DNA possibly contaminating the preparation by random breakage during the extraction procedure. Within these samples of nascent DNA, competitive PCR experiments for origin mapping were performed by the analysis of seven segments in the DHFR locus. Since, at first, we experienced the problem that several chosen primer pairs in this

region were not amplifying from total CHO DNA, the nucleotide sequence of a DNA region downstream of the DHFR gene (corresponding to nucleotides 1500 to 3210 of file *cgdhfrori* in GenBank) was re-determined by the analysis of overlapping PCR products. The new nucleotide sequence was found to differ in several positions from the published one; thus, the actual sequences for the synthesis of PCR primers were chosen according to the newly determined sequence.

Seven primer sets were synthesized (Table 2), whose localization is schematically shown in Figure 6; one of these sets is located within the 3' non-translated region of the DHFR gene, while the other six ones are positioned in a ~6 kb region (Caddle *et al.*, 1990) located ~15 kb downstream of the gene. For each primer set, a competitor DNA fragment was constructed, containing exactly the same sequence as the target genomic DNA segment, with the exception of a 20 bp insertion in the middle to allow resolution of the genomic and competitor amplification products by polyacrylamide gel electrophoresis. These competitors were constructed by a recombinant PCR procedure starting directly from the genomic amplification products, as already described (Diviacco *et al.*, 1992). The sequences of the internal primers used for the generation of these competitors, bearing, at their 5' ends, 20-nt tails complementary to each other which subsequently become the 20-bp insertion tags of the competitors, are given in Materials and Methods.

A competitive PCR experiment is shown in Figure 11. A fixed amount of nascent DNA was mixed with increasing amounts of competitor for each primer set, and amplified. The PCR products were resolved by gel electrophoresis and the intensity of each band (genomic and competitor) was determined by densitometric scanning. According to the principles of competitive PCR, the ratio between the two molecular species remains unchanged during the whole amplification process. Since the number of added competitor molecules is known, and the final ratio between the PCR products is measured, the initial amount of target DNA can be calculated in a simple way. These measurements are given on the right sides of each of the gels shown in Figure 11.

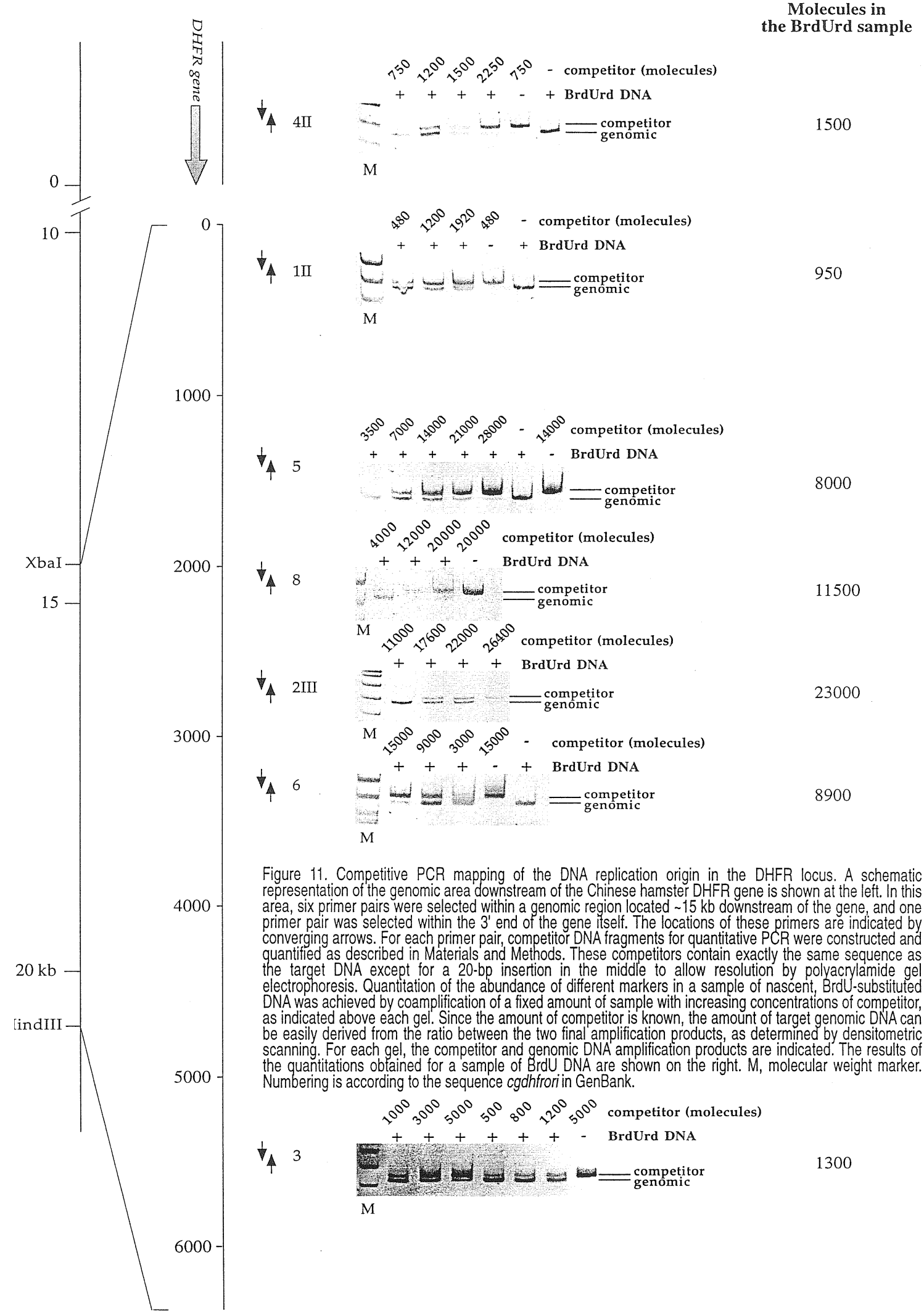


Figure 11. Competitive PCR mapping of the DNA replication origin in the DHFR locus. A schematic representation of the genomic area downstream of the Chinese hamster DHFR gene is shown at the left. In this area, six primer pairs were selected within a genomic region located ~15 kb downstream of the gene, and one primer pair was selected within the 3' end of the gene itself. The locations of these primers are indicated by converging arrows. For each primer pair, competitor DNA fragments for quantitative PCR were constructed and quantified as described in Materials and Methods. These competitors contain exactly the same sequence as the target DNA except for a 20-bp insertion in the middle to allow resolution by polyacrylamide gel electrophoresis. Quantitation of the abundance of different markers in a sample of nascent, BrdU-substituted DNA was achieved by coamplification of a fixed amount of sample with increasing concentrations of competitor, as indicated above each gel. Since the amount of competitor is known, the amount of target genomic DNA can be easily derived from the ratio between the two final amplification products, as determined by densitometric scanning. For each gel, the competitor and genomic DNA amplification products are indicated. The results of the quantitations obtained for a sample of BrdU DNA are shown on the right. M, molecular weight marker. Numbering is according to the sequence *cgdhfrori* in GenBank.

The average results of the quantifications obtained by different experiments are shown in Figure 12. At least four independent competitive PCR quantifications were performed for each primer set, by using three different nascent DNA preparations. The results obtained by these analyses revealed a distribution of abundance of genomic markers in the DHFR genomic region, peaking at the segment amplified by primer set 2III. This segment is enriched over 10-fold with respect to those localized at the two boundaries of the analyzed region (primer sets 1II and 3), or within the 3' end of DHFR gene itself (primer set 4II), which is localized approximately 17 kb apart.

Based on the distance between the two DNA segments (amplified by primer sets 8 and 6) flanking the segment amplified by primer set 2III, it can be concluded that a start site for DNA replication is localized within an approximately 800-bp region centered over fragment 2III (Figure 13).

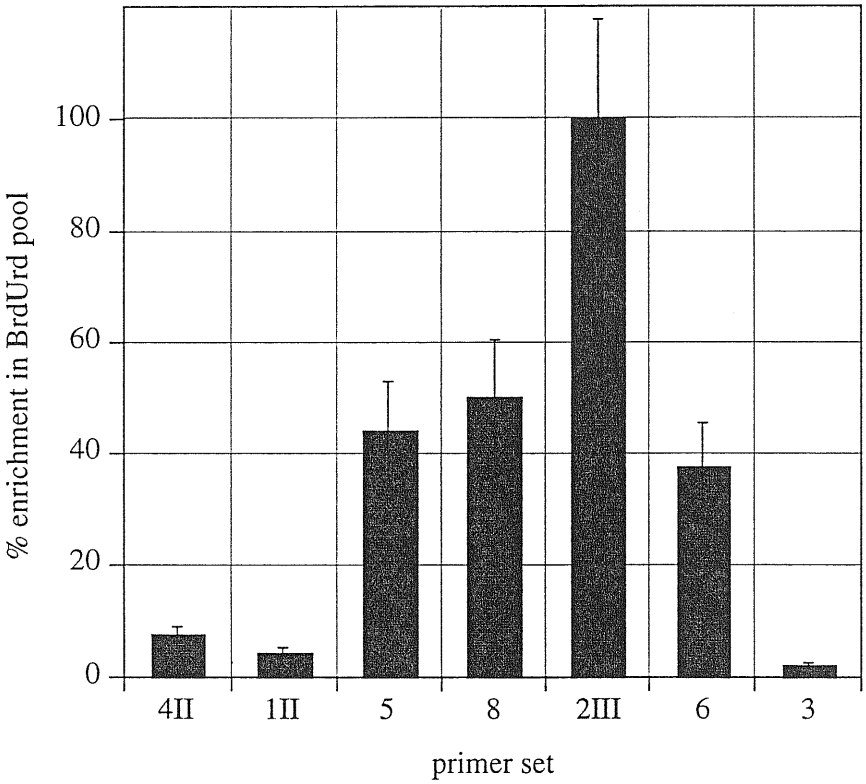


Figure 12. Quantitation of abundance of the analyzed genomic segments of Figure 11 in newly synthesized DNA samples of ~1000 nt. The results are expressed as a percentage of enrichment for each segment with respect to the most abundant one (the segment detected by primer set 2III). The results are the average of at least four independent determinations for each segment, obtained by using three nascent DNA preparations: the bars indicate standard deviations. A clear enrichment for the segment amplified by primer set 2III is evident, thus indicating that a start site for DNA replication is located in an ~800-bp region located between primer sets 8 and 6.

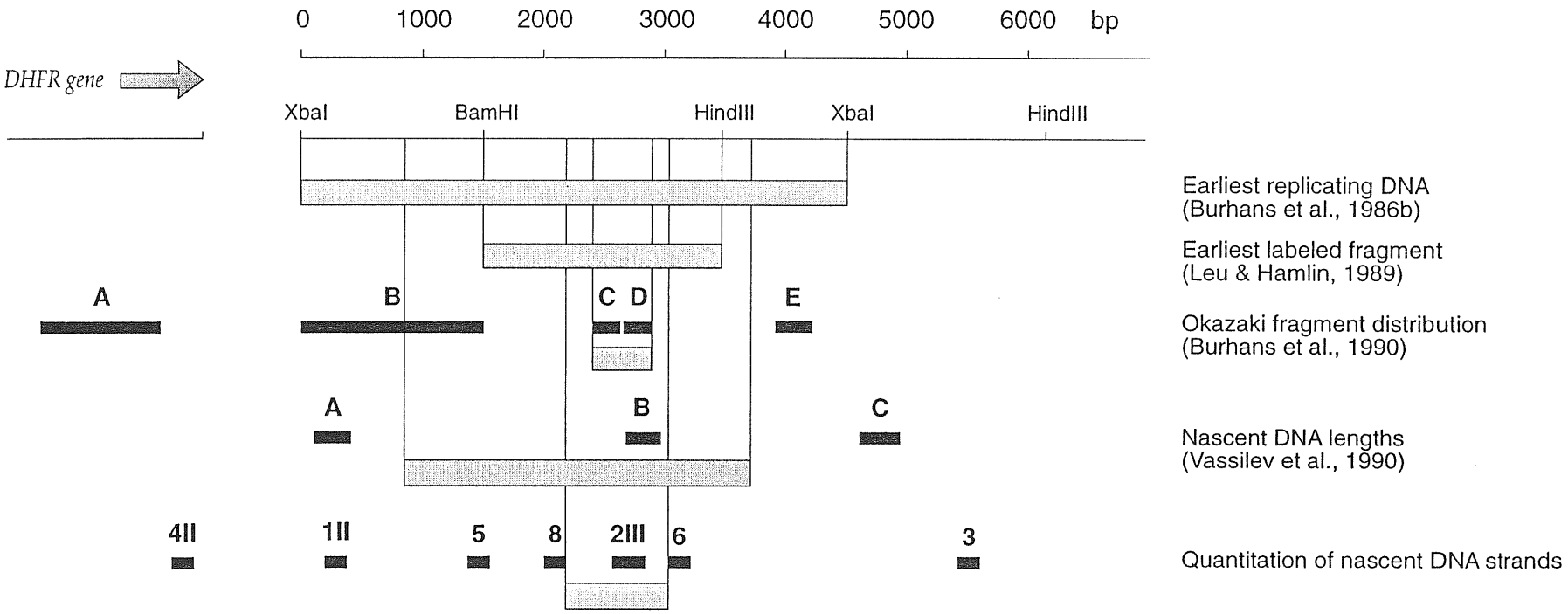


Figure 13. Localization of the start site for DNA replication identified by quantitation of nascent DNA by competitive PCR and by different methods applied to CHOC 400 or CHO K1 cells. A segment of ~6 kb located approximately 15 kb downstream of the DHFR gene is represented as a schematic restriction map. Probes used to determine the location of the origin region in different studies are indicated as solid bars. The origin regions defined by the use of these probes are indicated by gray.

PART II

A library of origin-enriched sequences

The experiments presented in the previous section of the thesis, as well as those obtained by other colleagues at ICGEB by mapping the lamin B2 origin of DNA replication, indicate that the nascent DNA sample prepared is a suitable and convenient source of origin-enriched DNA. This part of the thesis will deal with the utilization of this material for the construction of a library of nascent DNA to be used for mapping purposes.

A nascent DNA sample was prepared from $1x10^8$ exponentially growing, untreated HL-60 cells according to the procedure outlined above; isolation of nascent DNA was obtained by size selection of ~1 kb single strand DNA fragments on 5-30% sucrose gradients. In our experience, this is the critical size to avoid contamination by short Okazaki fragments (scattered along all the genome) and largest DNA stretches derived from random breakage of high molecular weight DNA. Since we recently demonstrated that BrdU labeling and further purification of BrdU-substituted DNA fragments can be omitted from the procedure (Kumar *et al.*, 1996) without changing neither the detectability of DNA fragments nor the relative abundance of the markers - that is the content of the sample -, nascent DNA was prepared without any label. Indeed, the omission of the BrdU-labeling step greatly simplifies the overall protocol and the manipulation of DNA sample (BrdU-substituted DNA is photo-sensitive and thus must be handled in the dark).

Before proceeding to the construction of the library, and to verify that this nascent DNA preparation was indeed enriched in origin sequences, the relative abundances of two DNA segments of the lamin B2 origin region were analyzed by competitive PCR amplification. The existence of an origin of replication in the lamin B2 region was demonstrated by competitive PCR quantification of nascent DNA in a variety of proliferating cells (Kumar *et al.*, 1996). By the analysis of adjacent DNA segments scattered along this region in nascent DNA preparations, the lamin B2 origin region

was identified by the peak of a gaussian distribution of segment abundance, being segment abundance inversely correlated with their distance from the origin. One of the primer set used, B48, detects the most abundant fragment of this region; therefore, it is very likely to coincide with a start site for replication. The other one, B13, detects a fragment, 4.5 kb far from the initiation site, poorly represented in nascent DNA samples. The relative abundance of the two markers in a nascent DNA preparation defines the origin activity, indicating at the same time, the enrichment of the sample in origin-specific DNA. These values are reported in Table 5, column a. Similarly to previously published values, the ratio between the B48 and B13 markers was ~15, thus indicating the quality of the nascent DNA preparation.

Library construction

After removal of the RNA primer, single-stranded nascent DNA was 3' tailed using terminal deoxynucleotidyl transferase (TdT) in the presence of dGTP. Subsequently, second strand synthesis was obtained by priming with an oligonucleotide containg a $(dC)_{14}$ stretch at its 3' end followed by extension using *Taq* DNA polymerase. The nucleotide sequences of the oligos used in this work is reported in Table 4; the overall strategy for library construction is shown in Figure 8. Double stranded, blunt ended DNA molecules obtained in this way were used as a substrate for ligation to a double-stranded linker. This linker is composed of two asymmetric strands, one 25 nt long (LLONG) and the complementary one 11 nt long (LSHORT), in order to allow only oriented ligation. Since it is expected that *Taq* DNA polymerase adds an additional 3' A to a proportion of fully elongated DNA strands (Clark, 1988), the linker used was indeed a mixture of two linker populations, one of which contained an extra T at the 3' end of oligo LLONG, in order to create a T overhang after annealing. The two oligos were used in a 1:1 ratio. The double stranded linker was in vast excess (~1000 fold) with respect to the estimated concentration of the double stranded nascent DNA preparation.

The effiency and quality of all these reactions was checked along the procedure. These controls are shown in Figure 14. An aliquot (1/10) of the dG-tailed material was used in a parallel extension reaction using either a cold second strand synthesis primer and inclusion of labeled dCTP or a labeled primer in the extension reaction. Lanes 1 and 2 in Figure 14, panel A show the results of these controls. Both lanes indicate that the obtained products are represented by a smear of fragments of estimated size in the range between 500 and 2000 bp. These results indicate that both the tailing and the extension procedures generate products with the expected efficiency and length.

Ligation was also monitored in a similar way, by using a hot preparation of oligo LLONG in a primer extension experiment (extension is expected to occur only if ligation had performed successfully). However, in this case no appreciable product smear could be visualized (not shown). The ligated product, however, was then submitted to 15 cycles of PCR amplification with oligo LLONG and oligo Sal/Cla (internal to the 5' tail of the oligo Sal/Cla-dC14 used for second strand synthesis), followed again by primer extension with oligo Sal/Cla, oligo Sal/Cla-dC14 and LLONG. In this case, a strong smear of the expected size could be visualized (Figure 14, panel B, lanes 1-3). Our interpretation of these observations is that the ligation step is highly inefficient in these conditions, yet, it produces a certain amount of ligated molecules which can be easily PCR amplified. The fact that linker ligation is the limiting step in this procedure is not surprising. The laboratory has a long experience in in vivo footprinting using ligation-mediated PCR (Demarchi *et al.*, 1992; Demarchi *et al.*, 1993; Dimitrova *et al.*, 1996), a technique which is based on a similar principle. Also in this case, linker ligation appears to be the limiting step of the procedure.

Evaluation of the amplification products and library cloning

In order to evaluate the quality of the PCR amplified product, as well as to ascertain that the ligation and PCR amplification steps had not introduced a bias in sequence

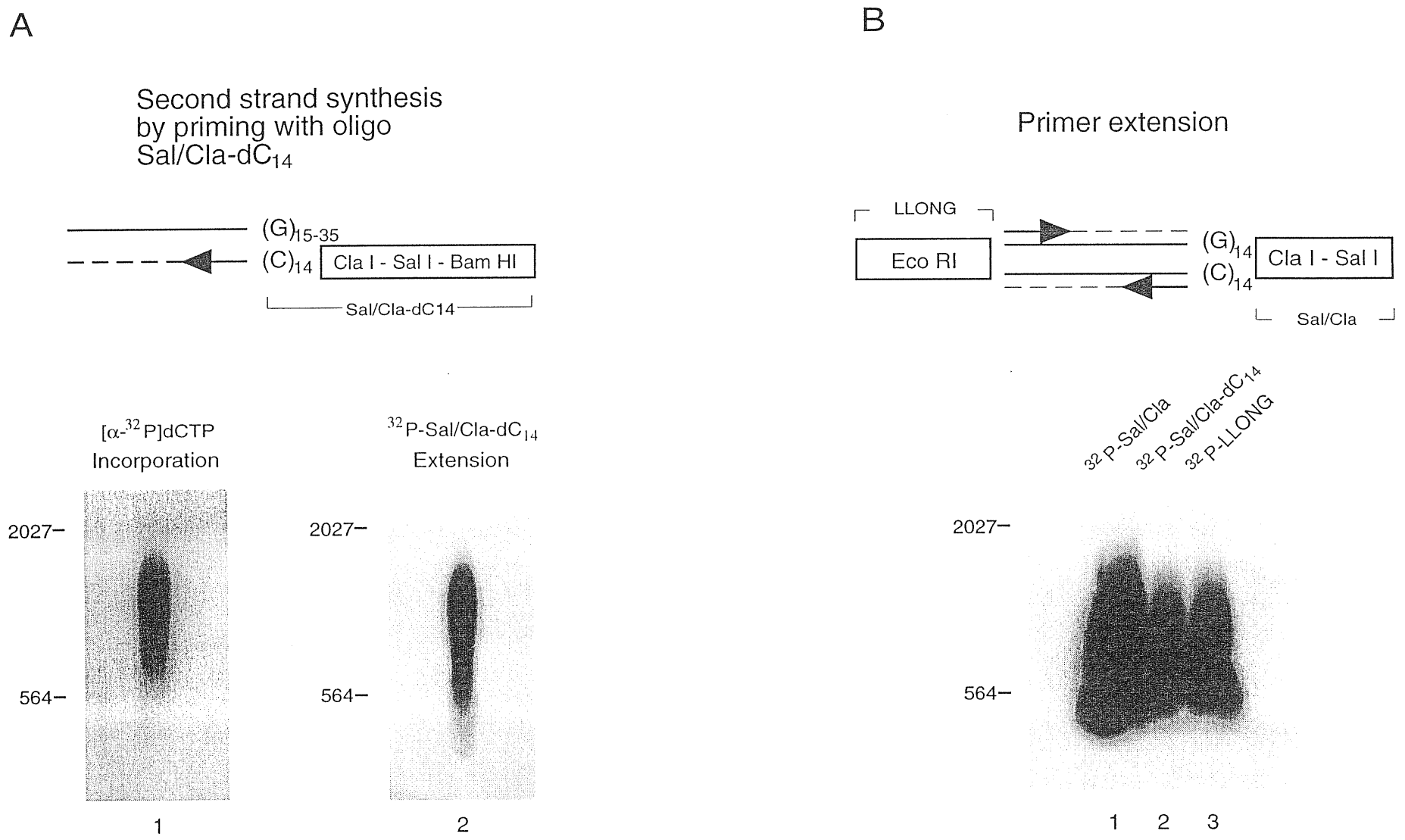


Figure 14. Analysis of nascent DNA preparations. The efficiency and quality of ori-library preparation was checked during the procedure outlined in Figure 8. An aliquot of the dG-tailed material was used in paralel extension reactions using either a second strand synthesis cold primer (oligo Sal/Cla-dC14) and inclusion of labeled dCTP (panel A, lane 1) or the same labeled primer in the extension reaction (panel A, lane 2). Both lanes indicate that the obtained products are represented by a smear of fragments of estimated size in the range between 500 and 2000 bp. Ligation was also checked in a similar way, by using a hot preparation of oligo LLONG in a primer extension experiment on the ori-library amplification product (panel B, lane 3). At the same time, hot oligos Sal/Cla and Sal/Cla-dC14 were used in the same experiment as a further control (panel B, lanes 1 and 2).

selection, a 1/50 aliquot of the PCR product was analyzed by quantification of the B13 and B48 markers. The values in Table 5, b represent the total number of B48 and B13 molecules in the PCR sample. With respect to the values found in the original nascent DNA preparation, a higher B48/B13 ratio was measured in this amplification product, being the B48 fragment 30 folds enriched over B13. Five μl of a 1:1,000 dilution of this first ori-library preparation were further amplified for 30 cycles with oligos LLONG and Sal/Cla. Again, B48 and B13 markers were quantified in a 1:100 dilution of the PCR product. These values are reported in Table 5 lane c (also in this case, the values correspond to the total number of B48 and B13 molecules in the PCR sample). The ratio between the two markers is now ~100 folds.
Therefore, as expected, during PCR some sequences are preferentially amplified. However, it seems that a positive selection of most abundant sequences (that is, most likely, DNA sequences from origins) has occurred against those which are less represented and therefore may be preferentially lost at any step or less efficiently amplified. Whatever the mechanism is, it results in a favourable increase of the signal-to-noise ratio. The total number of molecules, despite the overall loss during the procedure for library construction, was slightly increased after the first amplification, and efficiently amplified in the second round of PCR.
The PCR amplified products (first and second round of amplification) were cleaned from buffer and oligonucleotides by Centricon centrifugation and 1/10 was restricted with Eco RI and Sal I (these sites are present in the two linkers flanking the amplified sequences) and then ligated into the corresponding sites of plasmid pUC18. One/fifth of the ligation products was then transformed into E. coli DH5α by electroporation, using competent cells with an estimated transformation efficiency of $>5x10^8/\mu g$ plasmid DNA. Transformation generated ~50,000 total number of colonies. By taking into account the dilutions of the PCR and ligation products, this results in an estimated number of $\sim 2.5x10^6$ colonies for the total PCR reaction. The number of background colonies due to self-ligation of the vector (most likely due to inefficient digestion with one of the two enzymes) was estimated to be <7%.

	Nascent DNA	Nascent DNA Library First preparation	Nascent DNA Library Second preparation
B48 (molec.)	$2x10^6$	$1.8x10^7$	$3.0x10^7$
B13 (molec.)	$\sim 1.3x10^5$	$6.0\ x10^5$	N.D. (at least $<< 3.2x10^5$)
B48/B13	~15	30	>>100
	a	b	c

Table 5. Analysis of the ori-library. To evaluate the quality of the PCR amplified ori-library, quantitation of B48 and B13 markers in the lamin B2 region was performed by competitive PCR. The nascent DNA preparation used for library construction and two PCR products representing the ori-library were analyzed. In column a is presented the total number of molecules of B48 and B13 in the nascent DNA sample and their relative abundances before PCR amplification. B48 is ~15 folds enriched in respect to B13, as expected in our experience. The values in columns b and c represent the total number of the same markers respectively in the first PCR sample and after a further amplification step. Five µl of a 1:50 dilution of the first PCR product were analyzed (column b). Five µl of a 1:1000 dilution of this first ori-library preparation were further amplified. Again, B48 and B13 markers were quantified in a 1:100 dilution of the PCR product (column c). In this preparation B13 abundance is undetectable even performing 50 cycles of PCR amplification and starting from a 1:5 dilution of the PCR product. It is interesting to note that the ratio between B48 and B13 abundances is increased after library construction and amplification, as if a selection has occured against those sequences that are less represented in the original nascent DNA sample. The localization of B48 and B13 markers is reported in Figure 7.

Hybridization studies

The representativity of the library was also confirmed by hybridization experiments. Aliquots of the original library (library a in Figure 15, corresponding to 15 cycles of amplification of tailed nascent DNA products) and of the re-amplified preparation (library b) were spotted on nylon filters together with scalar amounts of total chromosomal DNA. Filters were probed with three subclones of the lamin B2 locus deriving, respectively, from the origin region (pBN1) and the two flanking fragments far from it (pSB16 and pBgB8). The B48 marker is the most abundant fragment in nascent DNA preparations as detected by competitive PCR, and is included in the pBN1 subclone; pSB16 and pBgB8 are located at the both sides of the origin. The results of these experiments are presented in Figure 15. Each filter contains aliquots of the library (1/20 and 1/200) and scalar amounts of total DNA (from 1 μg to 5 ng). The signals of hybridization of total genomic DNA and of the library preparations to each probe were quantified using an Instant Imager beta-counter machine. The results obtained are presented for each of the three probes in Figure 16, panel A. The hybridization values obtained from the different dilutions of total DNA, when plotted against the actual amount of blotted total DNA, are fitted by a straight line (correlation coefficients, r^2, from 0.90 to 0.99). According to the equations describing these lines, and considering the hybridization values obtained for the 15 cycles library, it was estimated that the dots contained an amount of DNA hybridizing to probe BgB8 corresponding to 44.7 ng of total DNA, an amount of DNA hybridizing to probe BN1 corresponding to 754.8 ng, and an amount of DNA hybridizing to probe SB16 corresponding to 381.2 ng. These results are reported in Figure 16 panel B. According to these results, it can be concluded that fragment BN1 is the most represented in nascent DNA, being 16.9 folds more abundant than the BgB8 fragment, which is ~5 kb far from the origin. On the contrary, fragment SB16, which is only 1.7 kb apart from the initiation site is 1.98 folds less abundant than BN1, as expected from the gaussian distribution of nascent DNA fragment lengths obtained by sucrose gradient isolation.

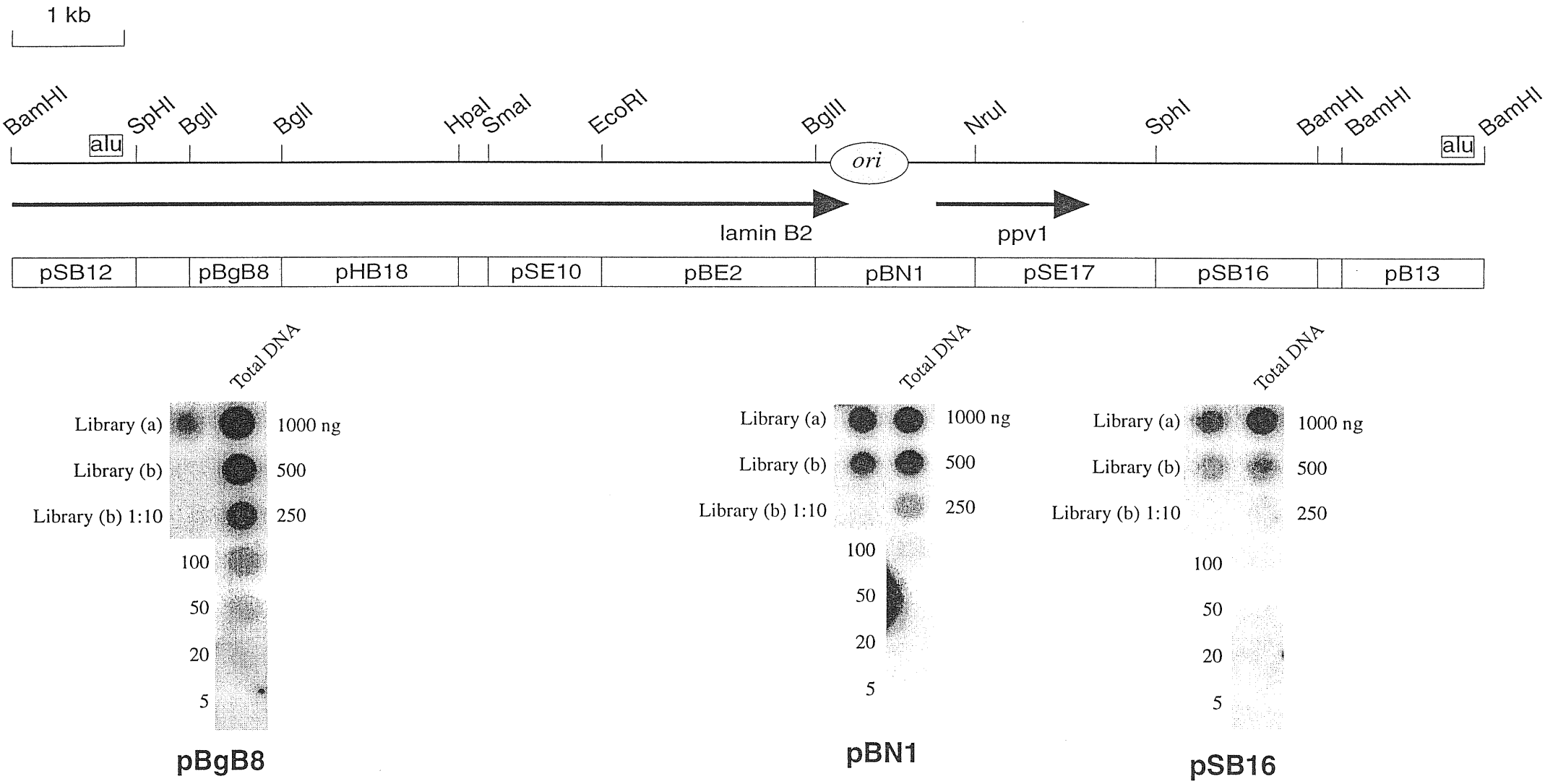


Figure 15. Hybridization of lamin B2 subclones to nascent DNA by dot blot experiments. Aliquots of the original library preparation (library a, corresponding to 15 cycles of amplification of tailed nascent DNA products) and the re-amplified preparation (library b) were spotted on nylon filters together with scalar amounts of total chromosomal DNA (from 1000 to 5 ng). Filters were probed with three subclones of the lamin B2 locus deriving, respectively, from the origin region (pBN1) and two flanking fragments far from it (pBgB8 and pSB16). Results obtained from quantification of the hybridization signals of total genomic DNA and of ori-library preparations are presented in Figure 16.

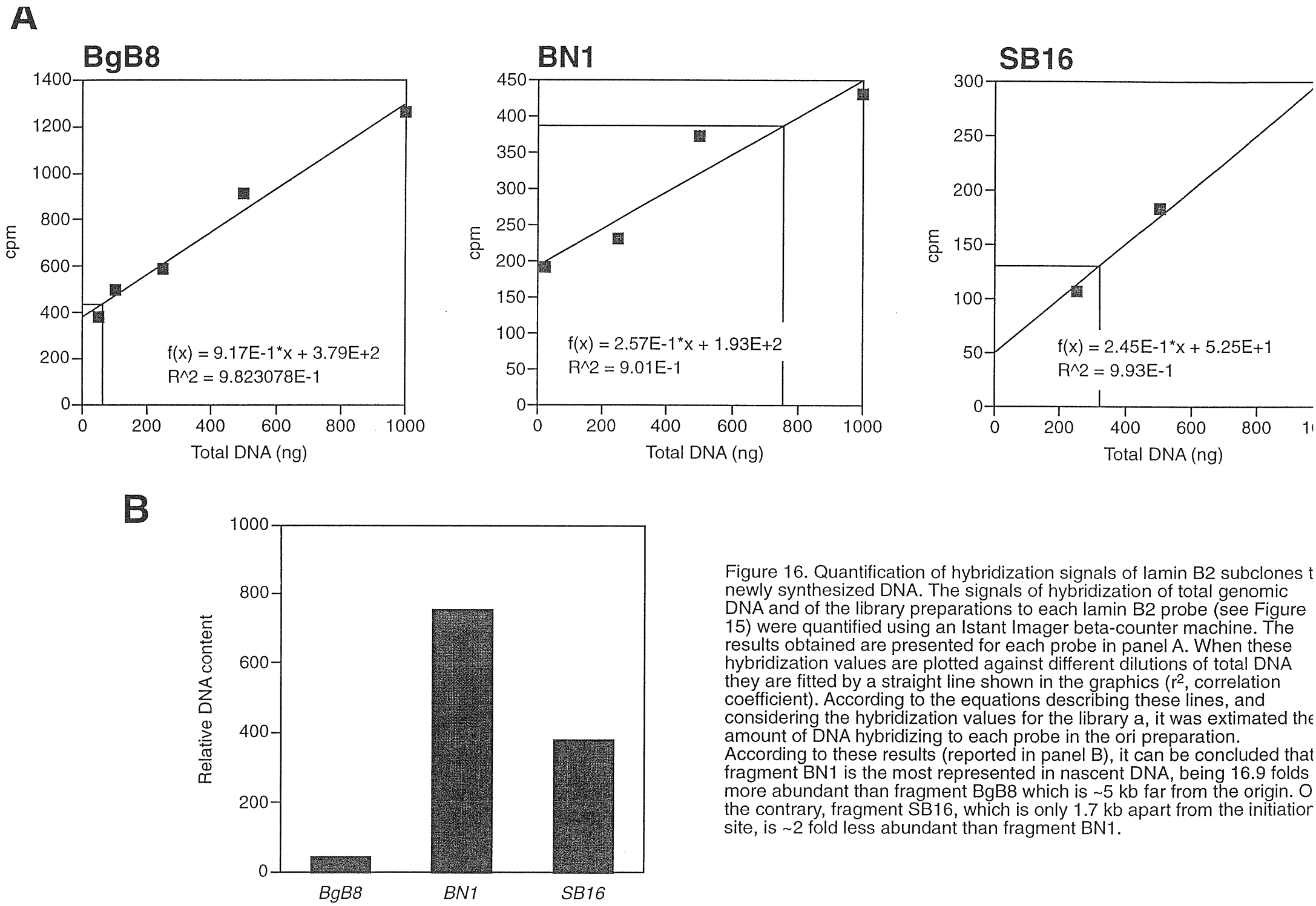


Figure 16. Quantification of hybridization signals of lamin B2 subclones t newly synthesized DNA. The signals of hybridization of total genomic DNA and of the library preparations to each lamin B2 probe (see Figure 15) were quantified using an Istant Imager beta-counter machine. The results obtained are presented for each probe in panel A. When these hybridization values are plotted against different dilutions of total DNA they are fitted by a straight line shown in the graphics (r^2, correlation coefficient). According to the equations describing these lines, and considering the hybridization values for the library a, it was extimated the amount of DNA hybridizing to each probe in the ori preparation. According to these results (reported in panel B), it can be concluded that fragment BN1 is the most represented in nascent DNA, being 16.9 folds more abundant than fragment BgB8 which is ~5 kb far from the origin. O the contrary, fragment SB16, which is only 1.7 kb apart from the initiation site, is ~2 fold less abundant than fragment BN1.

These results were even far more evident using the PCR preparation obtained by two subsequent cycles of amplification (library b in Figure 15). As discusses above, the ratio between B48 and B13 markers in this preparation was even more increased than in the 15 cycles library, and this is confirmed by the increase of the differential hybridization signals between region BN1 and regions SB16 and BgB8 (which is almost undetectable).

In conclusions, these results are in excellent agreement with those obtained by competitive PCR quantification and clearly indicate that the collection of nascent DNA sequences we obtained is indeed enriched in ori sequences.

Preliminary analysis of individual clones

After electroporation in competent DH5α cells, twenty clones were analyzed by PCR as samples of the cloning product. Preliminary results mostly show the existence of independent clones, again indicating that preferential selection of fragments nor by size nor by sequence content has occurred during library construction.

Characterization of individual clones is in progress.

A new method for the identification of origins of DNA replication

One of the most appealing properties of a library of origin-enriched sequences stems from its possible utilization for the identification of new DNA replication origins. For this purpose, the library can be used as a probe for hybridization to immobilized plasmids or cosmids containing fragments of the region of interest (most reasonably, in a contiguous arrangement). The principle of the method is that, being the library enriched in origin sequences, it will preferentially hybridize to immobilized fragments containing an origin of DNA replication. On the contrary, hybridization to non-origin sequences, under-represented in the probe, will poorly detected.

The validity of this new approach for origin identification was tested by hybridization of the library on plasmids containing fragments of the lamin B2 origin region. An example of these experiments is presented in Figure 17. Plasmids pBgB8, pBN1 and

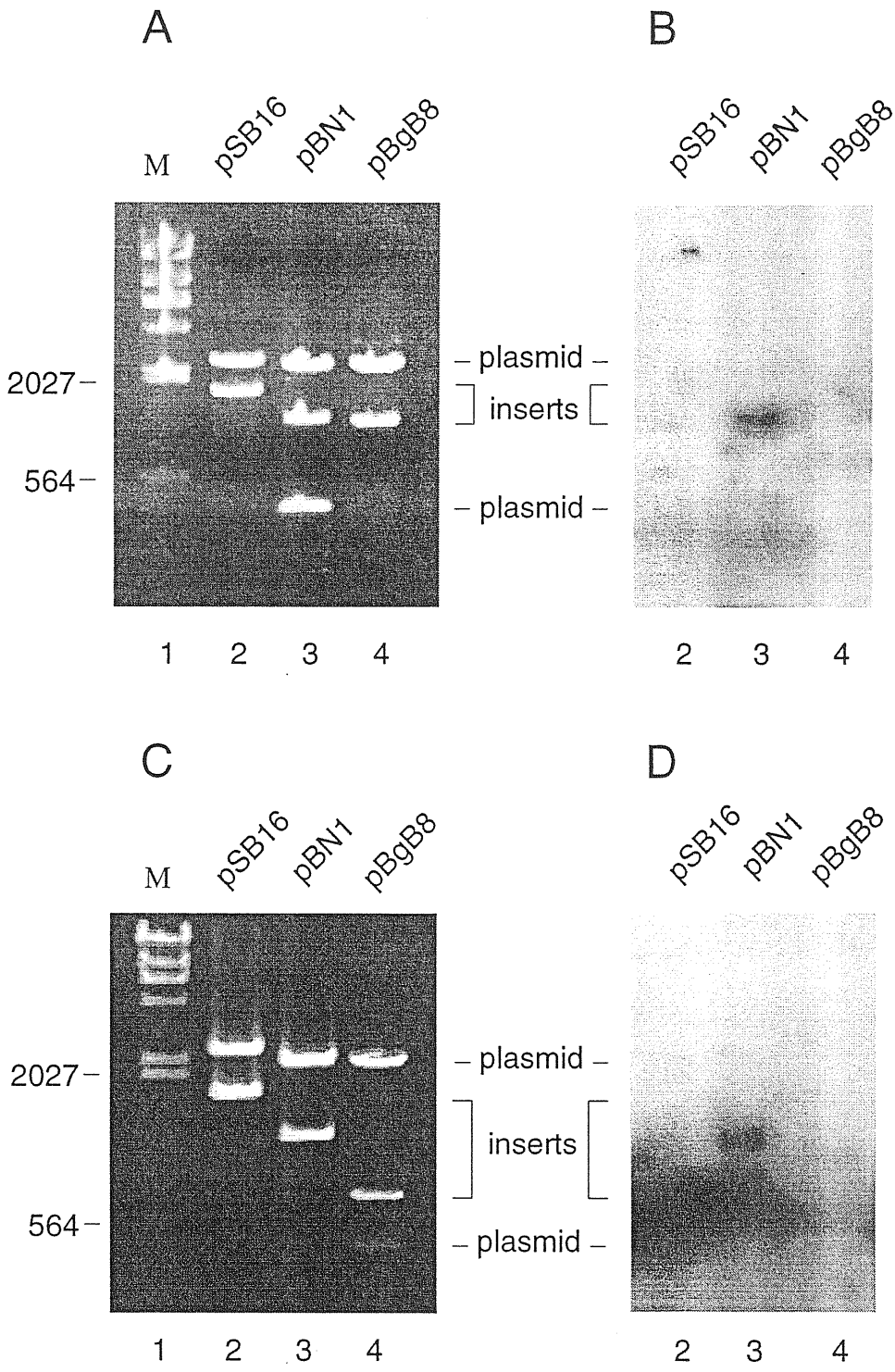


Figure 17. Southern hybridization experiments. Panels A and C are ethidium bromide-stained gels showing plasmid digests: subclone pSB16 was digested with SpHI and BamHI; sublcone pBN1 with PvuII; and subclone pBgB8 either with BglI (panel A) or PvuII (panel B).Vector fragments and inserts from the lamin B2 region are marked. The digests shown in panel A were probed with a ^{32}P-labeled ori-library preparation. The preferential hybridization of subclone pBN1, shown in panel B, likely suggests the presence of a replication initiation site in the pBN1 fragment. This finding was confirmed by using a different ori-library preparation, obtained from PCR amplification of the previous one, as a probe on similar digests (panel C and D). Also in this case, preferential hybridization to pBN1 subclone was found. M: molecular weight marker.

pSB16 (see Figure 15 for their localization in the lamin B2 genomic region) were digested, run on 1% agarose gel and transferred to nylon filters for hybridization to the random primer labeled library product. Panel A lanes 2-4 in Figure 17 show ethidium-bromide stained bands corresponding to the digestion products of plasmids pSB16, pBN1 and pBgB8 respectively. Panel B shows the results of hybridization of the same plasmids with the library probe. As it is evident from the Southern blot, the insert of plasmid pBN1 (lane 3) is recognized strongly by the probe. This is the plasmid encompassing the origin of DNA replication of the lamin B2. A faint band corresponding to the insert of plasmid pSB16 is also visible in lane 2.

A similar experiment was also performed with the origin-enriched library deriving from the second PCR amplification. Also in this case, the insert of plasmid pBN1 was the most detectable one (Figure 17, panels C and D).

DISCUSSION

Mapping DNA replication origins by quantifying the relative abundance of nascent DNA strands using competitive PCR

For my pre-doctoral work presented in this thesis, I contributed to the development, and extensively applied, the competitive PCR technique for the analysis of abundance of nascent DNA strands for origin mapping purposes. As detailed above, this technique is based the isolation of short [~1000 nucleotide (nt)], stretches of newly synthesized DNA followed by the measurement, within this sample, of the abundance of different DNA segments scattered along the genomic region of interest. Central to this procedure is the precise quantification of target DNA abundance. However, this is a challenging task. If it is assumed that about $3x10^4$ different origins are present in the haploid mammalian genome of somatic cells ($3x10^6$ kb), a 1 kb DNA fragment emanating from an origin of replication will be diluted in $3x10^6$ molecules of the same size in a random sample of DNA and in $3x10^4$ molecules in a sample of origin sequences. Hence, only very limited amounts of nascent DNA can be isolated routinely from the bulk of non-replicated DNA of mammalian cells. This is far below the detection limit of the hybridization procedures when using asynchronous cell populations.

These nascent DNA samples, however, are amenable to PCR amplification. A method based on conventional PCR has been developed for this purpose (Vassilev and Johnson, 1990; Vassilev *et al.*, 1990). This method, however, suffers from the difficulty of quantifying conventional PCR, which is sensitive to a number of predictable and unpredictable variables that profoundly affect its efficiency. Among these variables, are differences in the amplification kinetics for each primer set, the quality of the DNA samples and the presence of non specific amplification products. In addition, tube-to-tube variation occurs even under the most finely tuned experimental conditions (Gilliland *et al.*, 1990; Siebert and Larrick, 1992). Due to the

exponential nature of PCR, even a small unpredictable bias in amplification efficiency during the early cycles can lead to large differences in the final product yield. This can be a problem especially when starting with small amounts of target sequences, as is the case with nascent DNA samples. On the contrary, competitive PCR is immune to these problems. Among the advantages of competitive PCR for the quantification of limited amounts of DNA in biological samples are the following: the technique

i) permits the quantification of the absolute number of molecules;

ii) is not affected by the total yield of the PCR reaction;

iii) is independent of the kinetics of primer amplification;

iv) is not affected by the presence of non specific amplification products;

v) does not require the maintenance of the exponential phase of amplification (therefore, amplification products can be directly visualized by ethidium bromide staining of polyacrylamide gels).

Additionally, in our experience competitive PCR quantification generates reproducible results: when the same competitive PCR quantification is repeated several times, the coefficient of variation of the measurements (calculated by dividing the sample standard deviation by the sample mean) is usually <0.3.

Very few mammalian DNA replication origins have been characterized so far, and only a couple of them have been identified with a reasonably high resolution. It is therefore difficult to derive from these scarce data common features among DNA replication origins which may help the current research to identify the factors required for origin specificity and initiation of S phase. It must be said that most of the developed protocols for origin mapping in mammalian cells do not encourage identification of new origins, being very laborious and tricky to perform. Moreover, the results obtained after a usually huge amount of work are either poorly satisfactory because of the low resolution of origin definition or poorly convincing because of the suspicion that they may be biased by the experimental conditions used.

On the contrary, the described procedure for mapping mammalian origins by quantification of nascent DNA abundance using competitive PCR provides a very simple tool for the study of the process of DNA replication initiation in single copy regions of asynchronously growing cells. Most important, it does not require the use of any DNA metabolism-altering procedure, nor laborious experimental protocols. All of these advantages make this procedure virtually applicable to any cultured cell line, virus, plasmid and yeast and to all origin sequences both early and late replicating. The only requirement for its application is the availability of sequence information on the region to analyze, at least for the synthesis of PCR primers. However, since this is a high-resolution mapping method, it is suitable for the investigation of 10-20 kb regions, but inadequate for exploring genomic regions of hundreds of kilobases.

High resolution mapping of the origin of DNA replication in the hamster DHFR gene domain

The dihydrofolate reductase (DHFR) gene domain is the most thoroughly investigated mammalian origin of DNA replication. Replication studies of this region have been initially facilitated by using a cell line in which the DHFR gene is amplified 1000-fold (CHOC 400) and subsequently extended to the analysis of the single copy gene by using procedures that synchronize cells at the G1/S boundary. However, conflicting results have been obtained for this region as far as the localization of the replication origin is concerned. These results range from origin localization in a 450-bp stretch according to the study of the Okazaki fragments polarity switch (Burhans *et al.*, 1990), to origin delocalized in a 50-kb region according to the studies employing the 2-D gel technique (Dijkwel and Hamlin, 1995; Dijkwel *et al.*, 1991; Vaughn *et al.*, 1990). While it is arduous to reconcile these results, it should be pointed out that to date the vast majority of studies analyzing newly synthesized DNA emanating from replication start sites (by analysis either of its distribution along the chromosome or of its polarity) led

to the identification of precise sites for initiation of DNA replication. Among the identified origins are those located within the ribosomal protein S14 gene in Chinese hamster cells (Tasheva and Roufa, 1994), within the Syrian hamster CAD gene (Kelly *et al.*, 1995), in the proximity of the mouse adenosine deaminase gene (Carroll *et al.*, 1993; Virta-Pearlman *et al.*, 1993), of the Chinese hamster rhodopsin gene (Gale *et al.*, 1992), of the human β-globin gene (Kitsberg *et al.*, 1993), of the c-myc gene (Vassilev *et al.*, 1990) and of the human lamin B2 gene (Giacca *et al.*, 1994). In contrast, most of the studies analyzing the structure of replication intermediates by the 2D-gel technique led to the conclusion that origins are dispersed in wide genomic areas or, at best, confined to still large preferred initiation zones (Dijkwel and Hamlin, 1995; Leu and Hamlin, 1989; Little *et al.*, 1993; Vaughn *et al.*, 1990). Therefore, it seems that all these conclusions could be biased in either way by the method employed.

To avoid any possible artifact deriving from the experimental technique, to analyze the behavior of single-copy gene domains and to achieve sufficient resolution to allow subsequent protein-DNA interaction studies, it seemed important to develop an assay that could be performed in physiological conditions and by using the simplest and least artifact-prone procedure. All these demands are satisfied by the mapping procedure by competitive PCR quantification of short stretches of nascent DNA which can be applied to single copy genes, does not require synchronization or other DNA metabolism altering procedures and has a high resolution capacity (Giacca *et al.*, 1994). We have also recently demonstrated that even cell labeling with BrdU and subsequent purification of BrdU-substituted DNA fragments can be omitted from this procedure without changes in the final results (Kumar *et al.*, 1996). This observation definitely rules out the possibility of the introduction of a bias in fragment selection on the anti-BrdU column, or of artifactual DNA fragmentation by photo-damage of BrdU-substituted DNA. However, the finding that BrdU labeling can be omitted is not surprising, since the developed procedure for origin identification in asynchronous

cells relies in principle only on the selection of nascent DNA stretches according to their sizes.

The application of this method to the single copy DHFR locus of CHO K1 cells leads to the conclusion that replication starts at a defined region of ~800 bp (from nt 2206 to nt 3040 in the region sequenced by Caddle et al. (Caddle *et al.*, 1990)) located approximately 17 kb downstream of the gene. Results with the DHFR gene region were in excellent agreement with previously published mapping data in which nascent DNA strands also were labeled (Figure 13), and were subsequently confirmed in another laboratory (T. Kobayashi & M. DePamphilis, unpublished data). It is very unlikely that the distribution of genomic marker abundance found in the analyzed region and presented in Figure 12 could be representative of an initiation zone dispersed between marker 5 and 6 (< 2000 bp), with a preferred initiation site around marker 2III. In fact, the distribution found resembles the one expected by the gaussian distribution of the nascent DNA fragments of ~1000 nt obtained by sucrose gradient purification.

Analysis of the representation of different markers in the DHFR region in purified nascent DNA strands of increasing length was adopted also by Vassilev et al. using conventional PCR (Vassilev *et al.*, 1990). However, this method, although exploiting the extraordinary sensitivity of PCR, is subjected to all the uncontrollable variables affecting the quantitative outcome of conventional PCR (Diviacco *et al.*, 1992; Ferre, 1992; Siebert and Larrick, 1992). On the other hand, quantitative PCR using internal competitive DNA templates (Diviacco *et al.*, 1992; Sestini *et al.*, 1994) has to be considered one of the techniques of choice for precise quantitation of low abundance nucleic acids using PCR (Siebert and Larrick, 1992).

A further important advantage of origin mapping by competitive PCR of nascent DNA is that this technique, contrary to other methods (Handeli *et al.*, 1989), is able to identify origins at relatively high resolution. High resolution mapping is required in order to address the study of the protein-DNA interactions underlying the origin activation event using in vivo footprinting methods (Dimitrova *et al.*, 1996).

In conclusion, the results presented in this work reinforce the notion that DNA replication starts at a defined site downstream of the DHFR gene in Chinese hamster cells. The long-standing discrepancy between this conclusion and the one obtained by the 2D-gel analysis still needs to be understood.

A library of nascent DNA strands

In this thesis, the construction of a library of origin-enriched, nascent DNA strands is described. The purpose of this work was to obtain a tool for the rapid search of novel origins in large genomic areas, an essential pre-requisite for the understanding of the real anatomy of origins in mammalian cells, of their dynamics of regulation, and of their relationship with chromatin structure and replication regulation. By exploiting our experience in nascent DNA preparation and analysis, we obtained a collection of sequences deriving from origin regions. This library of origin-enriched sequences was constructed from a preparation of nascent DNA, using the same kind of protocol which proved successful in the above described origin mapping studies based on competitive PCR. Nascent DNA was prepared by size selection on 5-30% sucrose gradients collecting those fractions containing ~1 kb DNA. This is exactly the same protocol used to prepare nascent DNA for marker quantification by competitive PCR (Giacca *et al.*, 1997; Giacca *et al.*, 1994; Kumar *et al.*, 1996).

Methodological considerations

It is difficult a priori to estimate the expected representativity of a nascent DNA library. An hypothetical estimate can be done on the basis of the following assumptions:

i) replicons have an average size of 10^5 bp; assuming that the haploid human genome has ~$3x10^9$ bp, it derives that there are $3x10^4$ independent origin sequences;

ii) all origins are activated in all cells during each S phase; as a consequence, nascent DNA stretches from all origins have an equal probability of being represented in the library;

iii) nascent DNA is exclusively represented by origin sequences.

Given these assumptions, the library size can be estimated from the general equation:

$$N = \ln (1-p) / \ln (1-1/n)$$

where p is the probability of containing any particular DNA sequence of a sample of N clones, and n is the number of independent origin sequences.

For example, for p=0.99 and n=30,000:

$$N = \ln (1-0.99) / \ln (1-1/30{,}000) = \ln (0.01) / \ln (1-3.33\times10^{-5}) = 1.38\times10^{5}$$

i.e. 1.38×10^{5} clones need to be screened to look for any particular origin.

However, it should be noticed that these theoretical values are likely to be far from those found by our observations on the nature of nascent DNA (see next paragraph).

What is nascent DNA?

The quantitative PCR-based method for DNA replication origin mapping is based on the quantification of the relative differential abundance of different genomic segments of a given area in samples of nascent DNA. These samples are isolated by virtue of i) being shorter than the bulk of total DNA and larger than the expected size of Okazaki fragments - the latter ones derive from all the genome; ii) being synthesized in short time frame prior to isolation, since they contain BrdU-DNA. These fragments are likely to represent short leading strands issuing from the origin. Given this consideration, the expected distribution of abundance of genomic markers in this sample should peak at the origin only and be much decreased at distances far from

the origin itself. However, it is puzzling that in all our experimental observations we do find such a distribution, but the segments far from the origin itself are always represented with a ratio of 1/10-1/20 fold with respect to the segments at the origin. In other terms, segments far from the origin have a background level which do not tend to decrease to undetectability. This suggests to us that the sample of nascent DNA is either contaminated from randomly broken total DNA - which I find unlikely, since DNA is extracted as high molecular weight DNA, and the isolated fragments are ~1 kb in length - or, more likely, it contains low copy number oligomers of ligated Okazaki fragments (3 to 10 copies). These Okazaki fragments are obviously deriving from the lagging strand synthesis of any genomic region. With this respect, it should be observed that very little is known about the kinetics of ligation of Okazaki fragments. Old results in the literature indeed indicate that this is likely to occur at discrete steps, involving formation of progressively larger ligation intermediates (Lonn and Loon, 1983; Lonn and Lonn, 1985).

Additional considerations

In addition to the above reported considerations, it should also be noted that very little - if any - is known about the rate of origin activation in mammalian cells. In other words, it is not clear whether the same origin has to be expected to fire in each cell of a cell population and at each S phase. Consistent data from several laboratories in the last years have indicated that large gene domains retain the same replication timing during S-phase in subsequent cell cycles (Boggs and Chinault, 1994; Holmquist, 1987). This would suggest that a temporal order for origin activity exists and that it is maintained from generation to generation in a single cell type. However, this type of regulation probably requires control at a level higher than the origin - e.g. origin clusters belonging to the same replication focus -. Whether the same origin sequence is actually firing at the same cycle or multiple sequences can be randomly chosen in a single replicon each S-phase still requires further investigations. Obviously, this is an important matter with respect to the issue of library construction,

since, if multiple alternative origin sequences exist, the contribution of each individual origin to the pool of nascent DNA will be diluted. However, it is definitely too early to start addressing this problem, since very few mammalian origins have been so far identified at the nucleotide level.

Preliminary validation of the library of ori-enriched DNA

Before employing the ori-enriched PCR library for the identification of new origins of DNA replication, it was used in a reverse dot blot hybridization experiment, using probes from the lamin B2 genomic region. To this purpose, aliquots of the ori-library were spotted on nylon filters together with scalar amounts of human genomic DNA. Filters were hybridized with three fragments spanning the lamin B2 locus, one very close to the initiation site (BN1) and two distant fragments (BgB8 and SB16). By comparing the hybridization signals obtained for the genomic and nascent DNA for each probe, the amount of each fragment in nascent DNA was calculated. The use of total DNA as a reference ensures that the results were not biased by the different sequence content of the probes. Anyway, reproducible results were obtained using different ori-library preparations. Fragment BN1, which overlaps with the lamin B2 origin, in the nascent DNA library is ~17 folds more abundant than fragment BgB8, being the farthest from the origin.

A similar experiment has already been attempted in the past in our laboratory. Enrichment for lamin B2 origin sequences was measured in nascent, BrdU-substituted DNA prepared from HL-60 cells by comparing the hybridization of probes in this DNA relative to total DNA (Biamonti *et al.*, 1992). However, to obtain sufficient sensitivity in hybridization experiments, methods for the enrichment of nascent DNA were applied. Briefly, HL-60 cells were synchronized at the G1/S border with two blocks of aphidicolin and then released from the block allowing DNA replication to resume in the presence of BrdU. Total DNA was extracted, single stranded nascent

BrdU-DNA was isolated by sucrose gradients and further purified by immunoaffinity chromatography with anti BrdU antibodies. It should be pointed out again that cell synchronization is a means to enrich for newly replicated DNA, but at the same time it results in massive fragmentation of DNA increasing the background level. Similarly, BrdU labeling allows one to further purify nascent BrdU- substituted DNA from parental DNA; however, BrdU-DNA is light sensitive and can be easily damaged. In those experiments, almost 10^8 synchronized cells had to be processed in order to obtain a sufficient amount of newly synthesized DNA to be used in a single hybridization experiment with the probes. The film was developed after several weeks and the hybridization spots were almost at the detectability threshold.
On the contrary, the hybridizations shown in Figure 15 were performed using nascent DNA preparations obtained from asynchronous cell populations. Once tailed, nascent DNA is available in virtually unlimited amount, an can be handled easily. In addition, being the library produced from asynchronous cell populations and easily re-amplifiable by PCR, it can be extensively used to map any origin active in the original cell line (HL-60) without requiring any further purification of nascent DNA.

Potential use of the ori-enriched nascent DNA library: a new method for mapping origins of DNA replication in mammalian cells

Given the above reported considerations, and the poor understanding of the biology of origin activity in mammalian cells, the constructed library appears not to primarily be the source of clones for direct analysis of origin sequences, but to represent a source of probes for origin mapping in large genomic domains. The rationale of the method is to use the library as a probe in hybridization to immobilized plasmids or cosmid contigs corresponding to the region of interest. Since the library is enriched in origin sequences, it will strongly hybridize only to fragments corresponding to origins,

while hybridization to non-origin sequences, poorly represented in the library, will be less detectable.

For this purpose, some preliminary experiments were conducted to assess the feasibility of this approach. The method was applied to the lamin B2 region, where an origin of replication was identified by PCR quantification of nascent DNA (Giacca *et al.*, 1994). In this experiment, three single copy probes from the lamin B2 origin region were used, one corresponding to the initiation site (BN1), and the other two far from it (BgB8 and SB16). The results obtained showed the existence of a precise initiation site coincident with the one already described, since only BN1 fragment (which is the most abundant in nascent DNA as measured by PCR quantification) hybridizes to the nascent DNA library. These results were convincingly reproduced by using different preparations of the ori-library (Figure 17, panel B and D). These findings confirm by an independent method the mapping study on the lamin B2 gene domain which was obtained by competitive PCR. The above commented hybridization experiment also provides further support to the notion that initiation sites are precisely defined.

The use of DNA enriched in origin sequences as a hybridization probe on cloned fragments has been first described by Anachkova and Hamlin and applied to the DHFR amplicon (Anachkova and Hamlin, 1989). However, these experiments required cell synchronization at the G1/S border, enrichment of origin sequences by repeated trioxalen treatments and UV irradiations, and labeling of DNA with BrdU. With the omission of these steps, the amount of origin specific DNA recovered would be extremely limited for use as a hybridization probe. However, despite this protocol for enrichment of ori sequences, the amount of ori-specific DNA obtained was very low and most experiments required 2 to 5 weeks of exposure to X-ray film for adequate signal detection.

On the contrary, the approach we propose provides a very simple and sensitive method for the identification of replication origins. In particular, it does not require any labeling, nor chemical treatment, nor cell synchronization to enrich for origin specific

DNA, since this DNA is no more limiting; in fact, once tailed, nascent DNA can be cloned or simply re-amplified by PCR using specific primers. Nascent DNA is so abundant that, simply by labeling of 1/10 to 1/20 of the PCR product, a hybridization signal can be seen after an over night exposure. Dealing with so large amounts of nascent DNA greatly simplifies and speeds up the experimental work.

In conclusion, the described approach can be applied to any region of interest for which cloned portions are available, potentially providing a very simple new method for origin mapping in large chromosomal regions. The fragment that preferentially hybridizes to the ori-library either in dot blot or in Southern experiments will be the most abundant in nascent DNA; therefore it will very likely be lying the closest to the initiation site.

Experiments aiming at the identification of new origins of DNA replication in mammalian cells are in progress.

REFERENCES

Amaldi, F., Buongiorno-Nardelli, M., Carnevali, F., Leoni, L., Mariotti, D. and Pomponi, M. (1973) Replicon origins in Chinese hamster cell DNA. II. Reproducibility. *Exp. Cell. Res.*, **80**, 79-87.

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

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

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

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

Biamonti, G., Giacca, M., Perini, G., Contreas, G., Zentilin, L., Weighardt, F., Guerra, M., Della Valle, G., Saccone, S., Riva, S. and Falaschi, A. (1992) 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.

Blow, J.J. and Laskey, R.A. (1986) Initiation of DNA replication in nuclei and purified DNA by a cell-free extract of Xenopus eggs. *Cell*, **47**, 577-587.

Boggs, B.A. and Chinault, A.C. (1994) Analysis of replication timing properties of human X chromosomal loci by fluorescence in situ hybridization. *Proc. Natl. Acad. Sci. USA*, **91**, 6083-6087.

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

Brewer, B.J. and Fangman, W.L. (1987) The localization of replication origins on ARS plasmids in *S. cerevisiae*. *Cell*, **51**, 463-471.

Brewer, B.J., Sena, E.P. and Fangman, W.L. (1988) In Kelly, T. and Stillman, B. (eds.), *Analysis of replication intermediates by two-dimensional agarose gel electrophoresis.*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp. 229-234.

Brown, E.H., Iqbal, M.A., Stuart, S., Hatton, K.S., Valinsky, J. and Schildkraut, C.L. (1987) 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.

Burhans, W.C., Selegue, J.E. and Heintz, N.H. (1986a) Isolation of the origin of replication associated with the amplified Chinese hamster dihydrofolate reductase domain. *Proc. Natl. Acad. Sci. USA*, **83**, 7790-7794.

Burhans, W.C., Selegue, J.E. and Heintz, N.H. (1986b) Replication intermediates formed during initiation of DNA synthesis in methotrexate-resistant CHOC 400 cells are enriched for sequences derived from a specific, amplified restriction fragment.*Biochemistry*, **25**, 441-449.

Burhans, W.C., Vassilev, L.T., Caddle, M.S., Heintz, N.H. and DePamphilis, M.L. (1990) Identification of an origin of bidirectional DNA replication in mammalian chromosomes. *Cell*, **62**, 955-965.

Burhans, W.C., Vassilev, L.T., Wu, J., Sogo, J.M., Nallaseth, F.S. and DePamphilis, M.L. (1991) Emetine allows identification of origins of mammalian DNA replication by imbalanced DNA synthesis, not through conservative nucleosome segregation. *EMBO J.*, **10**, 4351-4360.

Caddle, M.S. and Calos, M.P. (1992) Analysis of the autonomous replication behavior in human cells of the dihydrofolate reductase putative chromosomal origin of replication. *Nucleic Acids Res.*, **20**, 5971-5978.

Caddle, M.S., Lussier, R.H. and Heintz, N.H. (1990) Intramolecular DNA triplexes, bent DNA and DNA unwinding elements in the initiation region of an amplified dihydrofolate reductase replicon. *J. Mol. Biol.*, **211**, 19-33.

Calza, R.E., Eckardt, L.A., DelGiudice, T. and Schildkraut, C.L. (1984) Changes in gene position are accompanied by a change in time of replication. *Cell*, **36**, 689-696.

Carroll, S.M., DeRose, M.L., Kolman, J.L., Nonet, G.H., Kelly, R.E. and Wahl, G.M. (1993) 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.

Clark, J.M. (1988) Novel non-templated nucleotide addition reactions catalyzed by prokaryotic and eukaryotic DNA polymerases. *Nucleic Acids Res.*, **16**, 9677-9686.

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

Cook, P.R. (1991) The nucleoskeleton and the topology of replication. *Cell*, **66**, 627-635.

Coverly, D. and Laskey, R.A. (1994) Regulation of eukaryotic DNA replication. *Annu. Rev. Biochem,* **63**, 745-776.

Demarchi, F., D'Agaro, P., Falaschi, A. and Giacca, M. (1992) Probing protein-DNA interactions at the long terminal repeat of human immunodeficiency virus type 1 by in vivo footprinting. *J. Virol.*, **66**, 2514-2518.

Demarchi, F., D'Agaro, P., Falaschi, A. and Giacca, M. (1993) In vivo footprinting analysis of constitutive and inducible protein-DNA interactions at the long terminal repeat of human immunodeficiency virus type 1. *J. Virol.*, **67**, 7450-7460.

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

DePamphilis, M.L. (1996) In DePamphilis, M.L. (ed.) *Origins of DNA replication*, Cold Spring Harbor Laboratory Press, New York, pp. 45-86.

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

Diffley, J.F.X. and Cocker, J.H. (1992) Protein-DNA interactions at a yeast replication origin. *Nature*, **357**, 169-172.

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

Dijkwel, P.A. and Hamlin, J.L. (1995) The Chinese hamster dihydrofolate reductase origin consists of multiple potential nascent-strand start sites. *Mol. Cell. Biol.*, **15**, 3023-3031.

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

Dijkwel, P.A., Vaughn, J.P. and Hamlin, J.L. (1994) Replication initiation sites are distributed widely in the amplified CHO dihydrofolate reductase domain. *Nucleic Acids Res.*, **22**, 4989-4996.

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

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

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

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

Ferre, F. (1992) Quantitative or semi-quantitative PCR: reality versus myth. *PCR Methods and applications*, **2**, 1-9.

Furst, A., Brown, E.H., Braunstein, J.D. and Schildkraut, C.L. (1981) α-Globin sequences are located in a region of early-replicating DNA in murine erythroleukemia cells. *Proc. Natl. Acad. Sci. USA*, **78**, 1023-1027.

Gahn, T.A. and Schildkraut, C.L. (1989) The Epstein-Barr virus origin of plasmid replication, oriP, contains both the initiation and termination sites of DNA replication. *Cell*, **58**, 527-535.

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

Giacca, M., Pelizon, C. and Falaschi, A. (in press) In "Mapping replication origins by quantifying the relative abundance of nascent DNA strands using the competitive

polymerase chain reaction". *Methods: A Companion to Methods in Enzymology*, Academic Press, San Diego.

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

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

Gilliland, G., Perrin, S., Blanchard, K. and Bunn, H.F. (1990) Analysis of cytokine mRNA and DNA: detection and quantitation by competitive polymerase chain reaction. *Proc. Natl. Acad. Sci. USA*, **87**, 2725-2729.

Grassi, G., Zentilin, L., Tafuro, S., Diviacco, S., Ventura, A. and Giacca, M. (1994) A rapid procedure for the quantitation of low abundance mRNAs by competitive RT-PCR. *Nucleic Acids Res.*, **22**, 4547-4549.

Hamlin, J.L. (1992) Mammalian origins of replication. *BioEssays*, **14**, 651-659.

Hamlin, J.L. and Ma, C. (1990) The mammalian dihydrofolate reductase locus. *Biochim. Biophys. Acta*, **1087**, 107-125.

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

Harland, R.M. and Laskey, R.A. (1980) Regulated replication of DNA microinjected into eggs of Xenopus laevis. *Cell*, **21**, 761-771.

Heintz, N.H. and Hamlin, J.L. (1982) An amplified chromosomal sequence that includes the gene for dihydrofolate reductase initiates replication within specific restriction fragments. *Proc. Natl. Acad. Sci. USA*, **79**, 4083-4087.

Heintz, N.H. and Stillman, B.W. (1988) 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.

Heinzel, S.S., Krysan, P.J., Tran, C.T. and Calos, M.P. (1991) Autonomous DNA replication in human cells is affected by the size and the source of the DNA. *Mol. Cell. Biol.*, **11**, 2263-2272.

Holmquist, G.P. (1997) Role of replication time in the control of tissue-specific gene expression. *J. Hum. Genet*, **40,** 151-173.

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

Huberman, J.A., Zhu, J., Davis, L.R. and Newlon, C.S. (1988) Close association of a DNA replication origin and an ARS element on chromosome III of the yeast *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **16**, 6373-6384.

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

Hyrien, O. and Mechali, M. (1992) Plasmid replication in Xenopus eggs and egg extracts: a 2D gel electrophoretic analysis. *Nucleic Acids Res.*, **20**, 1463-1469.

Hyrien, O. and Mechali, M. (1993) Chromosomal replication initiates and terminates at random sequences but at regular intervals in the ribosomal DNA of Xenopus early embryos. *Embo J*, **12**, 4511-20.

Jackson, D.A. (1990) The organization of replication centres in higher eukaryotes. *BioEssays*, **12**, 87-89.

Jacob, F., Brenner, S. and Cuzin, F. (1963) On the regulation of DNA replication in bacteria. *Cold Spring Harbor Symp. Quant. Biol.*, **28**, 329-334.

James, C.D. and Leffak, M. (1986) Polarity of DNA replication through the avian alpha-globin locus. *Mol. Cell. Biol.*, **6**, 976-984.

Kelly, R.E., DeRose, M.L., Draper, B.W. and Wahl, G.M. (1995) Identification of an origin of bidirectional DNA replication in the ubiquitously expressed mammalian CAD gene. *Mol. Cell. Biol.*, **15**, 4136-4148.

Kitsberg, D., Selig, S., Keshet, J. and Cedar, H. (1993) Replication structure of the human β-globin gene domain. *Nature*, **368**, 588-590.

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

Krysan, P.J. and Calos, M.P. (1991) Replication initiates at multiple locations on an autonomously replicating plasmid in human cells. *Mol. Cell. Biol.*, **11**, 1464-1472.

Krysan, P.J., Haase, S.B. and Calos, M.P. (1989) Isolation of human sequences that replicate autonomously in human cells. *Mol. Cell. Biol.*, **9**, 1026-1033.

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

Laskey, R.A., Fairman, M.P. and Blow, J.J. (1989) S phase of the cell cycle. *Science*, **246**, 609-614.

Leffak, M. and James, C.D. (1989) 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.

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

Linskens, M.H. and Huberman, J.A. (1990a) The two faces of higher eukaryotic DNA replication origins. *Cell*, **62**, 845-847.

Linskens, M.H.K. and Huberman, J.A. (1988) Organization of replication of ribosomal DNA in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **8**, 4927-4935.

Linskens, M.H.K. and Huberman, J.A. (1990b) Ambiguities in results obtained with 2D gel replicon mapping techniques. *Nucleic Acids Res.*, **18**, 647-652.

Little, R.D., Platt, T.H.K. and Schildkraut, C.L. (1993) Initiation and termination of DNA replication in human rRNA genes. *Mol. Cell. Biol.*, **13**, 6600-6613.

Lonn, U. and Lonn, S. (1983) Aphidicolin inhibits the synthesis and joining of short DNA fragments but not the union of 10-kilobase DNA replication intermediates. *Proc. Natl. Acad. Sci. USA,* **80,** 3996-3999.

Lonn, U. and Lonn, S. (1985) Accumulation of 10-kilobase DNA replication intermediates in cells treated with 3-aminobenzamide. *Proc. Natl. Acad. Sci. USA,* **82,** 104-108.

Mahbubani, H.M., Paull, J.K.E. and Blow, J.J. (1992) DNA replication initiates at multiple sites on plasmid DNA in Xenous egg extracts. *Nucleic Acids Res.*, **22**, 1457-1462.

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

Mechali, M. and Kearsey, S. (1984) Lack of specific sequence requirement for DNA replication in Xenopus eggs compared with high sequence specificity in yeast. *Cell*, **38**, 55-64.

Milbrandt, J.D., Heintz, N.H., White, W.C., Rothman, S.M. and Hamlin, J.L. (1981) Methotrexate-resistant Chinese hamster ovary cells have amplified a 135- kilobase-pair region that includes the dihydrofolate reductase gene. *Proc. Natl. Acad. Sci. USA*, **78**, 6043-6047.

Nasmyth, K. (1996) In DePamphilis, M.L. (ed.) *Origins of DNA replication*, Cold Spring Harbor Laboratory Press, New York, pp. 331-386.

Newlon, C.S. (1988) Yeast chromosome replication and segregation. *Microbiol. Rev.*, **52**, 568-601.

Newlon, C.S. (1996) In DePamphilis, M.L. (ed.) *Origins of DNA replication*, Cold Spring Harbor Laboratory Press, New York, pp. 873-914.

Orr-Weaver, T.L. (1991) Drosophila corion genes: cracking the eggshell's secrets. *BioEssays*, **13**, 97-105.

Romanowski, P. and Madine, M.A. (1997) Mechanisms restricting DNA replication to once per cell cycle: the role of Cdc6p and ORC. *Trends in Cell Biol*, **7**, 9-10.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual/II Edition*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

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

Sestini, R., Orlando, C., Zentilin, L., Gelmini, S., Pinzani, P., Giacca, M. and Pazzagli, M. (1994) Measuring c-*erb*B-2 oncogene amplification in fresh and paraffin-embedded tumors by competitive polymerase chain reaction. *Clin. Chem.*, **40**, 630-636.

Siebert, P.D. and Larrick, J.W. (1992) Competitive PCR. *Nature*, **359**, 557-558.

Spack, E.G., Lewis, E.D., Paradowski, B., Schimke, R.T. and Jones, P.P. (1992) Temporal order of DNA replication in the H-2 major histocompatibility complex of the mouse. *Mol. Cell. Biol.*, **12**, 5174-5188.

Stinchcomb, D.T., Struhl, K. and Davis, R.W. (1979) Isolation and characterization of a yeast chromosomal replicator. *Nature*, **282**, 39-43.

Tasheva, E.S. and Roufa, D.J. (1994) A mammalian origin of bidirectional DNA replication within the Chinese hamster RPS14 locus. *Mol. Cell. Biol.*, **14**, 5628-5635.

Trempe, J.P., Lindstrom, Y.I. and Leffak, M. (1988) Opposite replication polarities of transcribed and nontranscribed histone H5 genes. *Mol. Cell. Biol.*, **8**, 1657-1663.

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

Vassilev, L.T., Burhans, W.C. and DePamphilis, M.L. (1990) Mapping an origin of DNA replication at a single-copy locus in exponentially proliferating mammalian cells. *Mol. Cell. Biol.*, **10**, 4685-4689.

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

Vaughn, J.P., Dijkwel, P.A. and Hamlin, J.L. (1990) Replication initiates in a broad zone in the amplified CHO dihydrofolate reductase domain. *Cell*, **61**, 1075-1087.

Virta-Pearlman, V.J., Gunaratne, P.H. and Chinault, A.C. (1993) Analysis of a replication initiation sequence from the adenosine deaminase region of the mouse genome. *Mol. Cell. Biol.*, **13**, 5931-5942.

Waga, S. and Stillman, B. (1994) Anatomy of a DNA replication fork revealed by reconstitution of SV40 DNA replication *in vitro*. *Nature*, **369**, 207-212.

Yang, L. and Botchan, M. (1990) Replication of bovine papilloma virus type 1 DNA initiates within an E2 responsive enhancer element. *J. Virol.*, **64**, 5903-5911.

Zhu, J.G., Newlon, C.S. and Huberman, J.A. (1992) Localization of a DNA replication origin and termination zone on chromosome III of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **12**, 4733-4741.

APPENDIX

INTRODUCTION

Most of the interest in the field of DNA replication has recently been focused on the study of proteins involved in initiation of DNA replication. A search for factors binding to origins in yeast (ARSs, autonomously replicating sequences) led to the identification of the origin recognition complex (ORC). The ORC is a six subunit (Orc1-Orc6) initiation complex that specifically binds to origins and is essential for initiation of DNA replication in yeast (Bell and Stillman, 1992; Diffley and Cocker, 1992). Further studies have shown that homologs of ORC proteins exist also in higher eukaryotes. Drosophila Orc2 and Orc5 have been recently identified (Ehrenhofer-Murray *et al.*, 1995; Gossen *et al.*, 1995); Xenopus Orc1 and Orc2 have been shown to perform essential functions in DNA replication (Carpenter *et al.*, 1996; Romanowski *et al.*, 1996b; Rowles *et al.*, 1996); however, ORC-like complexes have not yet been isolated from human cell extracts and no function has been yet described for human ORC homologs (hOrc1 and hOrc2) so far identified (Gavin *et al.*, 1995).

In addition to ORC, the Cdc6 protein acts at a critical early step in replication, perhaps by interacting directly with the ORC and determining the frequency of origins firing (Cocker *et al.*, 1996; Coleman *et al.*, 1996; Liang *et al.*, 1995). Similarly, the Mcm (minichromosome maintenance) proteins are required for the initiation of DNA replication (for a review: Chong *et al.*, 1996; Kearsey *et al.*, 1995). Homologs of both Cdc6 and Mcm proteins have been identified in higher eukaryotes (Kubota *et al.*, 1997; Madine *et al.*, 1995; Romanowski *et al.*, 1996a; Sanders Williams *et al.*, 1997; Thommes *et al.*, 1997). These studies have suggested a model in which the

formation and activation of a pre-replicative complex (containing ORC, Cdc6, Mcm and probably other cell cycle proteins) at origins determine their proper utilization in S-phase (Coleman *et al.*, 1996; Rowles *et al.*, 1996; Stillman, 1996).

A project aiming at the identification of proteins binding to and activating origins in mammalian cells is ongoing in our laboratory. In the past few years, using an in vivo footprinting technique we identified specific protein-DNA interactions at the lamin B2 origin in human cells (Dimitrova *et al.*, 1996). Most interestingly, we detected an extended footprint of 70 bp in close correspondence to the replication initiation site which is correlated to replication, since in non proliferating cells this footprint was absent.

While we are trying to purify the complex responsible of the footprint in this region, we are investigating whether human ORC and Mcm homologs could bind to the same region. To this purpose, I have constructed an epitope tagged version of human Orc1 (Gavin *et al.*, 1995) and Mcm 3 (Thommes *et al.*, 1992) homologue proteins for in vivo cross-linking experiments to understand ORC and Mcm relationship to replication sites.

MATERIALS AND METHODS

Construction of HA-tagged proteins

The HA epitope (YPYDVPDYA) was added at the N-terminus of human Orc1 and Mcm3/P1 proteins by using PCR primers containing the tag sequence (Table 1-a). The upstream primer of each set includes (from 5' end to 3') a restriction enzyme site, the ATG start codon, the epitope tag sequence, and the DNA sequence for the amino-terminus of the target protein. The downstream primer includes (from 5' end to 3') a restriction enzyme site different from the one in the upstream primer, a stop codon, and the DNA sequence for the carboxy-terminus of the target protein. Amplification of the protein coding sequence by means of such a set of primers resulted in insertion of the HA epitope in frame with the coding sequence. To this purpose, *Pfu* DNA Polymerase (Stratagene) was used according to the instruction of the manufacturer. Primers were constructed according to human Orc1 sequence U40152 and human Mcm3/P1 sequence X62153 in GenBank.

RNA isolation and reverse transcription

Total RNA was isolated from human HeLa cells by guanidinium thiocyanate method (Chomoczynski and Sacchi, 1987). Reverse transcription was performed with 1 μg of RNA by using M-MLV Reverse Transcriptase (Promega) and oligo-dT primer (Gibco BRL) in a 25 μl reaction. The volume was then adjusted to 100 μl and 10 μl were used for subsequent PCR amplification.

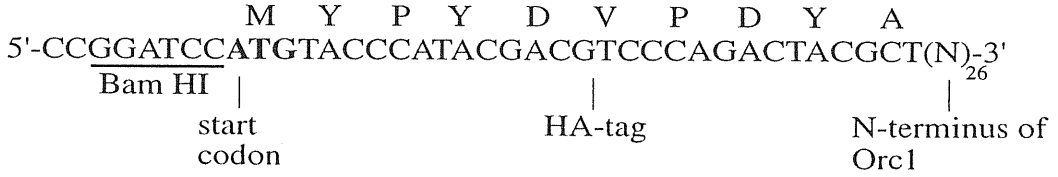
Primers for Orc1 cDNA amplification

Orc1Ntag	5'-CCGGATCCATGTACCCATACGACGTCCCAGACTACGCTATGGCACACTACCCCACAAGGCTGAA-3'
Orc1C	5'-CGCTCTAGACTATTACTCGTCTTTCAGCGCATACAGC-3'

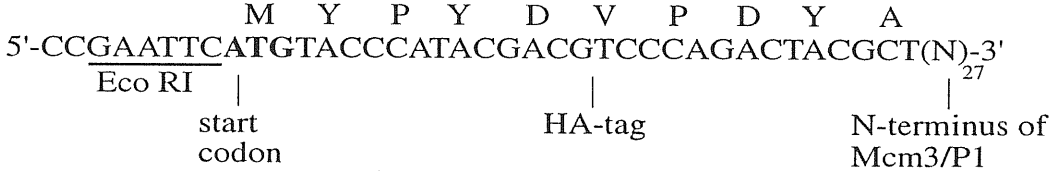
Primers for Mcm3/P1 cDNA amplification

P1Ntag	5'-CCGAATTCATGTACCCATACGACGTCCCAGACTACGCTATGGCGGGTACCGTGGTGCTGGACGAT-3'
P1C	5'-CGCTCTAGACTATCAGATGAGGAAGATGATGCC-3'

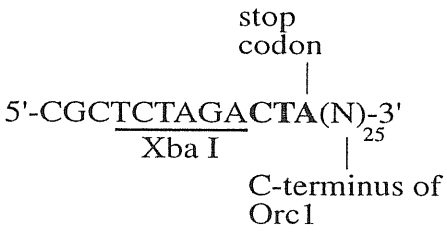
Structure of the upstream primer for tagging the amino-terminal end of Orc1 cDNA with the HA-epitope



Structure of the upstream primer for tagging the amino-terminal end of Mcm3/P1 cDNA with the HA-epitope



Structure of the downstream primer for Orc1 amplification



Structure of the downstream primer for Mcm3/P1 amplification

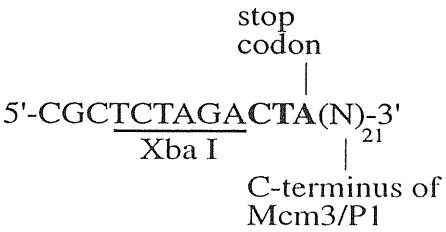


Table 1-a. Sequences and structures of primers for construction of HA-tagged human Orc1 and Mcm3/P1 proteins.

Construction of vectors that express HA-Orc1 and HA-Mcm3/P1

The PCR products corresponding to human HA-Orc1 and HA-Mcm3/P1 were purified and cloned in a pBS vector using Bam HI/ Xba I and Eco RI/ Xba I cloning sites, respectively. The cloned products were partially sequenced by the dideoxy-chain termination method (T7 Sequencing™ Kit-Pharmacia) and then cloned in a mammalian expression plasmid (pCDNA III vector, Invitrogen) using the same restriction sites. This vector confers neomycin (G418) resistance to transfected cells. The correct expression of the two tagged-proteins was verified by in vitro transcription and translation assay (TNT, Promega, Figure 1-a, panel A).

Transfection of plasmid DNA into mammalian cells

Plasmid DNAs were introduced into 293 mammalian cells by the calcium phosphate precipitation method as described (Wigler *et al.*, 1977). Cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum, 2 mM L-glutamine and 50 mg gentamicin per ml. HA-tagged proteins expression was verified 48 hrs after transfection by western blotting analysis (Figure 1-a, panel C).

Selection of stable clones expressing human HA-Orc1 and HA-Mcm3/P1 proteins

Transfected 293 cells were selected for neomycin (G418) resistance (1 mg/ml in complete medium, Gibco BRL) and single resistant clones were grown under selective conditions. The presence of the construct was analyzed in transfected cells by PCR using a primer set amplifying sequences from the neomycin gene. Then, three clones transfected with the HA-Mcm3/P1 pCDNA III construct were analyzed by

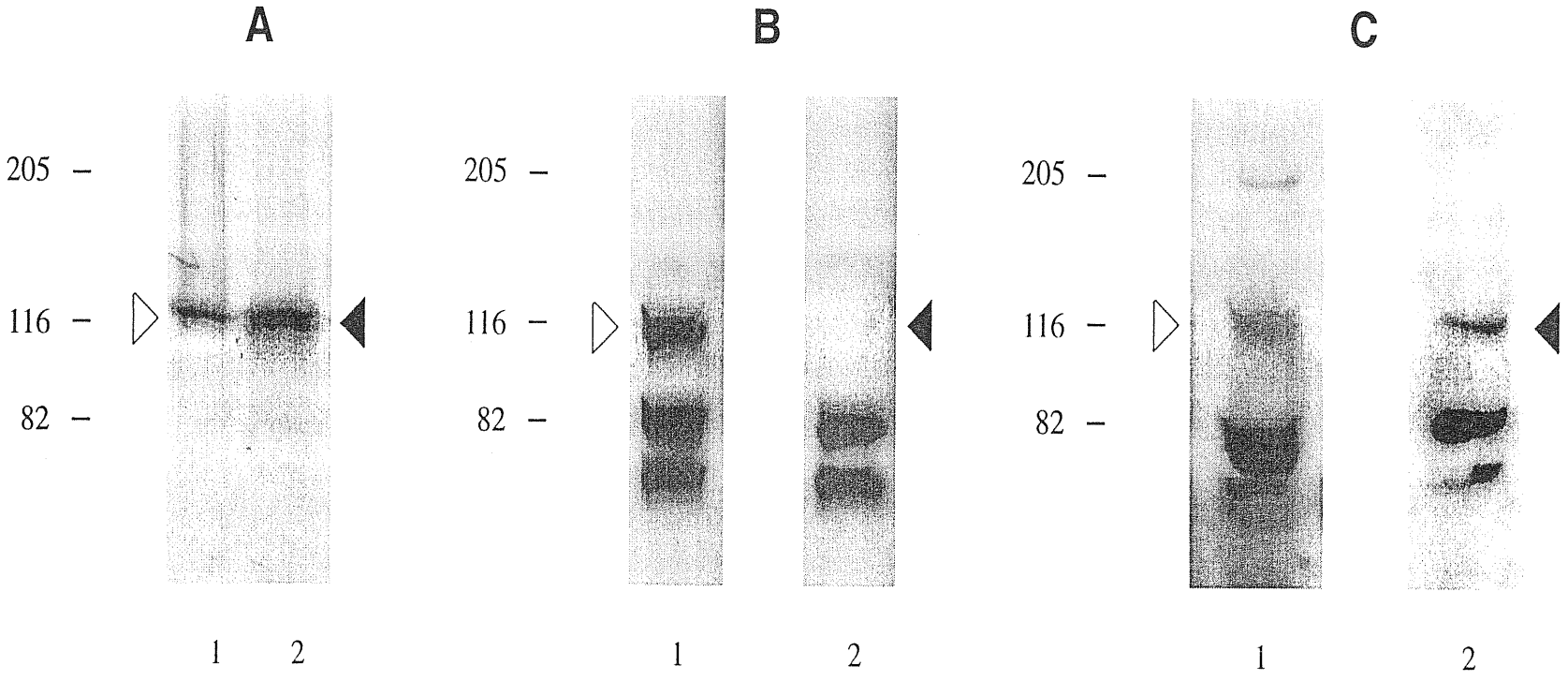


Figure 1-a. Analysis of HA-tagged human Mcm3/P1 and Orc1 proteins. Panel A: TNT reactions of Mcm3/ P1 and Orc1 performed in the presence of ^{35}S-methionine. Panels B and C: protein expression either in stable clones (panel B), or after transient transfection (48 hrs after transfection, panel C) analyzed by western blotting using a monoclonal anti-HA antibody. Lane 1 and 2 in each panel respectively represent Mcm3/P1 and Orc1 proteins. Molecular weight markers are presented at the left side of each panel. The empty arrowheads indicate the HA-Mcm3/P1 protein; the filled arrowheads indicate HA-Orc1 protein.

western blotting. All of them expressed the tagged-protein, although at different levels. Ten clones transfected with the HA-Orc1 expression plasmid were analyzed, but no one of them expressed the protein. Transfection with HA-Orc1 plasmid DNA was repeated and the bulk of cells after selection with G418 was analyzed. Unfortunately, HA-Orc1 positive cells were not found (Figure 1-a, panel B).

Preparation of cell lysates and western blot

Cells were washed twice with ice-cold phosphate-buffered saline (PBS) and lysed directly in SDS-loading buffer (whole cell lysates). To prepare nuclear lysates, cells were additionally washed once with ice-cold buffer A (10 mM Hepes pH 7.6, 15 mM KCl, 2 mM $MgCl_2$, 0.1 mM EDTA), nuclei were extracted with 0.2% NP-40 (Calbiochem, La Jolla CA) and lysed in SDS-buffer. Nuclear lysates from ~$2x10^6$ cells were separated on a 7.5% SDS-polyacrylamide gel and were transferred to a nitrocellulose filter (Amersham) by electroblotting. Filters were incubated in 5% milk-TBS buffer (5% w/v non-fat dried milk in 150 mM NaCl, 10 mM Tris pH 7.4) for 2 hr at room temperature, washed with TBS and then probed with mouse monoclonal antibody anti-HA (clone 12CA5, Boehringer Mannheim) at a concentration of 0.75 µg/ml in 5% milk-TBS for 1 hr at room temperature. After washing in TBS, filters were incubated in 5% milk-TBS with horseradish peroxidase-conjugated goat anti-mouse IgG (Amersham) for 1 hr at room temperature. Finally, filters were washed with 5% milk-TBS, TBS and TBS-0.1% Tween 20 and the immunoreactive proteins were detected by chemiluminescence (ECL, Amersham, UK).

RESULTS and DISCUSSION

With the aim to study the behavior of replication initiation proteins Orc1 and Mcm3/P1 in human cells, we constructed a tagged version of the same. By using primers containing the HA-tag sequence, protein coding sequences of human Orc1 and Mcm3/P1 were amplified resulting in the insertion of the HA tag at the N-terminus of both of them. Amplifications were carried out using appropriate primer sequences designed on the sequences of these proteins available in GenBank, and carrying suitable 5' ends for cloning into the BlueScript vector. The sequences of these primers are shown in Table I-a. These constructs were partially sequenced and were cloned in a mammalian expression vector under the control of the CMV promoter (vector pCDNA III).

The correct expression of the HA-tagged proteins was analyzed by in vitro transcription and translation experiments; this approach also offers a way to produce small amount of the tagged proteins. The TNT products are shown in Figure 1-a, panel A.

Plasmids containing the cDNAs for Orc1 and Mcm3/P1 in the eukaryotic expression vector were transfected into human kidney 293 cells. Cells were grown under neomycin selection and stable clones were isolated. However, we could obtain only stable clones producing HA-Mcm3/P1, while HA-Orc1 protein was present 48hr after transfection, but could not be found after a longer selection (Figure 1-a, panel B for stable cell clones and panel C for transient transfection). Three different clones have been isolated and analyzed for HA-Mcm3/P1 and 10 different clones for HA-Orc1. All clones are behaving in the same manner as far as the transfected proteins are concerned, as revealed by western blotting.

We believe that, being Orc1 mRNA level (and possibly also protein level) tightly regulated as a function of cell proliferation (Ohtani *et al.*, 1996) and being Orc1 most

likely important for DNA replication initiation, cells do not tolerate an unbalanced Orc1 expression; as a consequence, only cells in which HA-Orc1 expression is switched off can survive selection. According to this interpretation, it has also been reported that human Orc1 overexpression in *S. cerevisiae* leads to continuous DNA synthesis in the absence of mitosis (Wolf *et al.*, 1996). A similar perturbation of the re-replication control by overexpression of Orc1 may not be allowed in mammalian cells.

In conclusion, human Orc1 and Mcm3/P1 cDNAs have been cloned and HA-tagged proteins have been constructed. After transfection in 293 cells, the HA-proteins can be identified by using commercial antibodies against the HA tag. Either for transient transfection experiments (for HA-Orc1 and HA-Mcm3/P1) or by using stably transfected cell clones (for HA-Mcm3/P1), we believe that these reagents will constitute valid tools for the understanding of the relationship between the ORC and Mcm complexes and the dynamics of protein-DNA interactions at origins of DNA replications in human cells.

REFERENCES

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

Carpenter, P.B., Mueller, P.R. and Dunphy, W.G. (1996) Role for a Xenopus Orc2-related protein in controlling DNA replication. *Nature*, **379**, 357-360.

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

Chong, J.P.J., Thommes, P. and Blow, J.J. (1996) The role of MCM/P1 proteins in the licencing of DNA replication. *Trends in Biol. Sc.*, **21**, 102-106.

Cocker, J.H., Piatti, S., Santocanale, C., Nasmyth, K. and Diffley, J.F.X. (1996) An essential role for the Cdc6 protein in forming the pre-replicative complexes of budding yeast. *Nature*, **379**, 180-182.

Coleman, T.R., Carpenter, P.B. and Dunphy, W.G. (1996) The Xenopus Cdc6 protein is essential for the initiation of a singe round of DNA replication in cell-free extracts. *Cell*, **87**, 53-63.

Diffley, J.F.X. and Cocker, J.H. (1992) Protein-DNA interactions at a yeast replication origin. *Nature*, **357**, 169-172.

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

Ehrenhofer-Murray, A.E., Gossen, M., Pak, D.T.S., Botchan, M.R. and Rine, J. (1995) Separation of origin recognition complex functions by cross-species complementation. *Science*, **270**, 1671-1674.

Gavin, K.A., Hidaka, M. and Stillman, B. (1995) Conserved initiator proteins in eukaryotes. *Science*, **270**, 1667-1671.

Gossen, M., Pak, D.T.S., Hansen, S.K., Acharya, J.K. and Botchan, M.R. (1995) A Drosophila homolg of the yeast origin recognition complex. *Science*, **270**, 1674-1677.

Kearsey, S.E., Maiorano, D., Holmes, E.C. and Todorov, I.T. (1995) The role of MCM proteins in the cell cycle control of genome duplication. *BioEssays*, **18**, 183-190.

Kubota, Y., Mimura, S., Nishimoto, S., Masuda, T., Norjima, H. and Takisawa, H. (1997) Licensing of DNA replication by a multi-protein complex of PCP/P1 proteins in Xenopus eggs. *EMBO J.*, **16**, 3320-3331.

Liang, C., Weinreich, M. and Stillman, B. (1995) ORC and Cdc6 interact and determine the frequency of initiation of DNA replication in the genome. *Cell*, **81**, 667-676.

Madine, M.A., Khoo, C.Y., Mills, A.D. and Laskey, R.A. (1995) MCM3 complex required fr cell cycle regulation of DNA replication in vertebrate cells. *Nature*, **375**, 421-424.

Ohtani, K., DeGregori, J., Leone, G., Herendeen, D.R., Kelly, T.J. and Nevins, J.R. (1996) Expression of the HsOrc1 gene, a human ORC1 homolog, is regulated by cell proliferation via the E2F transcription factor. *Mol. Cell. Biol.*, **16**, 6977-6984.

Romanowski, P., Madine, M.A. and Laskey, R.A. (1996a) XMCM7, a novel member of the Xenopus MCM family, interacts with XMCM3 and colocalizes with it throughout replication. *Proc. Natl. Acad. Sci. USA*, **93**, 10189-10194.

Romanowski, P., Madine, M.A., Rowles, A., Blow, J.J. and Laskey, R.A. (1996b) The Xenopus origin recognition complex is essential for DNA replication and Mcm binding to chromatin. *Curr. Biol.*, **6**, 1416-1425.

Rowles, A., Chong, J.P.J., Brown, L., Howell, M., Evan, G.I. and Blow, J.J. (1996) Interction between the origin recognition complex and the replication licencing system in Xenopus. *Cell*, **87**, 287-296.

Sanders Williams, R., Shohet, R.V. and Stillman, B. (1997) A human protein related to yeast Cdc6p. *Proc. Natl. Acad. Sci. USA*, **94**, 142-147.

Stillman, B. (1996) Cell cycle control of DNA replication. *Science*, **274**, 1659-1664.

Thommes, P., Fett, R., Schray, B., Burkhart, R., Barnes, M., Kennedy, C., Brown, N.C. and Knippers, R. (1992) Properties of the nuclear P1-Protein, a mammalian homologue of the yeast mcm3 replication protein. *Nucleic Acids Res.*, **20**, 1069-1074.

Thommes, P., Kubota, Y., Takisawa, H. and Blow, J.J. (1997) The RLF-M component of the replication licencing system forms complexes containing all six MCM/P1 polypeptides. *EMBO J.*, **16**, 3312-3319.

Wigler, M., Silverstein, S., Lee, L.S., Pellicer, A., Cheng, Y.C. and Axel, R. (1977) Transfer of purified herpes virus thymidine kinase gene to cultured mouse cell. *Cell*, **11**, 223-232.

Wolf, D.A., Wu, D. and F., M. (1996) Disruption of re-replication control by overexpression of human ORC1 in fission yeast. *J. Biol. Chem.*, **271**, 32503-32506.