ISAS - INTERNATIONAL SCHOOL
FOR ADVANCED STUDIES

Fine Mapping of Human DNA Replication Origins

Thesis submitted for the Degree of
Doctor Philosophiae
(Academic Year 1995/96)

Candidate: Sanjeev Kumar

Supervisors: Prof. Arturo Falaschi
                    Dr. Mauro Giacca

Index

INTRODUCTION

DNA replication is an essential process to conserve genetic information through generations. Logistical complications underlie the replication of DNA especially in the case of higher eukaryotes. Not only billions of base pairs of this nucleic acid, need to be copied just once, with exquisite accuracy and at just the right time in the cell cycle; but also these bases are distributed over numerous chromosomes (46 in case of the humans) that have to maintain their integrity with each passing generation. Furthermore, the DNA replication starts at hundreds to thousands of sites, some of which are triggered early in the synthesis (S) phase of the cell cycle, while others are triggered late (see Huberman and Riggs, 1968; Marx, 1995).

What comes out of years of research in the field on the relatively simple genomes like those of prokaryotes and yeast, and today seems to hold true also for the complex metazoan genomes, is the presence of specific protein-DNA interactions in initiating DNA replication. In other words, replication begins at/near the sites where specific proteins bind specific DNA sequences to initiate DNA unwinding (see Kornberg and Baker, 1992; DePamphilis, 1993). This event is followed rapidly by initiation of DNA synthesis on one or both DNA templates. This synthesis most frequently proceeds by the standard replication fork mechanism as shown in Figure 1, where on the forward arm (leading strand) the synthesis is continuous while that on the retrograde arm (lagging strand) is discontinuous occurring via repeated synthesis

4

ORI

Forward Arm
Continuous DNA Synthesis

Retrograde Arm
Discontinuous DNA Synthesis

5'

3'

Okazaki Fragments

Okazaki Fragments

3'

5'

Retrograde Arm
Discontinuous DNA Synthesis

Forward Arm
Continuous DNA Synthesis
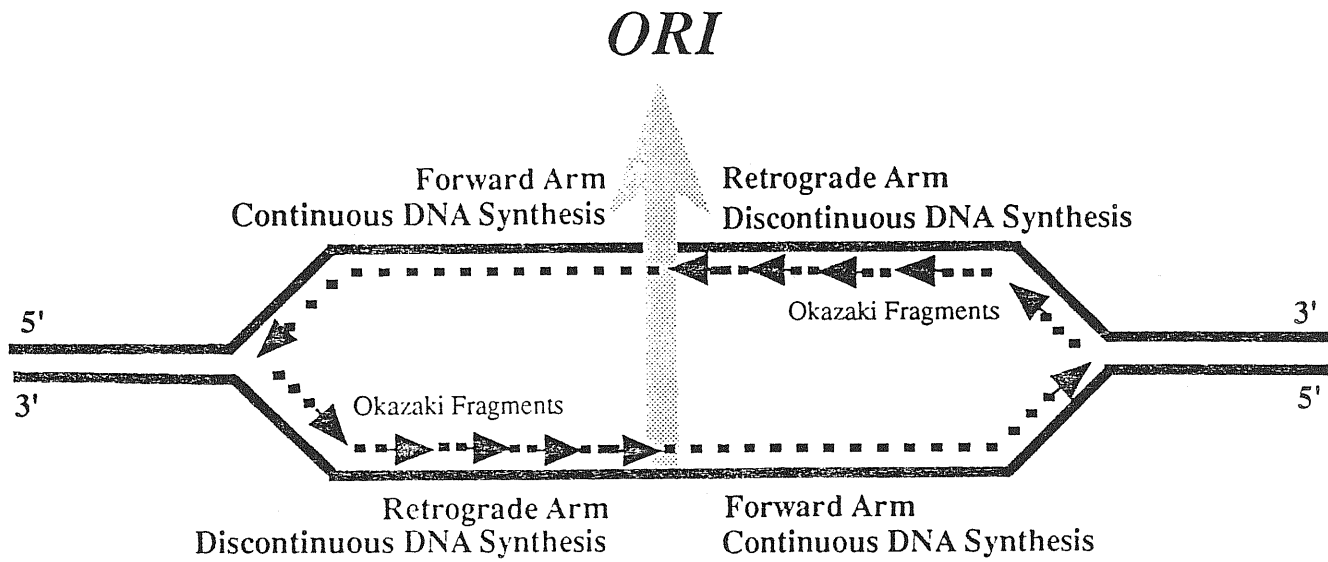
Figure 1. The replication fork model of DNA replication (for details see text).

Replication
Proteins

Transcription
Factors

Origin Recognition
Proteins

Transcription
Factors
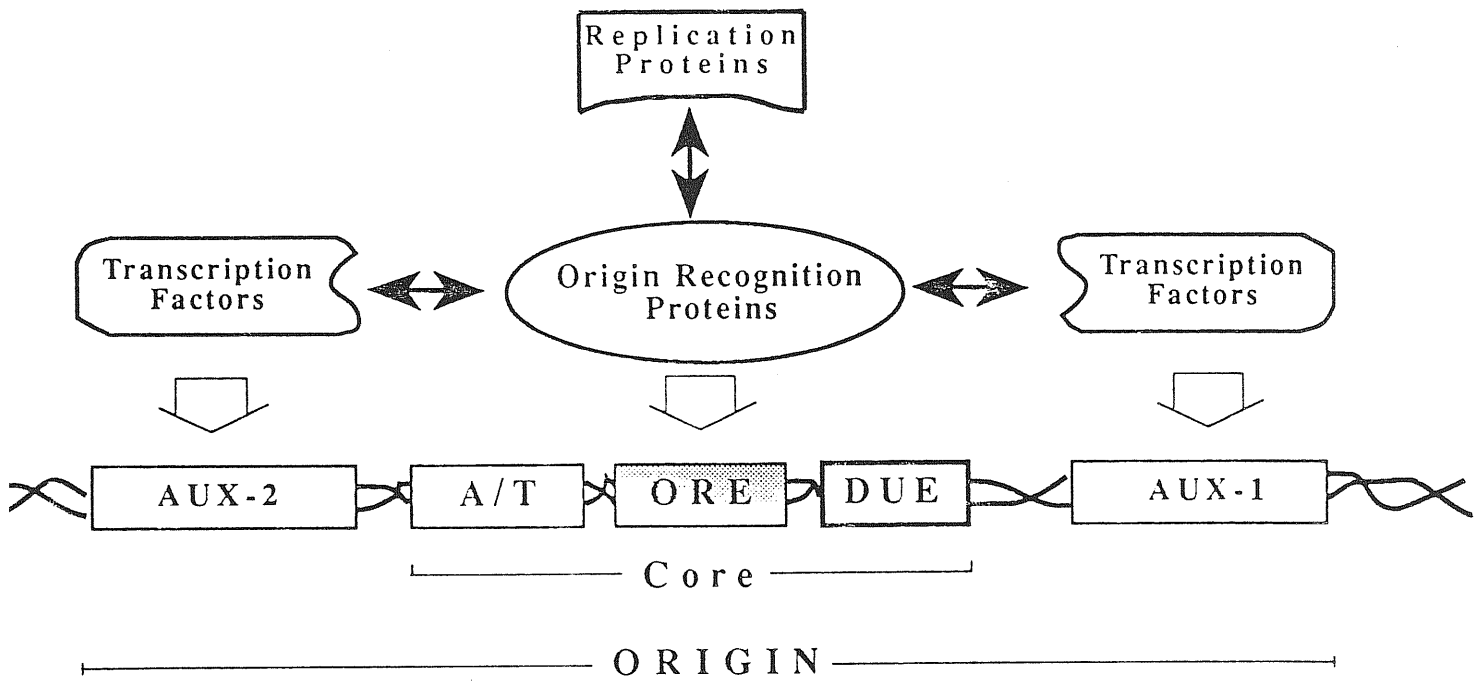
AUX-2   A/T   ORE   DUE   AUX-1

Core

ORIGIN

Figure 2. Origin-structure in simple genomes (for details see text).

and joining of short RNA-primed nascent DNA chains commonly referred to as Okazaki fragments.

1.1 Origins of replication in simple genomes

It is the above described melting of DNA to initiate replication that gives rise to the concept of an *origin* of replication, which defines the cis-acting site where the DNA synthesis begins. In simple genomes such as those of viruses and prokaryotes an *origin* is a precisely defined constant site.

Origins of replication (*ori*) in simple genomes share a common anatomy: an origin recognition element (ORE), a DNA-unwinding element (DUE), and one or more binding sites for specific transcription factor(s) as represented in Figure 2. The ORE and DUE together with an A/T rich element constitute what is normally referred to as the *ori* core. Spacing, orientation, and arrangement of these three core components are usually critical for *ori* function (see DePamphilis, 1993). The A/T rich element in some cases as in herpes simplex virus, may serve as the DUE itself. Origin recognition proteins bind specifically to their cognate DNA-binding site, the ORE. Table I lists a few typical examples of such proteins.

5

Table I. Proteins that specifically activate origins of replication in simple genomes (see DePamphilis, 1993; Bell, 1995; Donovan and Diffley, 1996)

Origin	Origin Recognition Protein(s)	Transcription Factors	
		Aux-2	Aux-1
SV40 *ori*	T-ag	Sp1, T-ag	T-ag
PyV *ori*	T-ag	AP1	T-ag
BPV *ori*	E1	E2	
Ad2 *ori*	(preTP:Ad DNA pol)	NF1, OCT1	
EBV *ori*P	EBNA-1	EBNA-1	
HSV *ori*S	UL9	several candidates	
mtDNA *ori*H	RNase MRP	mtRNA polymerase	
Yeast ARS	ORC	ABF1	ABF1

The DUE is an easily unwound DNA region. Although it is not a unique sequence like an ORE, its dependence on nucleotide sequence and not solely on the A/T content derives from the dependence of DNA unwinding on base-stacking interactions. DUE appears to be the site where DNA unwinding begins and studies on the SV40 genome show the very presence of the *ori* precisely where DNA melting in the DUE begins (Hay et al., 1982; Hay et al., 1984; Guo et al., 1991). *Ori* auxiliary (AUX) components consist of transcription factor-binding sites that facilitate *ori* core activity (Figure 2). The evidence is now compelling that *ori* auxiliary components must bind specific transcription factors (Table I) in order to facilitate *ori* activity. The reason for this presumably

involves specific protein-protein interactions between transcription factors and the origin recognition proteins. Transcription factors that activate one origin do not necessarily activate another, and the ability of a transcription factor to stimulate a promoter does not necessarily reflect its ability to stimulate an origin (for review see DePamphilis, 1993; Donovan et al., 1996).

1.2 Origins of replication in complex genomes

Fiber autoradiographic studies in the 1960s vividly demonstrated the presence of bi-directional origins of replication at around 100 kb intervals in mammalian DNA (Huberman and Riggs, 1968). However, it has been extremely difficult to prove that these sites actually correspond to genetic replicators. This difficulty in the identification of OBRs (Origin of Bi-directional Replication) in mammalian cells derives from the non-functionality of the conventional methods of *ori* search based on its ability to sustain autonomous replication. While the ARS (autonomously replicating sequence) assay led to the identification of the yeast replicators (Stinchomb et al., 1980; Chan et al., 1980), it has failed to give satisfactory results in origin-localization in the higher eukaryotes (Gilbert and Cohen, 1989; Burhans et al., 1990; Caddle and Calos, 1992). Despite a few reports that some cloned sequences may be more efficient "replicators" than others in such ARS assays

A

ori

5'
3'
3'
5'

B

Mapping the localization of nascent
DNA strands

C

Mapping the inversion of polarity of
leading or lagging strands

D

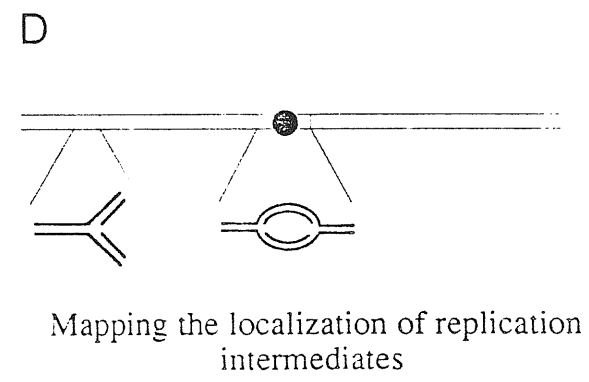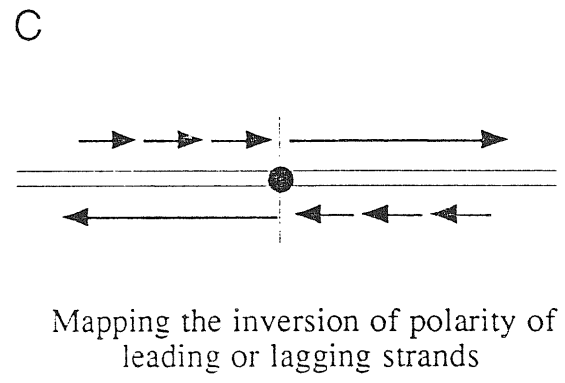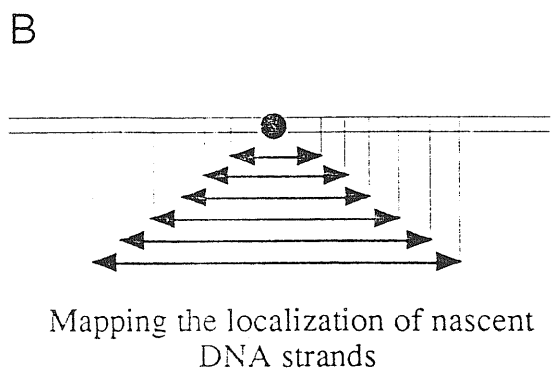Mapping the localization of replication
intermediates
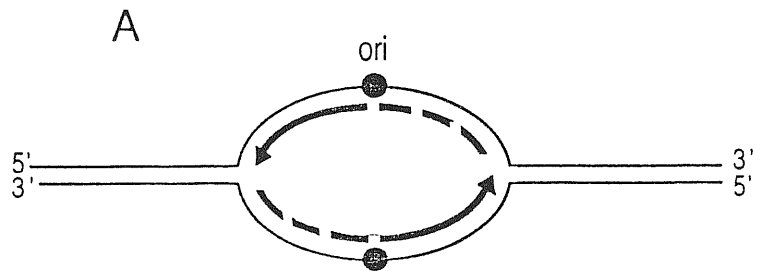
Figure 3. The basis of origin-mapping techniques (see section 1.3 for details).

(Frappier et al., 1987; Sudo et al., 1990; Mc Whinney et al., 1990; Virta-Pearlman et al., 1993; Berberich et al., 1995), it has yet to be shown that mutations in any such elements alter origin function. Indeed, most mammalian sequences seem to replicate to some extent when re-introduced into mammalian cells (Heinzel et al., 1991), which at the very least, results in an unworkable signal-to-noise ratio.

1.3 Approaches to map metazoan origins of replication

The failure of ARS studies to identify mammalian *oris* led to the devising of alternative strategies towards this end (see Falaschi et al., 1993; Falaschi and Giacca, 1994; Hamlin and Dijkwel, 1995). The techniques available today for origin-mapping in metazoan cells can be basically divided, depending upon the property of an OBR (*ori*) they exploit, into the following three categories (Figure 3B-D): structural analysis of the replication intermediates, analysis of the leading/lagging strand polarity and chromosomal mapping of nascent DNA strands.

1.3.1. *Structural analysis of the replication intermediates* (Figure 3D)

Two complementary two-dimensional (2-D) gel electrophoretic methods were introduced in 1987 that each provide an intimate

8
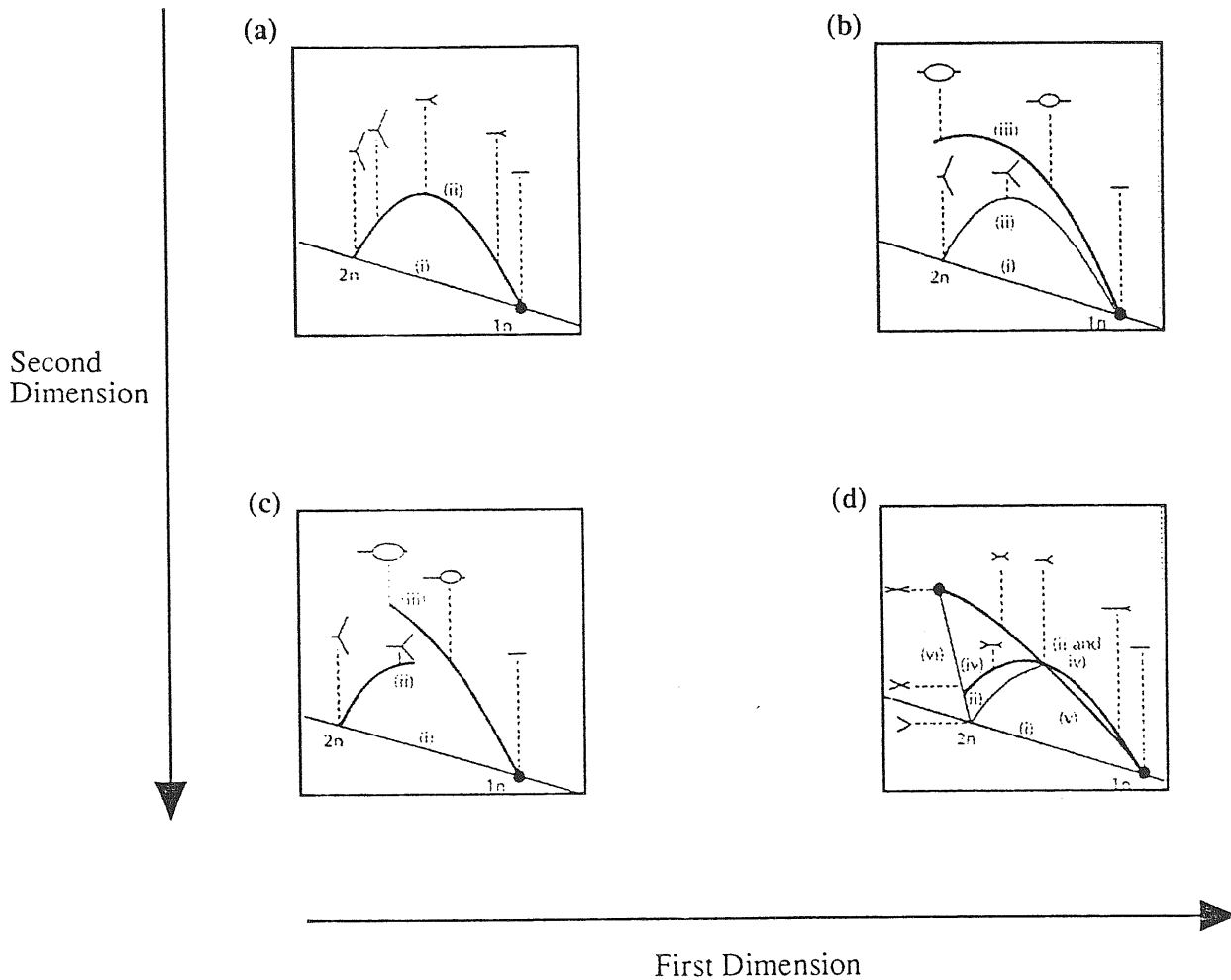
Figure 4. The principle of the neutral/neutral 2-D gel replicon-mapping technique. Idealized autoradiographic images obtained when a digest of replicating DNA is hybridized with probes for fragments that contain different intermediates. Curve (i) represents the diagonal of non-replicating fragments from the genome as a whole. (a) A complete single fork or simple Y-arc (ii) from a fragment replicated passively from an outside *ori*. (b) A fragment with a centered *ori* traces a curve (iii) that arches over the single fork arc curve (ii). (c) A fragment containing an off-centered *ori* traces the beginning of a bubble arc (curve iii) when the replication fork crosses one restriction site. (d) A fragment containing two approaching forks. If they meet asymmetrically or symmetrically, curves iv and v are obtained, respectively. If there is a fixed terminus in a fragment, the collected X-shaped structures would result in a concentrated spot somewhere on curve vi, on which would fall also any recombination structures present.

(but different) view of replication intermediates. Both techniques were devised originally to analyze yeast replicons and were used to show that ARS elements usually, but not always, serve as origins in the chromosome (see Brewer and Fangman, 1991; Newlon and Theis, 1993).

In the neutral/neutral 2-D gel method (Brewer and Fangman, 1987), replication intermediates are separated in the first dimension on the basis of molecular mass, which varies from 1n (unreplicated) to 2n (fully replicated), and in the second dimension on the basis of both mass and shape. Characteristic arcs are traced by fragments that contain a single fork (Figure 4a), an origin (Figure 4b,c) or a termination structure (Figure 4d). The replication intermediates in a region of interest are then visualized by hybridizing transfers of the digest with an appropriate radioactive probe.

The neutral/alkaline 2-D gel method (Nawotka and Huberman, 1988) determines the direction of fork movement through a chromosomal locus. A restriction digest of replicating DNA is separated by mass in the first dimension. The nascent strands are then released from their templates with alkali and are separated by size in the second dimension. The direction of fork movement through any given restriction fragment can be determined by probing sequentially with adjacent probes: an origin-proximal probe traces a complete diagonal, whereas an origin distal probe detects only the nascent strands (Figure 5). Logically it can be

9

(a)

Double-stranded
replication intermediates

First dimension
(neutral)

2n

In

Single-stranded
nascent chains

Second dimension
(alkaline)

0

Specific
probes

(b)

DNA

*

Resulting autoradiograms
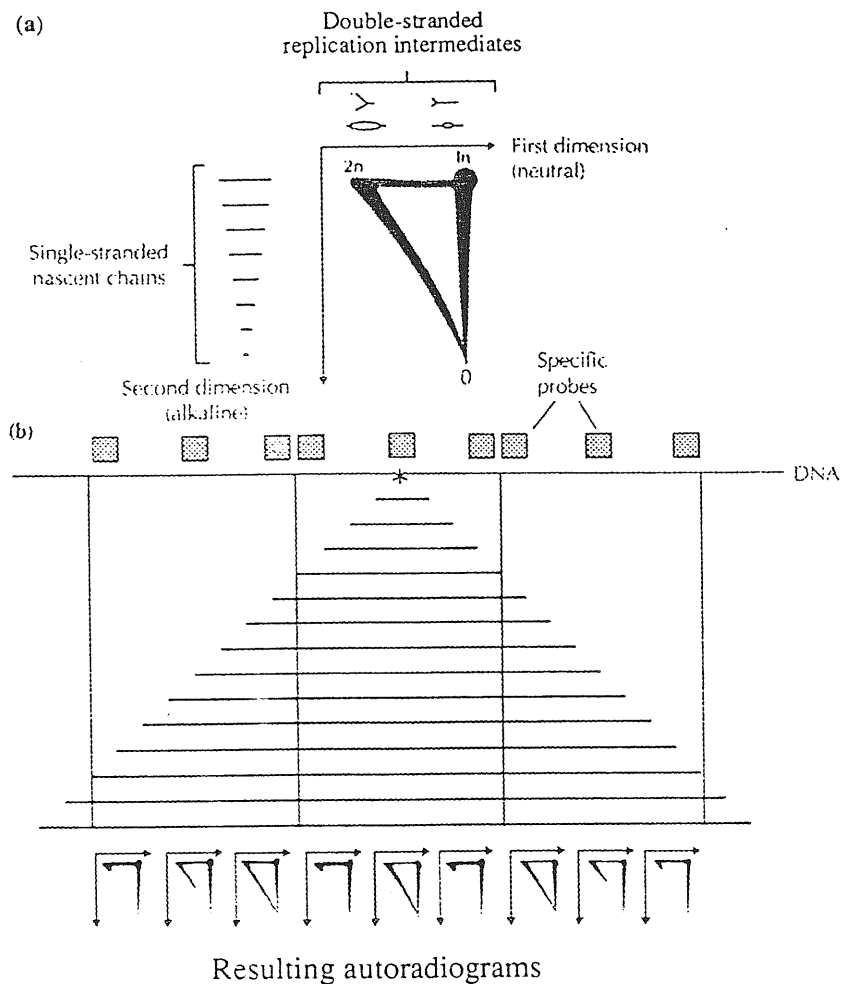
Figure 5. Principle of the neutral/alkaline 2-D gel replicon-mapping technique. (a) Idealized patterns of migration of double-stranded replication intermediates in the first (neutral) dimension and the single stranded nascent chains in the second (alkaline) dimension. (b) Three adjacent restriction fragments are demarcated by the vertical lines. An *ori* is positioned in the central fragment (asterisk), and nascent strands of different sizes are indicated. The panels below show the autoradiograms that would be obtained when each of the indicated probes (shaded boxes) is used to illuminate its cognate restriction fragment.

deduced that a region between divergently moving forks must contain an origin.

2-D gel based origin-mapping methods have been utilized successfully to map replicons in simple genomes (see Vassilev and DePamphilis 1992; Benard et al., 1996, Dhar et al., 1996; Little et al., 1995). However, when applied to complex genomes the 2-D gel based methods become problematic due to their relatively low resolution limits and the potential ambiguities in the interpretation of the results (Linskens and Huberman, 1990). Despite this consideration, there still exist a few convincing reports on origin localization using these methods (Table II). All such reports, however. are puzzling since in most cases the hybridization patterns obtained indicate the presence of an origin-region rather than a fixed *ori*-site. The resulting hybridization patterns obtained in 2-D gel-based methods usually are more complex to analyze than those expected from the model. suggesting the possible presence of unusual DNA structures responsible for the unexplained hybridization spots. A good example is the DHFR *ori* in the chinese hamster ovary (CHO) cells where using methods other than those based on the 2-D gel technique an origin was mapped to a precise location downstream from the DHFR (dihydrofolate reductase) gene in a chinese hamster cell line with multiple copies of the DHFR region (Handeli et al., 1989; Burhans et al., 1990; Pelizon et al., 1996); the application of the 2-D gel methods, on the contrary, showed a

10

delocalization of origins in a broad initiation zone of 55 kb (Vaughn et al., 1990; Dijkwel and Hamlin, 1992).

Table II. Metazoan *oris* mapped with methods based on 2-D gel electrophoresis

Region	Organism	Main conclusion	References
DHFR locus	Chinese hamster	55 kb initiation zone with preferred delocalized *ori*-sites	Vaughn et al.,1990 Dijkwel et al., 1992
Chorion locus	Drosophila	*ori* contained in 1 kb region	Heck et al., 1990
Histone gene	Drosophila	multiple initiation sites in the 5 kb histone transcription unit	Shinomiya et al., 1993
rDNA locus	Human	multiple initiation sites in 31 kb non transcribed spacer (NTS)	Little et al., 1993
	Drosophila	initiation-zone in the NTS	Hyrien et al., 1995
Puff II/9A	Sciara sp.	around 1 kb initiation zone	Liang et al., 1993 Liang et al., 1994
Polymerase α gene	Drosophila	10 kb initiation zone 3' to pol α gene	Shinomiya et al., 1994

The 2-D gel mapping approach has been best applied to highly amplified loci and that too utilizing synchronization procedures, typically with aphidicolin, to enrich for the newly synthesized DNA. Aphidicolin, an inhibitor of at least three eukaryotic DNA polymerases (Wang, 1991), blocks progression but not initiation of DNA synthesis (Huberman, 1981; Mosca et al, 1992). Moreover, inhibition of DNA replication can cause reinitiation events

(Mariani and Schimke, 1984), possibly leading to multiple-branched initiation intermediates and to branch migration of nascent DNA (Dijkwel et al, 1991). To rule out these potential artifacts, another synchronizing agent mimosine (Lalande 1990) has been utilized in the studies that still tend to confirm the delocalization of origins (Dijkwel and Hamlin, 1992; Mosca et al., 1992). It is however important to note that the mechanism of action of mimosine in cell synchronization is yet unknown. Moreover, the utilization of 2-D gel approach in animal cells requires previous enrichment of replication intermediates by DNA purification procedures, including benzoylated naphthoylated(BND)-cellulose chromatography (Levine et al., 1970) or selective isolation of the nuclear matrix (Dijkwel et al., 1991); procedures that could easily alter the structure of replication intermediates. for example, by introducing breaks into replication bubbles consequently masking the presence of an active *ori* (Linskens and Huberman, 1988). The 2-D gel mapping results, therefore, should be interpreted with a cautious skepticism, especially when applied to the higher eukaryotic cells.

1.3.2. *Analysis of the leading/lagging strand polarity* (Figure 3C)

Since DNA polymerase synthesizes DNA only in the 5' to 3' direction, an OBR is also the site where continuous DNA synthesis on the leading strand and discontinuous DNA synthesis on the lagging strand, switch template and invert polarity (Figure 3B).

The lagging strand origin-mapping assay (Burhans et al., 1990) relies on this polarity switch at the origin in the template for the Okazaki fragment synthesis. Template bias can be determined by hybridizing radiolabelled Okazaki fragments to the positive and negative strands of recombinant M13 clones containing fragment pairs from a region of interest. Although this method has been used to successfully map some metazoan origins (Table III), it suffers from an important drawback: the Okazaki fragments are scattered uniformly along the genome and thus it is not feasible to use this method to detect *ori* in single copy loci in physiologically growing cells. In order to obtain a sufficiently high signal-to-noise ratio, therefore, the cells need to be synchronized, permeabilized and labelled with very high amount of ^{32}P-dNTPs (Burhans et al., 1990), creating the same problems as listed in the earlier section.

A leading strand polarity-switch based assay, on the other hand, relies on the observation that the protein synthesis inhibitor emetine selectively inhibits lagging strand synthesis (Burhans et al., 1991). By using emetine-resistant bromodeoxyuridine-labelled (leading strand) DNA as a probe on positive and negative strands of recombinant M13 clones from the region in question, it is possible to determine the switch site for the leading strand synthesis and hence, also the origin. The main drawback of this method besides its being tedious, is the lack of knowledge on the exact mode of emetine action and whether such a treatment of

the cells that alters cellular metabolism considerably, also affects the physiological pattern of DNA replication initiation.

Table III. Metazoan *oris* mapped with methods based on leading/lagging strand polarity switch

Region	Organism	Main conclusion	References
DHFR locus	Chinese hamster	*Ori* site within a 15 kb region	Handeli et al., 1989
APRT locus	Chinese hamster	*Ori* site within a 6 kb region	Handeli et al., 1989
ADA gene	Mouse	11 kb origin region 5' to ADA gene	Caroll et al., 1993
β-globin	Human	*Ori* mapped in a 2 kb fragment	Kitsberg et al., 1993
Ribosomal protein S14 gene (RPS14)	Chinese hamster	origin mapped in a 2.5 kb region overlapping the gene	Tasheva et al., 1994
CAD gene	Syrian and Chinese hamsters	origin localized in a 5 kb region in the CAD transcription unit	Kelly et al., 1995
c-myc	Human	origin mapped within 2.5 kb region upstream of the gene	Waltz et al., 1996

1.3.3. *Chromosomal mapping of nascent DNA strands*

DNA replication initiation is characterized by the production of short nascent DNA strands. These can be distinguished by the bulk DNA on the basis of their being synthesized earlier than any

other DNA sequence within the same replicon. Since these nascent strands represent an extremely small portion of the total genome the methods based on such an approach aim at obtaining a considerable enrichment of nascent strands sufficient enough to give a detectably high signal in origin localization.

Utilization of the replication timing studies in mapping mammalian *ori*s is particularly limited by the resulting low resolution. Although several studies have been conducted on the replication timing of contiguous loci in selected chromosomal regions and excellent information on the replication properties of wide regions of the genome could be obtained, yet no *ori* could be mapped precisely (Calza et al., 1984; Brown et al., 1987; Dhar et al., 1988; Taljanidisz et al., 1989; Gale et al., 1992; Selig et al., 1992; Spack et al., 1992).

Increased resolution in such studies can be obtained when applied to highly amplified regions where there is naturally a higher abundance of the nascent strands as compared to the single copy regions. Such studies on the CHOC 400 cells (a chinese hamster cell line with more than 1000 copies of an approximately 240 kb long genomic region encompassing the DHFR gene) suggested the presence of two origins of replication spaced about 22 kb apart downstream from the DHFR gene (Ma et al., 1990). The method consisted in synchronizing the cells at G1/S border and labelling the nascent DNA fragments with a radioactive nucleotide precursor immediately after the cells' entry into S

15

phase. The radiolabelled products were then analyzed either by direct visualization after specific restriction enzyme digestion (Heintz and Hamlin, 1982), or used as probes against cloned DNA fragments of the region (Burhans et al., 1986). A similar method, where the nascent-strand abundance in asynchronously growing cells was quantitatively evaluated using densitometric scans of the resulting southern-hybridization signals, recently led to the identification of an *ori* within the naturally repeated human ribosomal DNA (Yoon et al., 1995). Though utilizing an in-gel renaturation procedure to eliminate most of the background due to non amplified genomic sequences the technique could be further refined (Leu and Hamlin, 1989); the major drawback of such replication timing studies i.e. their low resolution in *ori* mapping, continues to exist; and the techniques even in the case of highly amplified loci like that of DHFR in CHOC 400 cells have been complemented with still other kinds of origin mapping methods (see Hamlin and Dijkwel, 1995).

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

16

major drawback is the potential alteration of the physiological controls of DNA replication upon psoralen treatment.

Employing high temperature to extrude nascent DNA from replication bubbles in synchronized cells (Zannis-Hadjopoulos et al., 1981) one can purify the nascent strand population in the double stranded form by CsCl density gradient centrifugation if the synthesis of DNA has been allowed to occur in the presence of a heavy precursor such as BrdUrd (bromodeoxyuridine). Such a method has been utilized to map an *ori* region near the chicken α-globin gene (Razin et al., 1986). A similar approach where the label was instead Hg-dCTP and affinity chromatography as the purification technique, was used to construct a library of monkey DNA sequences putatively enriched in *ori* (Kaufmann et al., 1985).

In vitro *run off* technique is another interesting method where temporal order of replication of a defined chromosomal region, and hence the origin, is mapped. This technique consists of allowing isolated nuclei to elongate DNA chains initiated in vivo in the presence of a heavy DNA precursor. Since DNA in isolated nuclei does not undergo reinitiation events, the extent of incorporation of the label in each molecule is directly proportional to its distance from the *ori*. Such a method has been the basis of a few reports where, although at low resolution, some *ori*s have been localized (James and Leffak, 1986; Heintz and Stillman, 1988; Trempe et al., 1988; Leffak and James, 1989; McWhinney and Leffak, 1990).

Yet another kind of approach that surpasses all others mentioned above in its high sensitivity for *ori*-mapping, is the use of polymerase chain reaction (PCR) to estimate the relative abundance of various DNA regions in the nascent strand pool (Vassilev and Johnson, 1990; Vassilev et al., 1990; Ariizumi et al., 1993; Shinomiya et al., 1993; Virta-Pearlman et al., 1993; Taira et al., 1994; Tasheva et al., 1994). While all the other methods described before, suffer from the major drawback of poor sensitivity, the conventional-PCR based method is hampered by its inability to provide indisputable quantitative results since many predictable and unpredictable variables affect its outcome (Diviacco et al. 1992). Table IV lists various reports on origin-localization using methods based on analysis of nascent DNA strand size.

If the drawbacks that marr the usage of otherwise highly sensitive PCR-based techniques in *ori*-mapping could be overcome, such techniques would become very advantageous towards this end. The following section infact, describes a novel easy, highly sensitive, efficient and precise method based on competitive PCR devised in our lab, and its successful application in identifying a human *ori*.

18

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

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

19

STUDIES ON THE LAMIN B2 ORIGIN

2.1. INTRODUCTION

2.1.1. A novel *ori*-mapping method

During the past few years in our lab a novel method for ori-mapping was devised (Diviacco et al., 1992). This method utilizes competitive PCR to map the relative abundance of various sets of DNA markers in the nascent DNA pool. As shown in Figure 6, an origin of bidirectional DNA replication in a cell culture chronically labelled with ^{14}C-thymidine is pulse labelled with ^{3}H -deoxycytidine and cold BrdUrd for different times. The pulse labelled DNA is isolated, denatured and fractionated on neutral sucrose density gradients. The density gradient separation serves to separate the nascent DNA rich in origin-sequences from the latter since the latter DNA fragments having originating recently are very short. This BrdUrd substituted nascent DNA fraction is then purified through a column containing an immobilized monoclonal antibody against BrdUrd, yielding more than 95% of ^{3}H-labelled DNA and practically none with ^{14}C-label (Contreas et al., 1992). These fragments are subsequently assayed for the abundance of selected markers with quantitative PCR. The nearer a marker is to the origin the higher is its abundance in the nascent DNA pool, the abundance peaking at the origin.
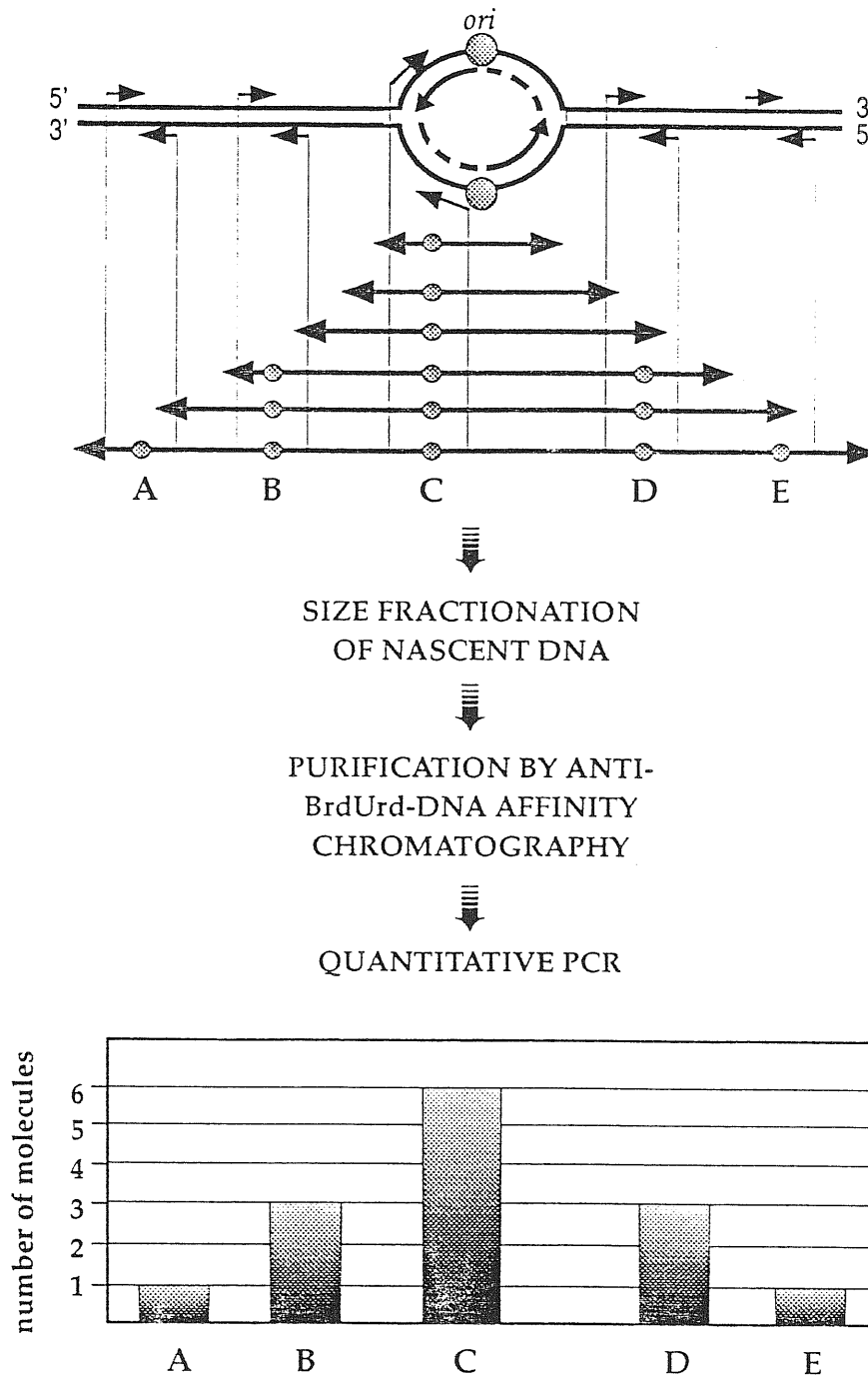
Figure 6. The novel *ori*-mapping procedure based nascent DNA strand analysis with quantitative-PCR. Short fragments of nascent DNA, synthesized in the presence of BrdUrd, can be isolated and separated from the bulk DNA on a neutral sucrose gradient. Further purification of this DNA can be done by immunoaffinity chromatography with anti-BrdUrd antibodies. Within this population of newly synthesized DNA molecules, the abundance of selected fragments (A through E) scattered within a genomic region can be precisely known by quantitative-PCR. The pair of primers (small arrows on the topmost figure) amplifying the highest number of molecules in the sample, is closest to the origin (marker C).

The quantitative PCR method devised in our lab takes advantage of the addition of a competitor molecule to the amplification reaction. in this way, any kind of predictable and unpredictable variations that affect the reaction for a given template would affect in the same way also the competitor since the competitor has the same sequence as that of the template except for a minimum negligible insertion of 20 base pairs. Such an insertion could then make this competitor molecule easily distinguishable from the template when electrophoresed on a polyacrylamide gel. Since the ratio between the amounts of template and the competitor in the reaction remains constant throughout the reaction, coamplifying a fixed amount of sample with known amounts of competitor one can easily calculate the amount of specific template molecules present in the sample.

The competitor molecules are constructed easily by PCR as shown in Figure 7. For each template two internal primers with contiguous 3' ends are synthesized one for the upper and the other for the lower strand with a tail of unrelated but complementary 20 nucleotides at the 5' end. These primers are used in two separate PCRs (Figure 7b, b'), each in combination with the corresponding external primer for the template. The two partial products are then eluted from the gel after PAGE, mixed together, annealed and amplified with external primers to get the desired competitor molecule with a 20 bp insertion in the middle.
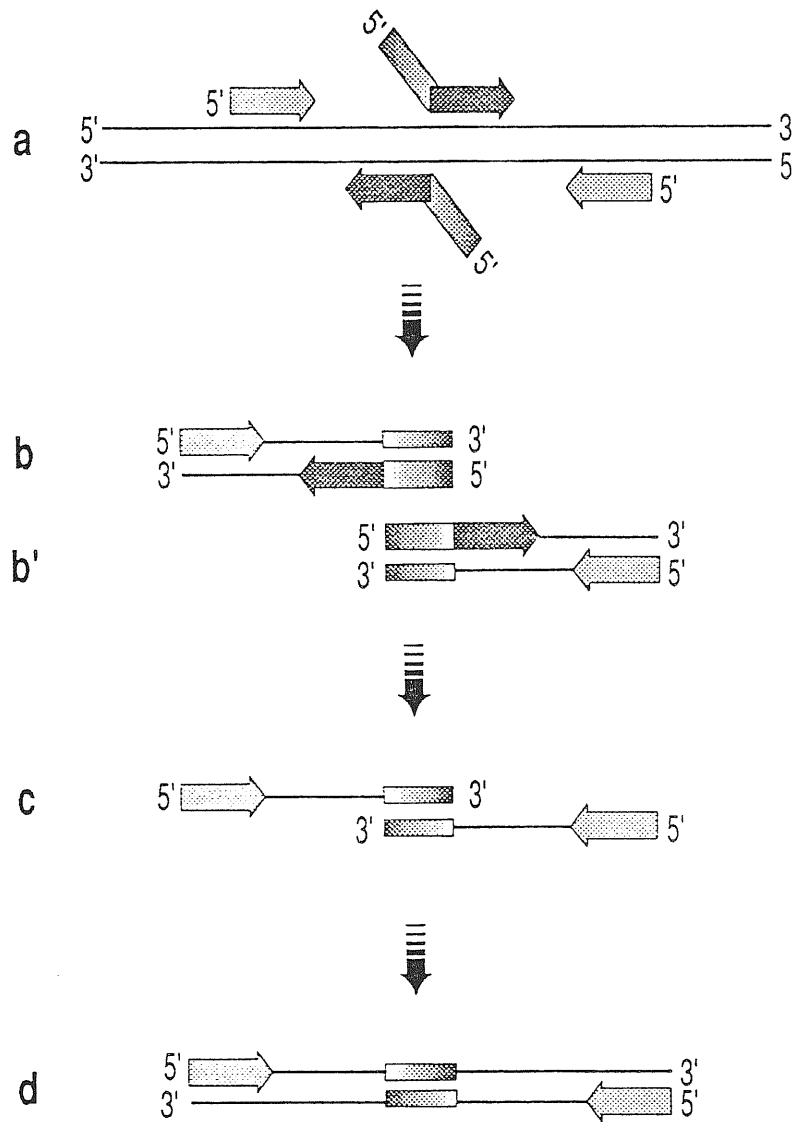
Figure 7. Construction of the competitor molecule. For a detailed explaination see section 2.1.

The validity of the above described method in mapping an *ori* efficiently was ascertained by its use in mapping an already well known *ori*, that of the SV40 (Giacca et al., 1994).

2.1.2. The Lamin B2 *ori*

In the previous years a library of DNA synthesized at the onset of the S-phase from HL60 (human myeloid leukemia cell line) cells was constructed in our laboratory (Tribioli et al., 1987). These cells were synchronized by two cycles of blocks with aphidicolin and the BrdUrd-containing newly replicated DNA isolated on sucrose gradients immediately after the second block. This early-replicating DNA was then cloned to get a library that is presumably rich in origins that are activated at the onset of S-phase. A particular sequence (B48) from this library was synthesized in the very first minute of the onset of S-phase in synchronized HL60 cells and mapped to the G-negative band p13.3 of chromosome 19 (Biamonti et al., 1992a). A stretch of this DNA was sequenced (14 kb) and was shown to contain the terminal portion of the lamin B2 gene and another unidentified transcript (ppv1) firing in the same direction.

The above sequence was then screened for the presence of an *ori* using the quantitative PCR method. An *ori* was localized, with the most abundant marker, the B48 primer set, to the non-

transcribed region between the end of the lamin B2 gene and the beginning of the ppv1 transcript in synchronized HL60 cells (Biamonti et al., 1992b). The origin region was subsequently narrowed down to about 470 nucleotides and, given the high efficiency of the method, was localized also in the non-synchronized cell population (Giacca et al., 1994). Figure 8 shows a schematic representation of the region with the location of the various primer sets used in the study.

At this point we wondered if the same origin was active also in human cells that have a different histological derivation from that of HL60. This part of the thesis describes the experiments performed to address this question. The results show that our origin-mapping method could be simplified further and that the same lamin B2 *ori* is active in asynchronously growing cells of a variety of human cell lines with distinct derivation (two neuroblastoma, one epithelial, one fibroblast) and even in actively growing primary cells (PHA-stimulated peripheral blood lymphocytes) but not in quiescent cells (unstimulated peripheral blood lymphocytes).

SB12

lamin B2 ppv1

SE10 BN1 B48 SE17 B13

origin

1 2 3 4 5 6 7 8 9 10 11 12 13 14 kb
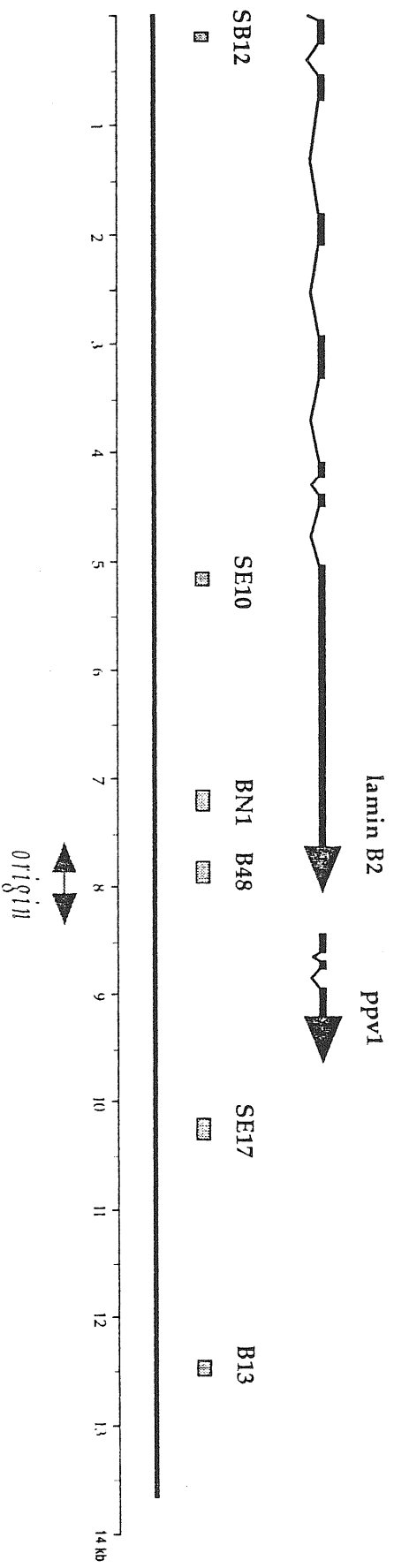
Figure 8. Schematic representation of the human genomic region analyzed in this study. The 13.7 kb region of the human chromosome 19 investigated contains the 3' end of the lamin B2 gene and another tandemly arranged small gene (ppv1). Bold straight portions of the respective gene transcripts shown by arrowed lines represent exons while the vents represent introns. The gray boxes show the locations of the segments amplified by the various primer sets used. The sequence of the core portion of this region (humlambbb) is available in GenBank (accession number M94363).

2.2. MATERIALS AND METHODS

2.2.1. Cell cultures and DNA labelling

HL-60 (myeloid), IMR32 (neuroblastoma) and SKNMC (neuroblastoma) cells were cultured in RPMI 1640 with 15% fetal calf serum; HeLa (epithelial), IMR90 (lung fibroblasts) and SKNBE (neuroblastoma) cells were cultured in Dulbecco's modified Eagle medium with 12% fetal calf serum. Both media were supplemented with 2 mM glutamine and 50 μg/ml gentamicin. Peripheral blood mononuclear cells were obtained by density centrifugation of 200 ml peripheral blood from a normal donor over Ficoll-Hypaque (Sigma, St. Louis, MI). DNA from resting lymphocytes was obtained from 10^8 non-adherent cells after a 30-minute incubation in complete RPMI 1640 medium.

Lymphocytes were activated by incubation in culture medium containing 1 μg/ml phytohemagglutinin (PHA, Sigma) and 10% interleukin-2 (IL-2, Cellular Products Inc., San Diego, CA). Five days after stimulation, DNA was extracted from 10^8 cells.

When appropriate (see below) cells (1 x 10^8) were pulse-labeled for 10 min. at 37°C either with 1 μM (final concentration) ^{3}H-deoxycytidine (21.5 /mmol, Amersham, UK) and 100 μM (final

concentration) cold 5'-bromodeoxyuridine (BrdUrd; Boehringer Mannheim GmbH, Mannheim, Germany), as described (Giacca et al., 1994) or with 1 μM (final concentration) ^{3}H-thymidine (15.1 Ci/mmol, Amersham).

2.2.2. Extraction and purification of newly synthesized DNA

Total DNA was extracted from exponentially growing, asynchronous cell cultures by standard procedures (Sambrook et al., 1989), denatured by a 10 minute incubation in boiling water, and size-separated (300 μg per gradient) on 35 ml of 5-30% neutral sucrose gradients for 20 hours at 26000 rpm in a Beckman SW28 rotor at 20°C. In parallel, a reference tube with a double stranded size marker DNA (containing four different sized DNA fragments in the range of 500 to 5000 bp) was also run on an identical gradient. Fractions of 1 ml each were collected and those from the gradient with the markers were run on 1% agarose gel at 30 V overnight. Using the separation pattern of the marker DNA on the gel as a sedimentation velocity reference, the fractions containing ssDNA of ~ 1 kb were selected in each case and dialyzed against Tris-EDTA (0.5M Tris pH 8; 0.01M EDTA) for at least 8 hours. These fractions were then used to quantify ssDNA by competitive PCR using different sets of primers in the lamin B2 region. Only in the case of SKNBE cells, a 10 minute incubation of the cells in 100 μM BrdUrd and 1 μM ^{3}H-deoxycytidine in complete RPMI 1640 medium was performed

prior to DNA extraction, and the size-selected fractions were purified by immunoaffinity chromatography using anti-BrdUrd antibodies (Contreas et al., 1992) before proceeding to quantitative PCR. This is exactly the same procedure as the one reported previously (Giacca et al., 1994); see Figure 9A. The variations to this standard procedure (i.e. omission of the immunoaffinity step, omission of the BrdUrd labelling step, labelling with ^{3}H-thymidine 1μM, etc.) are reported in Table V. Analysis of all samples in Figure 11 was according to the simplified protocol outlined in Figure 9B.

2.2.3. PCR amplification and competitor construction

The primers used for competitive PCR experiments in the lamin B2 genomic area (primer sets B13, B48, BN1, SB12, SE10 and SE17) and within the β-globin gene (primer set PCO) have already been described (Giacca et al., 1994). These primers amplify segments of 100-300 bp; their position in the lamin B2 genomic area is shown in Figure 8. For each primer set, competitor DNA segments were constructed carrying the corresponding genomic sequence with the addition of 20 extra nucleotides in the middle, to allow gel electrophoretic resolution of the amplification products. The procedure for competitor construction was the same as described in reference: (Diviacco et al., 1992).

Competitive PCR experiments were carried out by the addition of scalar quantities of competitor to a fixed amount of each nascent DNA preparation and subsequent amplification with the appropriate primer sets. This procedure and the individual PCR profiles have already been described (Giacca et al., 1994)).

2.3. RESULTS

2.3.1. A simplified origin mapping procedure

The original method for DNA replication origin mapping (Giacca et al., 1994), as described in Figure 9 protocol A, consisted of: i) bromodeoxyuridine (BrdUrd) and ^{3}H-deoxycytidine pulse-labelling of the newly synthesized DNA; ii) DNA extraction and size fractionation of short nascent DNA strands (~1 kb); iii) further purification of newly synthesized, BrdUrd-substituted DNA by immunoaffinity chromatography using anti-BrdUrd antibody coupled to sepharose beads; and iv) quantitation of DNA segment abundance in the purified sample by competitive PCR. By this procedure, nascent DNA (i.e. short DNA stretches emanating from the origin) is firstly purified from total newly synthesized DNA (deriving from all regions in the genome, the cells being unsynchronized) according to its size and then further enriched from the bulk unreplicated genomic DNA by affinity

Asynchronous
exponentially growing
cells

A   B

BrdUrd and ^{3}H-Cyt   No label
pulse label for 10 min.

DNA extraction

Denaturation and size fractionation
on a neutral sucrose gradient

Immunoaffinity
chromatographic
purification of BrdUrd-
rich ssDNA

DNA quantitation by competitive PCR with
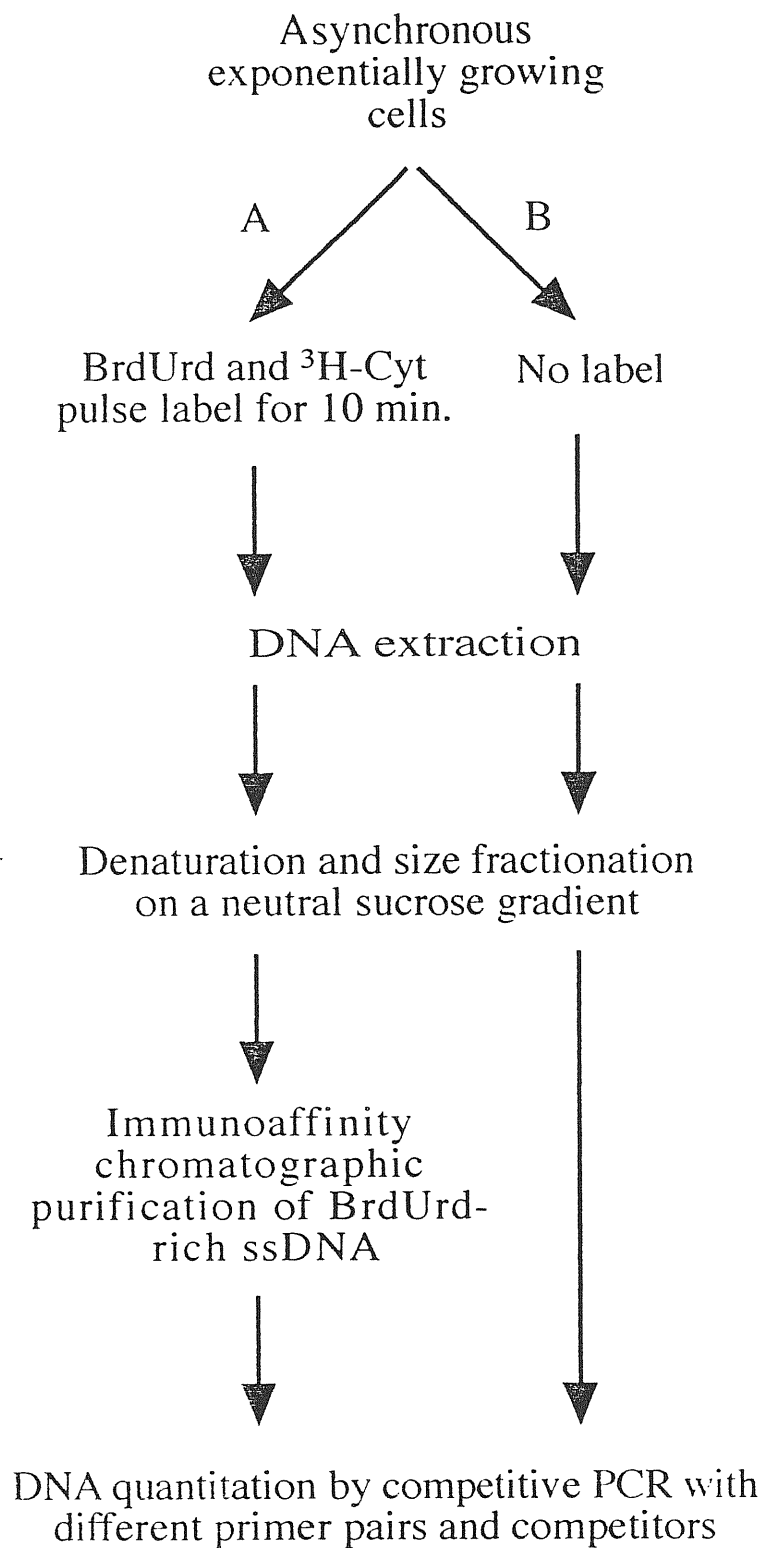different primer pairs and competitors

Figure 9. Flow chart diagram showing the original (A) and the simplified (B)
procedures for preparing newly replicated DNA fractions for *ori*-mapping studies

chromatography. Since the origin mapping procedure by fragment abundance detection relies in principle on size selection only, we attempted a simplification of this procedure by omitting the affinity chromatography step. Accordingly, nascent DNA strands of about 1 kb in size isolated from exponentially growing asynchronous cells of the neuroblastoma cell line SKNBE were directly submitted to quantitative PCR analysis (Figure 9B) using six sets of primers spanning the region of interest (Figure 8) and one control primer set (primers PCO3 and PCO4), which detects an unrelated region in the β-globin gene. As shown in Table V (rows A and B), the omission of the BrdUrd-purification step did not give any appreciable differences neither in the detectability of DNA segments by competitive PCR nor in the relative abundances of the markers. On the contrary, omission of this step, besides simplifying the procedure, also avoids the introduction of a possible bias in fragment selection on the anti-BrdUrd affinity column and reduces the possibility of DNA fragmentation by photodamage.

Given the above reported results, BrdUrd labelling was no longer utilized, and the simplified procedure of Figure 9 protocol B was consistently followed. Line C of Table V reports the results obtained by the application of this procedure to the HL-60 cell line, where the lamin B2 origin had originally been mapped within ~500 bp encompassing primer set B48 (Giacca et al., 1994). Comparison of the marker abundance distribution in this cell line

with that obtained in SKNBE cells indicates that also the latter utilizes the lamin B2 *ori*.

Table V. Comparison of the complete and simplified procedures for origin mapping by quantitation of DNA segment abundance in samples of nascent DNA

MARKERS

		SB12	SE10	BN1	B48	SE17	B13	PCO
Cell Line	Method Used	Relative Abundance *(arbitrary scale)*						
SKNBE	A	<500	300	900	1600	400	120	200
	B	450	300	900	1200	N.D.	120	200
HL60	C	400	600	1200	3000	850	400	500
	D	400	500	1200	3000	900	400	500
		Absolute Abundance *(number of primer specific molecules per 10^6 total ssDNA molecules)*						
	D*	4.8	7.2	14.4	36.1	10.2	4.8	6.0

A: Complete procedure as shown in Figure 9 protocol A and explained under Materials and Methods.

B: Same as A but with the omission of the immunoaffinity chromatography step (Figure 9 protocol B).

C: Same as B but without any labeling of cells (BrdUrd or radioactivity).

D: Same as C with a 10 min incubation of the cells with ^{3}H-thymidine prior to DNA extraction.

D*: Same as D, but the results are expressed as absolute amounts i.e. number of primer specific molecules per 10^6 total ssDNA molecules.

We have previously shown. by utilizing the BrdUrd procedure in asynchronous cells, that the lamin B2 *ori* was present with a
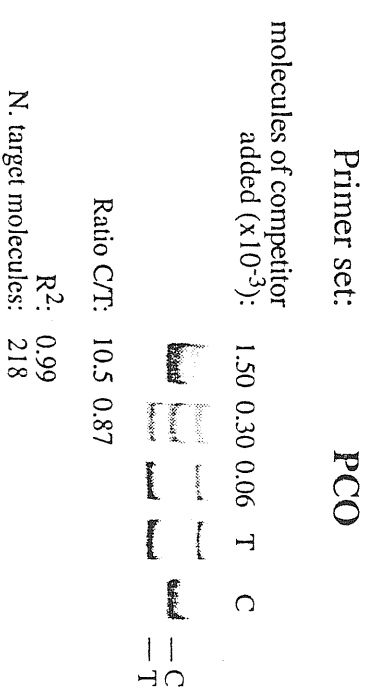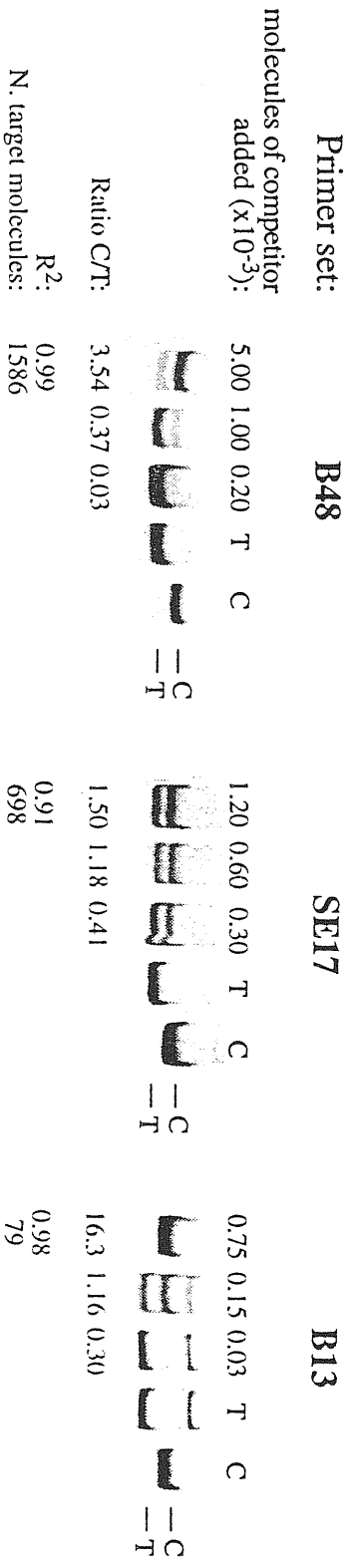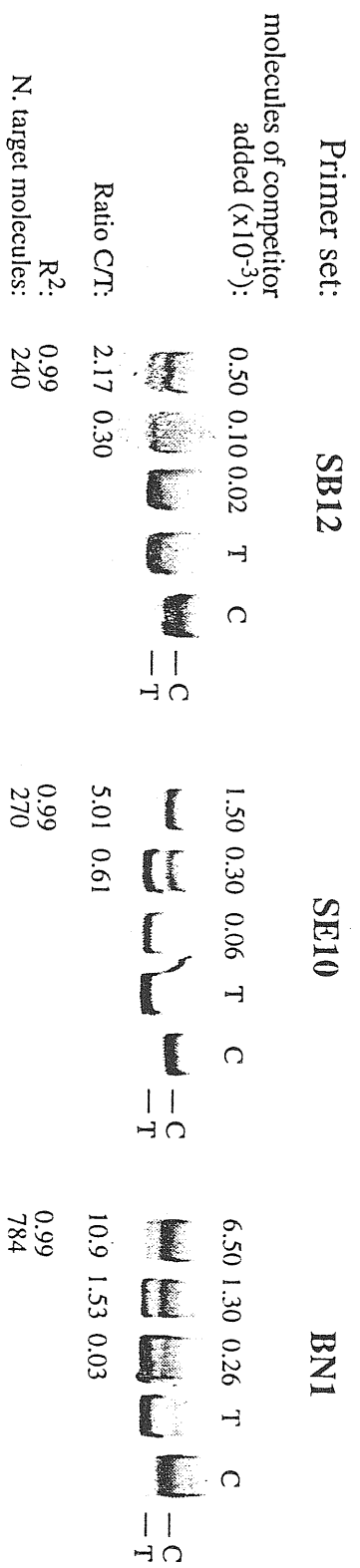
3 0

Primer set: SB12

molecules of competitor added (x10^{-3}):	0.50	0.10	0.02	T	C
Ratio C/T:	2.17	0.30			—C —T
R^2:	0.99				
N. target molecules:	240				

Primer set: SE10

molecules of competitor added (x10^{-3}):	1.50	0.30	0.06	T	C
Ratio C/T:	5.01	0.61			—C —T
R^2:	0.99				
N. target molecules:	270				

Primer set: BN1

molecules of competitor added (x10^{-3}):	6.50	1.30	0.26	T	C
Ratio C/T:	10.9	1.53	0.03		—C —T
R^2:	0.99				
N. target molecules:	784				

Primer set: B48

molecules of competitor added (x10^{-3}):	5.00	1.00	0.20	T	C
Ratio C/T:	3.54	0.37	0.03		—C —T
R^2:	0.99				
N. target molecules:	1586				

Primer set: SE17

molecules of competitor added (x10^{-3}):	1.20	0.60	0.30	T	C
Ratio C/T:	1.50	1.18	0.41		—C —T
R^2:	0.91				
N. target molecules:	698				

Primer set: B13

molecules of competitor added (x10^{-3}):	0.75	0.15	0.03	T	C
Ratio C/T:	16.3	1.16	0.30		—C —T
R^2:	0.98				
N. target molecules:	79				

Primer set: PCO

molecules of competitor added (x10^{-3}):	1.50	0.30	0.06	T	C
Ratio C/T:	10.5	0.87			—C —T
R^2:	0.99				
N. target molecules:	218				

Figure 10. Results of the quantitative-PCR experiments on ssDNA pool from asynchronously growing IMR90 cells with the indicated primer sets. For each primer set, a fixed amount of newly synthesized DNA was co-amplified with the amount of competitor indicated on the top of each gel. PCR products were resolved on PAGE, stained with ethidium bromide and photographed. The intensity of the bands corresponding to the PCR products for the competitor (C) and the target genomic (T) DNAs was assessed by densitometric scanning of each lane. The ratio between the two products (C/T) is linearly correlated with the quantity of competitor added to the reaction; the correlation coefficient (R^2) approaching the value of 1. From the equation of the line fitting the experimental points (Diviacco et al., 1992), the number of target molecules in each case was calculated.

competitors, as indicated at the top of each gel. According to the principles of competitive PCR (Diviacco et al., 1992), the ratio between the competitor and genomic template amplification products (shown at the bottom of each gel) is linearly related to the number of competitor molecules initially added to the reaction. Since the latter amount is known, the concentration of each investigated genomic segment in the sample could be calculated easily from the interpolation of the regression line.

For each genomic segment to be quantified, competitive PCR experiments were firstly carried out with 10-fold scalar dilutions of the corresponding competitor, to roughly estimate the target concentration; more precise quantification was subsequently obtained by using 5-fold or 2-fold competitor dilutions, as shown in Figure 10. Similarly, precise quantification of each competitor was independently obtained by competitive PCR experiments with a fixed amount of total human genomic DNA.

The results obtained for the IMR90 cell line are graphically shown on the upper left panel of Figure 11. The other panels of figure 11 show the results of similar experiments performed on cell lines of different histological derivation. These include the IMR32, SKNBE and SKNMC neuroblastoma, the HL-60 myeloid and the HeLa epithelial cell lines. In all these cells, a clear enrichment could be detected for the genomic fragment corresponding to the B48 marker - the one corresponding to the lamin B2 *ori* region (Giacca et al., 1994) -, scoring 6-10 fold higher than the markers localized
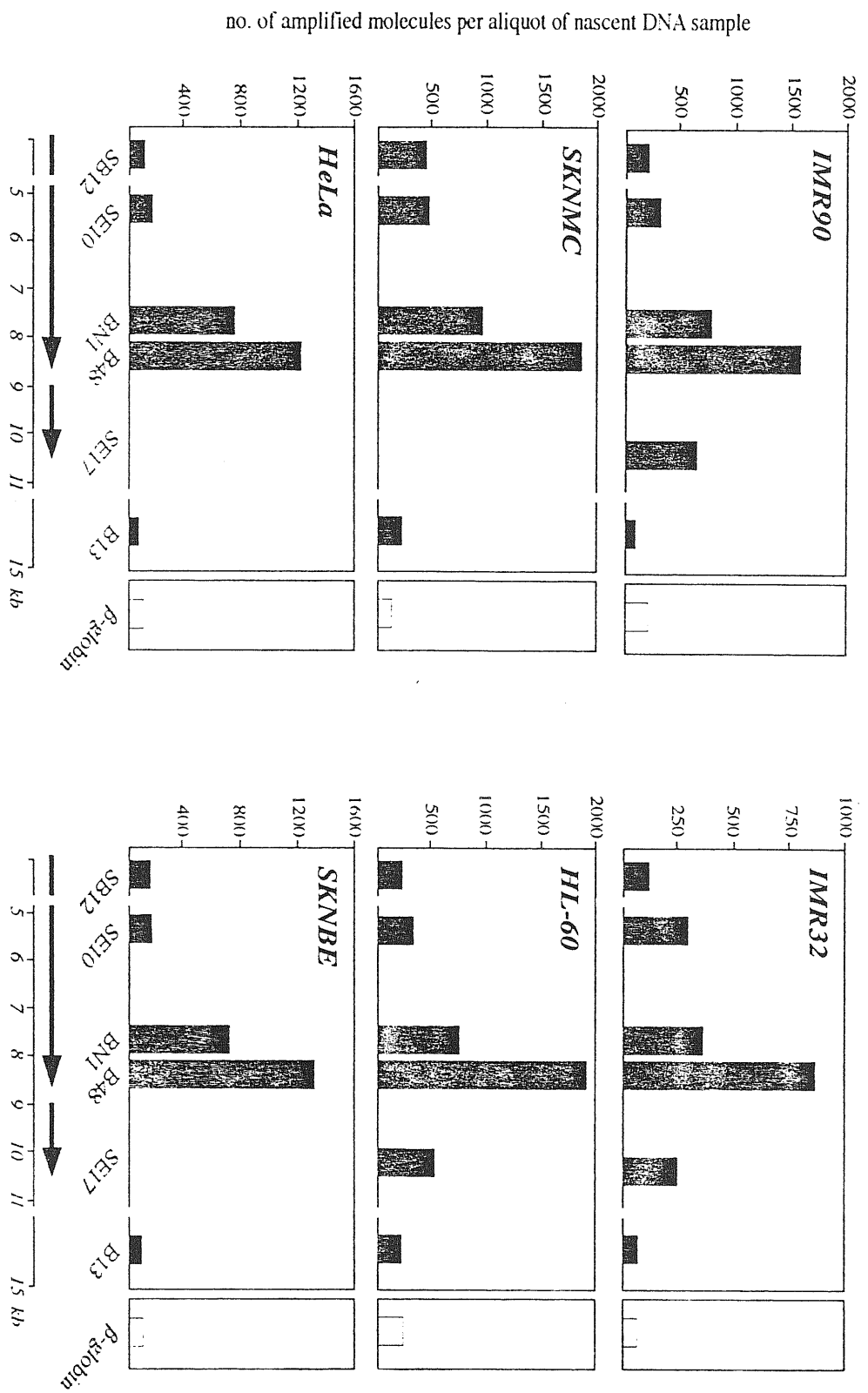
32

no. of amplified molecules per aliquot of nascent DNA sample

Figure 11. Graphical representation of the results obtained by the quantification of different segments in the lamin B2 gene domain by competitive PCR of nascent DNA samples from the indicated cell lines. PCR quantitation results are given in terms of an arbitrary value, i.e. number of primer specific molecules amplified in a fixed volume of the ssDNA pool tested for each cell line. Quantification of the β-globin gene copy number was obtained by the PCO primer set.

>5 kb apart on each side (SB12 and B13), and those of the β-globin gene control region.

This distribution of DNA segment abundance in samples of nascent DNA clearly suggests that the lamin B2 *ori* is active in all the analyzed cell lines.

2.3.3. Absence of origin activity in quiescent primary cells

In addition to the results reported above, which were obtained in established cell lines that replicate at a continuous rate in the absence of inhibitory treatments, the lamin B2 *ori* activity was studied in peripheral blood lymphocytes from a normal individual. These primary cells are largely quiescent in the absence of stimulation. When size-selected single stranded DNA was analyzed from these cells, no *ori* activity could be detected in the lamin B2 region, as expected (Figure 12A). However, when the PCR mapping experiment was performed on nascent DNA samples extracted from the same lymphocytes at five days after stimulation (obtained by the addition of phytohemoagglutinin and interleukin-2), a clear evidence of activation of the lamin B2 *ori* could be observed, peaking at the same region as in the cell lines (Figure 12B). This observation, besides providing an important control to the results obtained in the established cell lines,
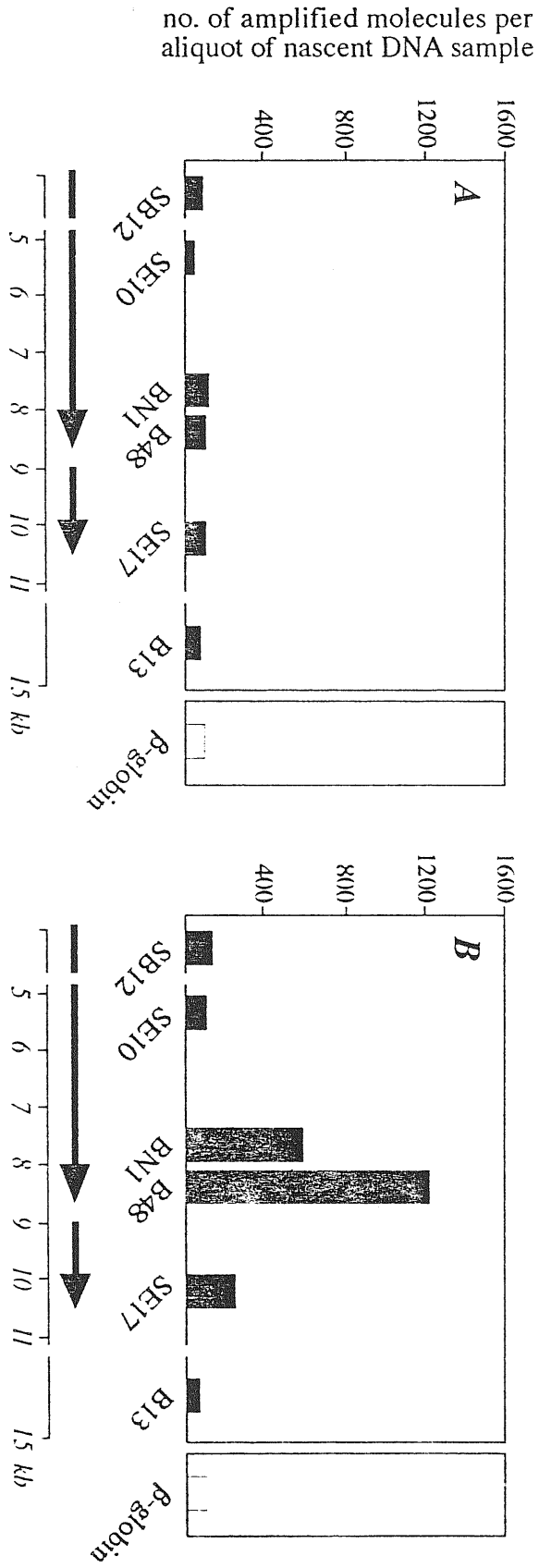
33

no. of amplified molecules per aliquot of nascent DNA sample

Figure 12. Quantitative PCR results for the indicated primer sets in 1 kb ssDNA pools obtained from quiescent (A) and proliferating (B) human peripheral blood lymphocytes from a normal donor.

represents the first observation of origin usage in primary human cells.

2.4. DISCUSSION

The use of BrdUrd pulse-labelling of newly replicated DNA and of the anti-BrdUrd antibodies in the process of template DNA preparation for origin-detection experiments was originally introduced (Vassilev and Russev, 1988; Contreas et al., 1993) with the aim of obtaining a high signal-to-noise ratio. It was seen that the method can be further simplified and that the use of BrdUrd and immunoaffinity chromatography can be omitted. This result can be attributed to careful size selection of the nascent DNA strand pool to be used as well as to the reduction in the background (i.e. pool contamination by aspecifically broken DNA) resulting from the elimination of the BrdUrd-substituted areas in the DNA which are prone to breakage. The value of approximately 3×10^4 different *oris* inferred by our data in human cells agrees with the occurrence of an origin approximately every 100 kb, as originally suggested by the fiber autoradiographic studies on the mammalian DNA replication fork movement (Huberman and Riggs, 1968). This new simplified procedure for origin mapping now provides a simple tool for the study of the process of DNA replication in single copy domains of untreated, asynchronously growing cells. Its extensive application will be suitable for the identification of novel origins in mammalian cell DNA and will

assist in the understanding of the functional organization of the genome.

As far as the lamin B2 *ori* in the chromosome 19p13.3 location is concerned, the results reported here show that this origin is used to initiate DNA replication in a variety of human cells. Of particular interest is the observation that this *ori* is activated also in primary peripheral blood lymphocytes, providing further support to the physiological significance of the observations obtained in the established cell lines. The absence of origin activity in unstimulated, quiescent lymphocytes goes in parallel to our previously reported observation where also HL-60 cells, when differentiated in vitro to a non proliferating state, gave no evidence for lamin B2 *ori* activity (Giacca et al., 1994). Once again, these observations reinforce the validity of the developed method for origin identification.

The universal use of this origin in all the analyzed cells is not surprising. Given the close relationship between transcriptional activity and initiation of DNA replication (DePamphilis, 1993), possibly related to the accessibility of chromosomal domains to proteins involved in both processes, the observation that this origins lies in a constitutively expressed gene domain, coding for a house-keeping protein (Biamonti et al., 1992b) predicts origin usage by all cells. Further support to this notion is the observation that, in synchronized cells, this *ori* fires within the first minutes of S-phase (Biamonti et al., 1992a).

The observation that the same precise and rather narrowly defined region of the chromosome works as an *ori* in very different cell types gives still more weight to the conception that *ori*s in higher organisms are constituted by cis-acting sites of spatially and temporally programmed interactions with trans-acting specific protein factors. A study of the nature of the proteins involved in the definition and activation of the human lamin B2 *ori* has been initiated (Dimitrova et al., 1996).

A NEW HUMAN *ORI*

3.1. INTRODUCTION

Having developed an efficient and fast method to map an origin of DNA replication in complex genomes, we planned to localize new human *oris*. To address this problem, we started to select a well studied short region of human genome (at least 10 kb in size) where the presence of an *ori* could be hypothesized. As explained at the end of this section, one of such well defined interesting sequences is presented by a human DNA region on chromosome 19 that contains a preferred site for integration by adeno-associated viruses.

3.1.1. The life cycle of the adeno-associated virus (AAV)

The AAV belongs to the Parvoviridae family of viruses that infects both vertebrates and invertebrates and is characterized by the small single-stranded DNA genome (Siegl et al., 1985). The virion is icosahederal with a diameter of 20-30 nm, is non-enveloped and contains a DNA genome of about 5 kb. AAV constitutes the dependovirus genus of the family because of its usual requirement for a helper virus coinfection for a productive infection to occur (Figure 13). The virus is widespread and is notable because infection has never been associated with any disease (Blacklow et al., 1968).

AAV

Helper virus co-infection

Integration

Helper virus infection

Rescue
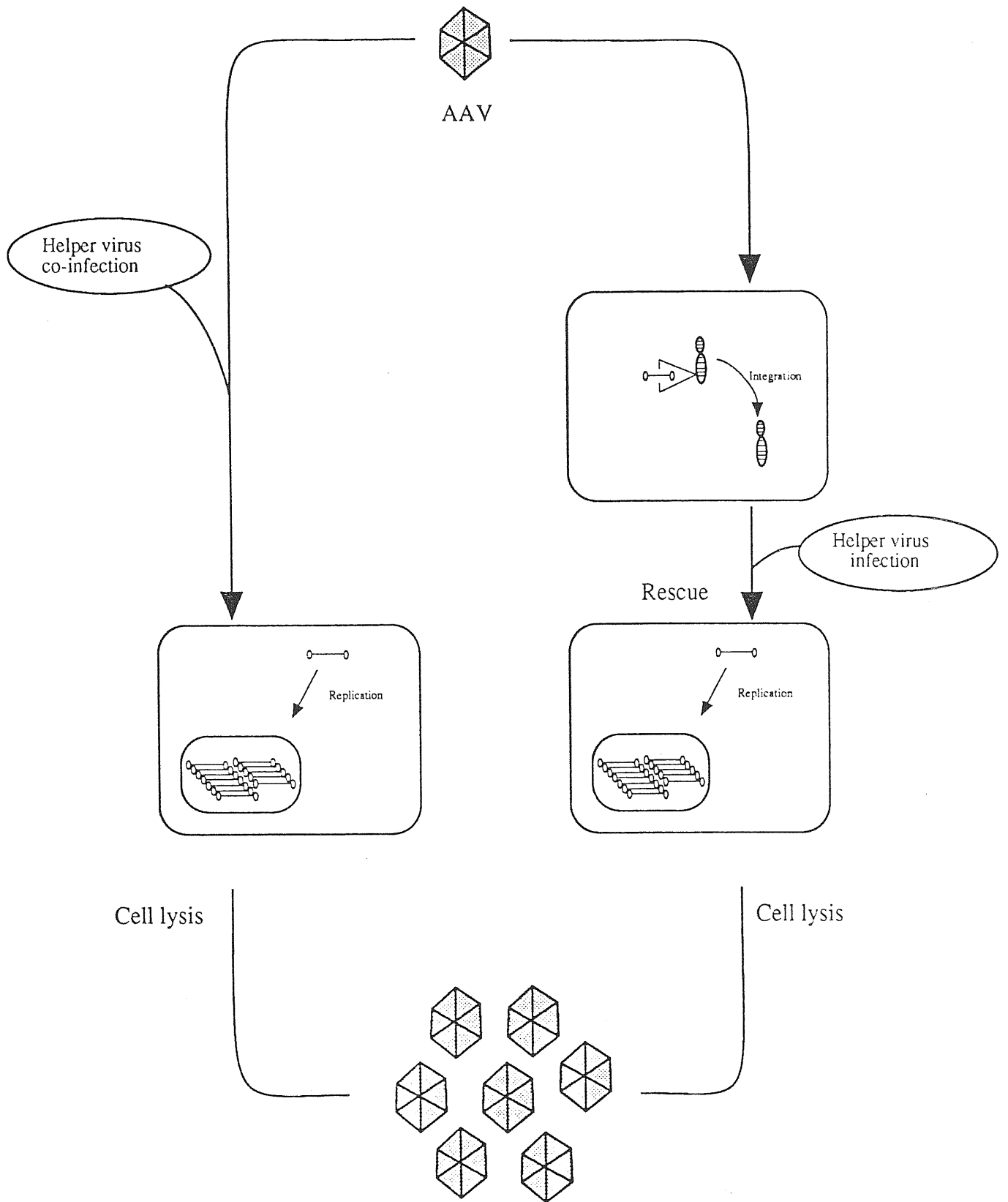
Replication

Replication

Cell lysis

Cell lysis

Figure 13. The life cycle of AAV

The helper functions needed for a productive AAV infection in a cell culture have been best characterized for adenovirus (Ad) coinfection where all the early functions (E1A, E1B, E2A, E4) except E2B and E3 contribute to the helper effect (Richardson and Westphal, 1981; West et al., 1987). Thus, while in the Ad help in AAV productive infection is characterized by those Ad functions that effect gene-expression (both cellular and viral), that of Herpes simplex virus (HSV) is shown to occur by HSV enzymes involved in DNA synthesis (Mishra and Rose, 1990; Weindler and Heilbronn, 1991). This remarkable variability in the functions provided by different helper viruses has challenged the earlier notion of denoting AAV as a "defective" virus. The current hypothesis is that if AAV infects a healthy cell, the genome is programmed to repress the gene expression required for productive infection so that the viral genome can integrate into the host genome to establish a latent infection. Subsequent exposure of the cell to conditions which activate stress response genes also causes activation of the AAV genome, with rescue and replication of the integrated DNA and production of progeny virions. This hypothesis has been supported by repeated observations that the exposure of several cell lines to genotoxic stimuli (e.g. UV irradiation or chemical carcinogens) makes the cells permissive for productive AAV infection in the absence of a helper virus (Yakobson et al., 1987; Yakinoglu et al., 1988).

The 4680 nucleotide sequence of the AAV genome has been determined (Srivastava et al., 1983). It contains two large open

reading frames (orf). The orf in the 5' half encodes four non-structural, regulatory proteins and that in the 3' half encodes three structural proteins, both with overlapping amino acid sequences (Figure 14). The 5' orf is commonly referred to as the *rep* gene because almost every nonsense mutation in the orf inhibits DNA replication (Hermonat et al., 1984; Tratschin et al., 1984; Yang et al., 1992). Transcription with two different promoters at map positions (mp) 5 and 19 produces two kinds of transcript (Rep 78 and 52) each of which in turn gives rise to yet another transcript (Rep 68 and Rep 40) with one splicing event. The two larger Rep proteins have been implicated in the regulation of gene expression (Labow et al., 1986; Beaton et al., 1989) and DNA replication (Labow and Berns, 1988). Functions of the two smaller proteins are less certain and are implicated in the packaging of the linear single strands (Chejanovsky and Carter, 1989). Rep68/78 binds to specific sites on the AAV genome (in the inverted terminal repeat (itr) and in each of the three promoters), makes a site-specific nick in the genome and can function as an ATPase and helicase (Im and Muzyczka, 1990).

There are three viral coat proteins, Vp1-3, which range in size from 65-85 kd (Figure 14). All three are translated from a transcript whose promoter is at mp 40 (Green et al., 1980). There are alternatively spliced mRNAs; Vp1 is translated from the minor species and the other two from the more abundant mRNA. Vp3 accounts for about 80% of the protein mass of the virion and is translated from the first in-frame initiator codon in the major

40

MAP UNITS

20   40   60   80

1000   2000   3000   4000

NUCLEOTIDES

Rep 78

Rep 68
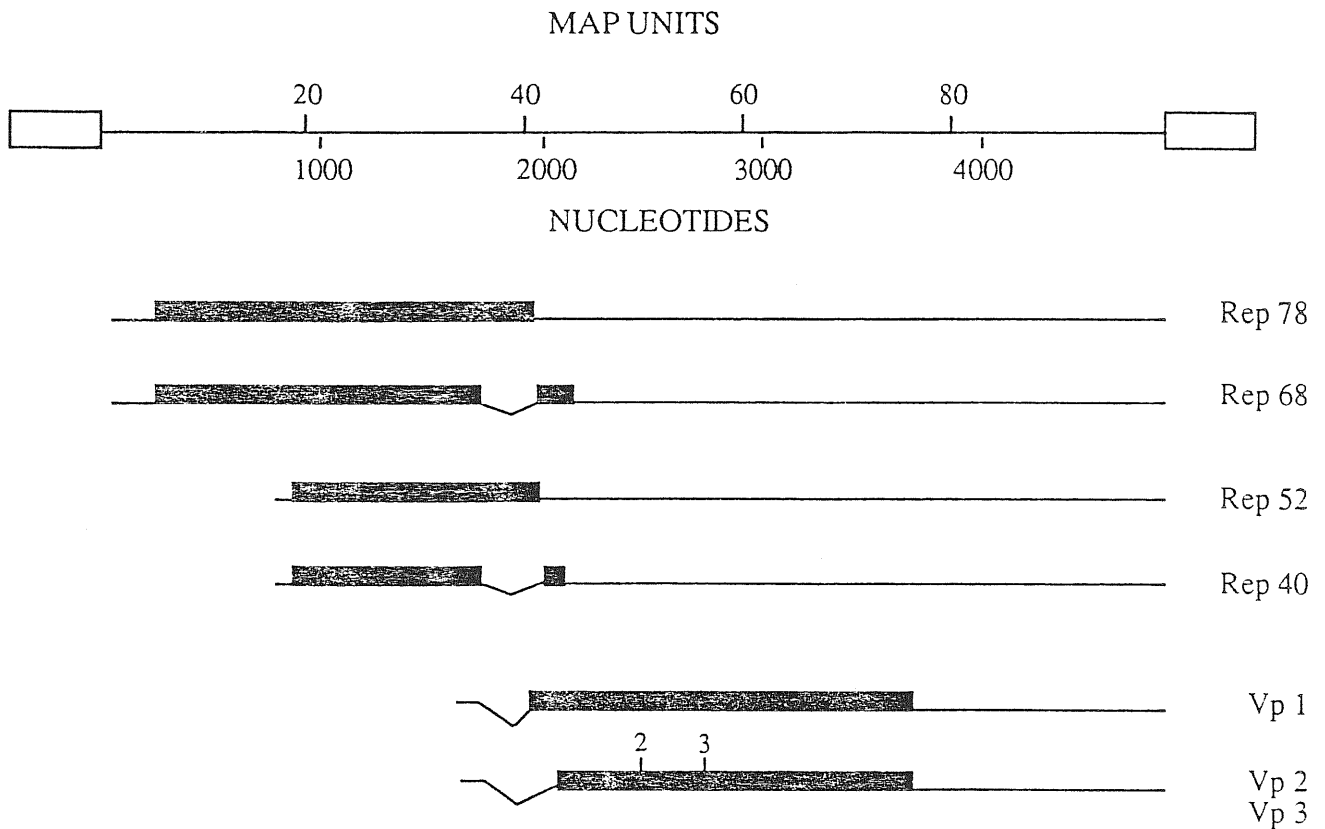
Rep 52

Rep 40

2   3

Vp 1

Vp 2
Vp 3

Figure 14. Genetic map of AAV. The promoter at map position 5 (p5) initiates a transcript which, unspliced, encodes Rep 78 and spliced, encodes Rep 68. p19 at map position 19 initiates a transcript that is either translated intact, yielding Rep 52, or spliced, to yield Rep 40. Both products are in-frame with the larger Rep proteins. p40 initiates all three capsid mRNA species. Vp1 is produced by an alternative splicing event used for the Rep proteins (Rep 68 and 40). Utilizing the same mRNA as Vp3, Vp2 results from the utilization of an alternative ACG start codon.

RNA species while Vp2 is initiated from an upstream, in-frame ACG codon (Beccera et al., 1988), which presumably accounts for its lower abundance with respect to the other cap proteins.

A remarkably special feature of the AAV genome is the presence of palindromic, inverted terminal repeats (*itrs*). The *itr* is 145 nucleotides long of which the first 125 constitute an overall palindrome interrupted by two smaller 21 nucleotide palindromes immediately flanking the axis of symmetry of the overall palindrome. When folded on itself to maximize potential base pairing, the overall palindrome forms a T-shaped structure in which there are only seven unpaired bases. Genetic experiments suggest that for some of the steps in the AAV life cycle, the potential conformation of the *itr* may be more important than the actual sequence. Like the Rep proteins, the *itr* is involved in every step of the AAV life cycle: regulation of gene expression, DNA replication, site specific integration and rescue of the genome from the integrated state (Cheung et al., 1980; Samulski et al., 1991; Kotin et al., 1992).

3.1.2. Site specific integration of AAV to establish latency

AAV latency was discovered by Hoggan et al. in 1972 when they noticed that 20% of primary African green monkey kidney cell lots and 1-2% of primary human embryonic kidney cell lots would

produce AAV upon infection of the cells with purified Ad. In the absence of a helper virus AAV readily establishes a latent infection (Figure 13) in continuous cultures of human cell lines, where the viral DNA has been shown to be integrated into the cellular DNA (Handa et al., 1977; Cheung et al., 1980) at a preferred site mapped to 19q 13.3-ter position (Kotin et al., 1990; Kotin et al., 1991; Samulski et al., 1991). Such a site specific integration of a foreign DNA into the host DNA by non-homologous recombination is unique to AAV and is of extreme interest for its potential advantage in gene therapy (Berns and Linden, 1995).

3.1.2a The AAVS1 sequence

A 4 kb DNA fragment containing the human DNA segment involved in site-specific integration of AAV has been sequenced and characterized (Kotin et al., 1992). This pre-integration site sequence, also referred to as the AAVS1 sequence, shows many interesting features (Figure 15A). In addition to a very high G+C content (65% overall and 82% in the first 900 bp stretch), there is an abundance also of repetitive elements. Most striking is an array of 10 basic repeat units from nt 3660 to 4021. This array was identified as a minisatellite sequence that appears at about 60 sites in the human genome, all on the q arm of the chromosome 19 (Das et al., 1987; Kotin et al., 1991). A number of sequence motifs reported to function as cis-acting signals are also
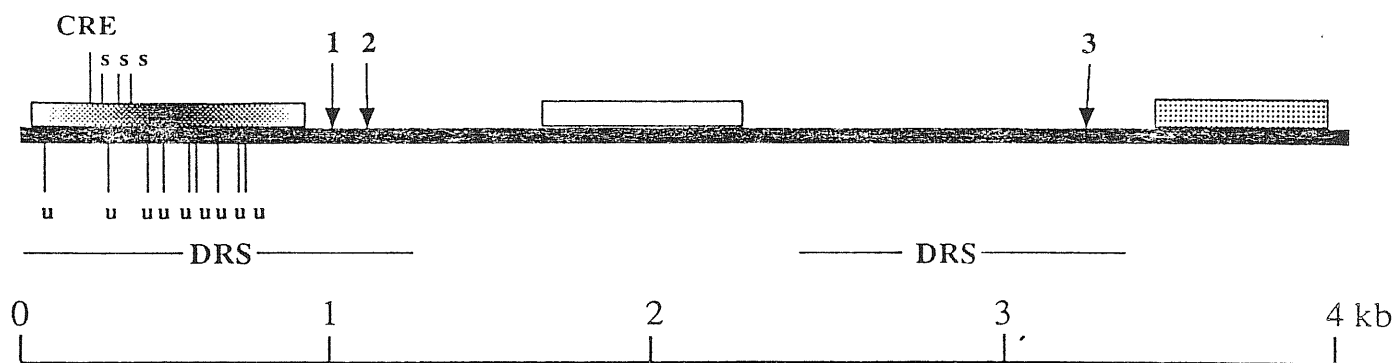
(A)
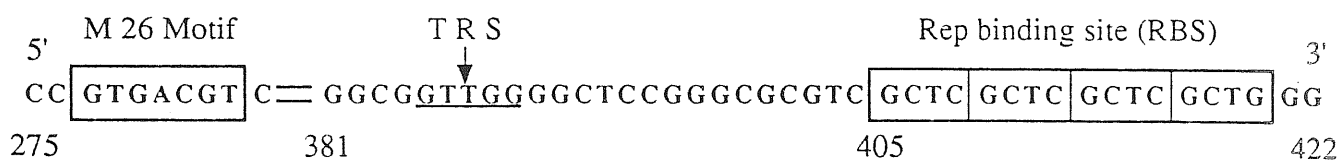
(B)

Figure 15. (A) A graphical summary of the AAVS1 is shown with respect to a nucleotide-scale in kilobases. The radially shaded box indicates the CpG island. s: Sp1 binding site; u: UBF1 binding site; CRE: cyclic AMP response element. The open box indicates the partial cDNA clone. The hatched box indicates the position of the chromosome 19q specific minisatellites. 1-3 indicate locations of the sites involved in the recombination event. DRS are regions with high frequency of short direct repeats. (B) A part of the 510 bp AAVS1 with the three putative signal-motifs responsible for directing site-specific AAV integration and resulting DNA sequence rearrangements is shown. The nucleotide positions described are with respect to the AAVS1 sequence. TRS is the putative terminal resolution site where Rep protein could nick.

present. The first 900 nt, having an 82% G+C content fit the parameters described for CpG islands (Bird, 1987). CpG islands have been found upstream of several housekeeping genes and are mainly associated with TATA-less promoters. A sequence corresponding to a cyclic AMP response element (CRE), TGACGTCA (Montminy et al., 1986), is between nt 279 and 286 in the CpG island. The sequence for the upstream binding factor (UBF 1), CGGCC, or its reverse complement, GGCCG (Bell et al., 1988), is located at nine positions within the CpG island at nt 87, 347, 455, 497, 585, 587, 603, 689 and 690. Three additional putative UBF 1 binding sites are at nt 2130, 2782 and 3441.

A partial cDNA clone was obtained from a cDNA library, spanning nt 1620 to 2318 of the AAVS1 sequence (Kotin et al., 1992). Since no polyadenylation signal is found in the sequence, it is likely that if the sequence is transcribed for expression, the mRNA would extend beyond the determined sequence. The partial cDNA product is yet unidentified.

To address whether sequence motifs previously reported as being associated with recombination are present at or near the sites of AAV integration and associated rearrangements, comparisons using 100 nt stretches of the AAVS1 sequence as probes with available databases have been made. The only sequence with >70% homology to an AAVS1 probe is within the short terminal repeat region of Herpes Simplex Virus 1. Sequences consisting of five or more As and/or Ts have been described as preferred sites

of non-homologous recombination in mammalian cells (Konopka, 1988). The pentanucleotide sequence TTTAA is found at the recombination site at nt 1028-1032. The sequence TTTTA is located at nt 3370-3374 within 10 nt of the recombination site 3360-3363. The involvement of topoisomerase I (Topo I) in non-homologous recombination has been described (see Bullock et al., 1985). A highly preferred Topo I cleavage site, CTT (Konopka, 1988), is on both substrates (AAVS1 and AAV) at the viral-cellular junction involving cellular nt 1027-1030 and AAV nt 54-51. The Topo I cleavage site CTC is present at the cellular-viral junction involving cellular nt 1144-1146 and AAV nt 157-155; and at the junction between cellular nt 3360-3362 and 1144-1146.

3.1.2b DNA rearrangements associated with viral DNA integration

Extensive DNA rearrangements in the integrated viral sequences, involving inversions and translocations have been described (Kotin and Berns, 1989). Comparisons of the AAVS1 pre-integration sequence to the flanking cellular sequences found in one latently infected cell line characterized (Kotin and Berns, 1989), demonstrated that the flanking cellular sequences had also undergone rearrangements. The right cellular-viral junction corresponded to the cellular AAVS1 sequence from nt 1620 to nt 1144-1147 and the AAV sequence nt 154-157, the exact point of recombination being in doubt because of a four base identity in

both genomes at the site of recombination. The left junction was, instead, generated by at least two recombinational events; one involving only cellular sequences and a second involving both cellular and viral ones. The latter recombination involved nt 1026-1030 of AAVS1 and nt 51-55 of the AAV terminal repeat. In this instance there is an overlap of five homologous bases between the two DNAs at the site of recombination. The cell-cell recombination involved nt 3360-3363 and nt 1144-1141. As a consequence of the recombination and inversion involving only cellular sequences, interestingly, a structure resembling one cross-arm of the sort found in AAV terminal repeat is generated.

3.1.2c A cellular DNA sequence is responsible for site-specific AAV integration

The site-specific integration of AAV into the q arm of chromosome 19 in established human cell lines implies a specific recognition signal. In a recent study (Giraud et al., 1994) such signals have been localized in a 510 bp fragment at the very start of the 5' end of the AAVS1 sequence. Interestingly, this fragment lies around five to seven hundred bp upstream from the virus-cell recombinant junctions mapped in the AAVS1 and described above. A recent discovery of particular interest is that the 5' end of AAVS1 DNA contains a binding site for the AAV rep gene products, Rep 68 and Rep 78. This binding site is located between nt 353-468 within the 510 bp region that directs recombination

45

with AAV (Figure 15B). The observation that *rep* gene products may be involved in the site-specific AAV integration has been long indicated by the lack of sequence-specificity of integration of the recombinant AAV-based vectors not expressing Rep. Moreover, in cell free reactions, Rep68/78 forms a complex between the AAVS1 DNA segment bordered by nt 353-468 and the AAV genome (Weitzman et al., 1994). A putative terminal resolution site (TRS), i.e. the site where Rep 68/78 nicks DNA in AAV, has been found also in the 510 bp fragment, 14 bp upstream from the Rep-binding site (RBS). If Rep were able to nick DNA at this site then this might serve as a step to initiate recombination. It is likely that the Rep binding motif might be present at many locations in the cellular genome which implies that this might be just one of the necessary signals needed to achieve the site-specificity of AAV DNA integration. In addition to the Rep binding site an M26 (a yeast enhancer) motif is also present upstream from the putative TRS and forms part of the 510 bp fragment essential to get site-specificity in AAV integration.

Recently (Linden et al., 1996), the recombination signal has been further narrowed down to a 33 nucleotide sequence containing the cellular RBS separated by separated by an eight nucleotide spacer from the cellular TRS . Interestingly, this recombination signal is separate from the signal governing sequence rearrangements in AAVS1 upon integration of AAV. The M26 motif was seen to co-segregate with the rearrangements in the

46

cellular DNA in the study. However, since the M26 motif-associated rearrangements are also shown to be AAV-independent, further studies are on to define the appropriate signal(s) .

The AAVS1 lies in a region of chromosome 19 that has been shown to be associated with myotonic dystrophy (Aslanidis et al., 1992) and a common breakpoint in B-cell leukemia (McKeithan et al., 1987; Korneluk et al., 1989), and undergoes sister chromatid exchange at an unusually high frequency. These observations, together with all other interesting characteristics of AAVS1 sequence described above, led us to screen the region for a possible origin of DNA replication (*ori*) in or near it. The following sections deal with the work done towards this aim.

3.2 MATERIALS AND METHODS

All the procedures for cell cultures, DNA extraction and PCR quantifications were same as already described in the section 2.3. D6-7374 cells, a kind gift from Prof. Berns, were cultured in Dulbecco's modified Eagle medium with 12% fetal calf serum. Both media were supplemented with 2 mM glutamine and 50 µg/ml gentamicin. The nascent DNA strand pools of peripheral blood lymphocytes (both from activated and quiescent cells) were same as those described in section 2.2 for quantification experiments
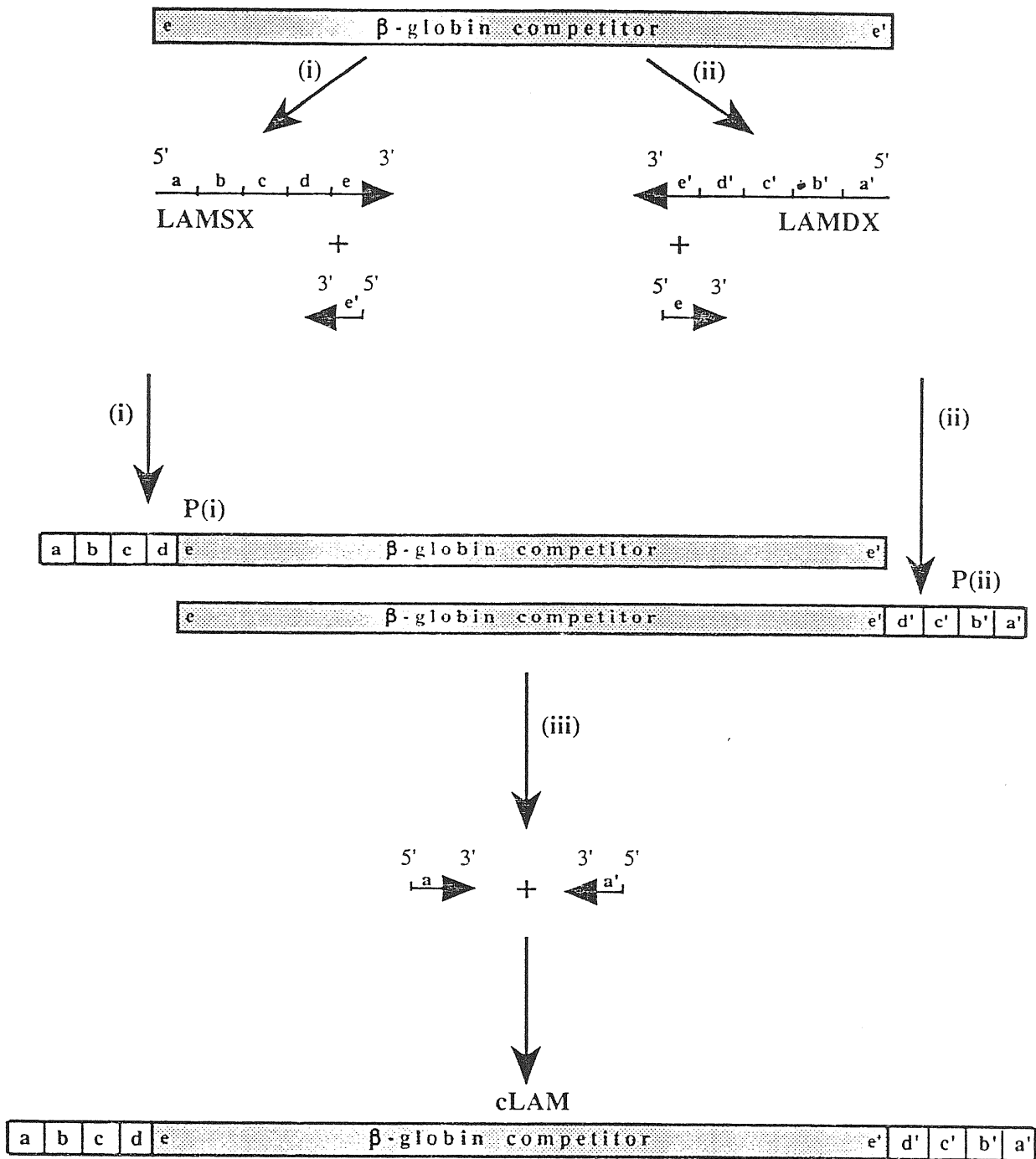
Figure 16. Schematic representation of the construction of cLAM by PCR. PCR reaction (i) using the primers LAMSX and e' gives rise to the partial product P(i) whereas the reaction (ii) using the primers e and LAMDX gives rise to the partial product P(ii). P(i) and P(ii) are run on a polyacrylamide gel, eluted out in a buffer and used in a PCR reaction (iii) where primers a and a' amplify the partials to the full length cLAM. The long competitor contains the β-globin competitor (i.e. the β-globin template with a 20 bp insertion in the middle) flanked by the right (a', b', c' and d') and left (a, b, c, d) primers that represent the right or (DX) and left (or SX) primers of the B48, B13, SE10 and SB12 primer sets respectively. cLAM is, therefore, amplified by B48, B13, SE10, SB12 and the β-globin (PCO3-4) primer sets.

with the lamin B2 origin. The details of the primer sets used in the study are described in Table VI. All the primer sets from SK1 to SK7 amplified the genomic DNA at the standard PCR profile i.e. denaturation at 94°C, annealing at 55°C and extension at 72°C, all for 30 seconds each and for a total of 50 cycles. FISH on the D6-7374 cells was performed using total AAV genome as a probe following the standard procedure (Lichter et al., 1990) at CNR Pavia with the help of Dr. Elena Raimondi..

Table VI. Details of the oligonucleotides used in the study

Region	Template Size (bp)	Primer name	Sequence	Position (nt.)
SK1	213	SK1DX	5'-AAATGGAGTCCATTAGCAGAA-3'	235-215
		SK1SX	5'-AGTCTGCTATTCATCCCCTTTA-3'	23-44
SK2	188	SK2DX	5'-TGCTCCGAAAGAGCATCCTT-3'	1572-1552
		SK2SX	5'-TGCTTTCTCTGACCAGCATT-3'	1385-1404
SK3	259	SK3DX	5'-TTAAGGAAAGAAGGATGGAT-3'	2825-2806
		SK3SX	5'-ACAAAATCAGAATAAGTTG-3'	2567-2585
SK4	193	SK4DX	5'-CTCACCAAGTGGTTCATAAA-3'	3566-3547
		SK4SX	5'-TTTAGCGCTGAAACCCTCAGT-3'	3364-3384
SK5	159	SK5DX	5'-GAACGAAGCCGTGGGC-3'	4730-4712
		SK5SX	5'-CTCAGTGAACTGGAGTGT-3'	4572-4589
		C5N2DX	5'-tail1-GCACACCTGTGTGCCTGGG-3'	4621-4602
		C5NSX	5'-tail2-AGCAGCTGTCTCACCCCTC-3'	4622-4640
SK6	247	SK6DX	5'-ACCCCAGCCCACCCCAATG-3'	3250-3233
		SK6SX	5'-TTCCTGGCTCTGCTCTTCA-3'	3004-3022
		C6NDX	5'-tail1-AAAGGTCAGCCTGGTAGAC-3'	3157-3138
		C6NSX	5'-tail2-GGTGATTCTCCTCCAACTG-3'	3158-3177
SK7	128	SK7DX	5'-GTACAGGCATCCCTGTGAAA-3'	4067-4048
		SK7SX	5'-TGGACTCCTTCATCTGA-3'	3940-3956
		C7DX	5'-tail1-CCCCTCCCCATTCAACCC-3'	4002-3985
		C7SX	5'-tail2-TTGGGCCTGGACTCTGGA-3'	4003-4020
BIG competitor	293 (with SK1)	BIGDX	5'-SK1DX-SK4DX-SK2DX-SK3DX-PCO4-3'	
		BIGSX	5'-SK1SX-SK4SX-SK2SX-SK3SX-PCO3-3'	

Note: Tail1 = 5'-ACCTGCAGGGATCCGTCGAC-3'; Tail2 = 5'-GTCGACGGATCCCTGCAGGT-3'
SK5 primer set lies outside the AAVS1 and was chosen from the sequence sent kindly by Dr. R. M. Linden.
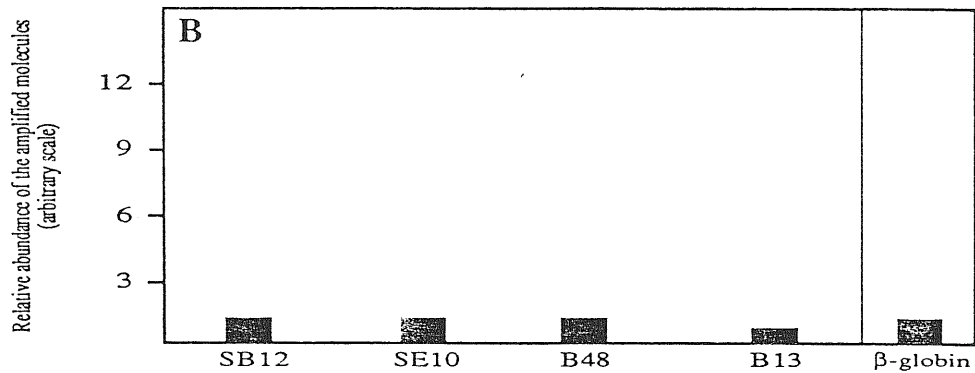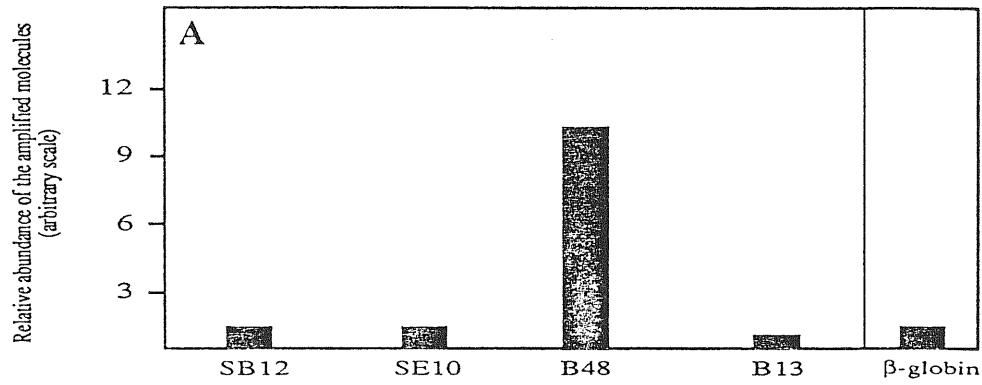
Figure 17. Graphical representation of quantitative PCR results with cLAM on ssDNA from activated (A) and quiescent (B) peripheral blood lymphocytes from a normal donor. Primer set names are plotted on the X-axis while the relative abundance of the segment amplified by each primer set is shown on the Y-axis in terms of an arbitrary scale i.e. the number of specific molecules amplified by a given primer set in a 10 μl aliquot of the sample DNA pool.

3.3. RESULTS

3.3.1. Further simplification of the ori-mapping procedure

In the *ori*-mapping method used in our group the final quantification of the abundance of various markers spanning the region of interest is performed by competitive PCR of the nascent DNA pool. An important step in this quantification is the exact quantification of the concentration of the different competitor preparations. This is obtained by their quantification against a fixed amount of total human genomic DNA. To simplify this procedure, a common competitor molecule (cLAM) which could be amplified by all the primers of interest, was constructed and tested on the lamin B2 origin. As shown in figure 16, this competitor contains the forward and reverse primers of all the primer-sets of interest, attached in a head to tail fashion so as to obtain amplification products of sizes different than those of the cellular DNA. Figure 17 shows the results obtained using cLAM for *ori*-localization in activated and quiescent lymphocytes. Usage of such a big competitor molecule (having a sequence composition that is widely different from that of the template) produced the same results as were obtained by the usual method where separate competitor molecules were used for each primer set used in quantification experiments (Figure 12).
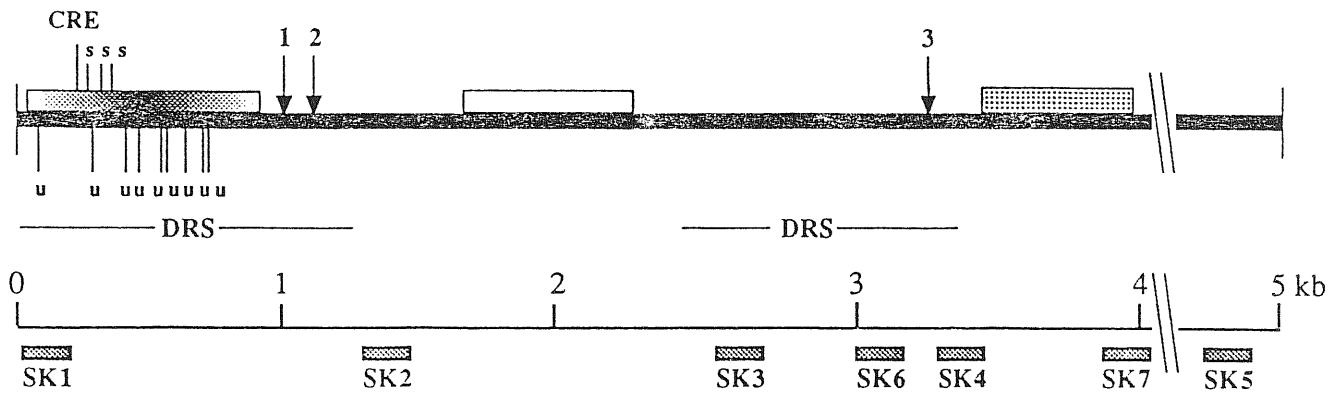
49

Figure 18. The locations of the various DNA segments amplified by the primer sets used in the study is shown (gray boxes) with respect to the AAVS1 sequence. The primer set SK5 lies outside the AAVS1 clone and is thus separated by a double-line break from it. All the symbols on the AAVS1 sequence are same as in figure 15 (A).
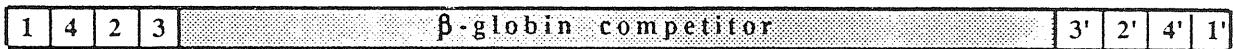

Figure 19. Schematic representation of the big competitor (cBIG) used in the study. The 293 bp long competitor contains the β-globin competitor (i.e. the β-globin template with a 20 bp insertion in the middle) flanked by the right (1', 2', 3' and 4') and left (1, 2, 3, 4) primers of the primer sets SK1-4 respectively in the order shown in the figure. cBIG molecule is, therefore, amplified by SK1-4 as well as β-globin primer sets.

3.3.2. Presence of an *ori* in the AAVS1

Several primer sets were constructed spanning the AAVS1 region (Figure 18). Except for the primer set SK5 that lies around 1 kb downstream of the end of the AAVS1, all other primer sets lie within the AAVS1. The primer-sets SK1-4 together with the control primer set PCO 3-4 (β-globin) amplify the common big competitor cBIG (Figure 19). For the primer sets SK5-7 separate competitor molecules were made with the usual 20 bp insertion described earlier. Table VI shows the details of the primers used in the study.

Quantitative PCR was performed using the above-mentioned primer sets and competitors on newly synthesized DNA pools of around 1 kb fragments obtained from proliferating (activated) and quiescent human peripheral blood lymphocytes. In the activated lymphocytes a 40 fold excess was seen in the abundance of the amplification product of primer set SK4 as compared to that of the control set PCO3 and 4 and the other primer sets (SK1, 2 and 5) located at least a kilobase away on either side of SK4 region (Figure 20A), indicating the presence of an active *ori*. The clear-cut absence of this abundance in the non-proliferating lymphocytes (Figure 20B) confirmed the presence of an *ori* at/near the primer set SK4.
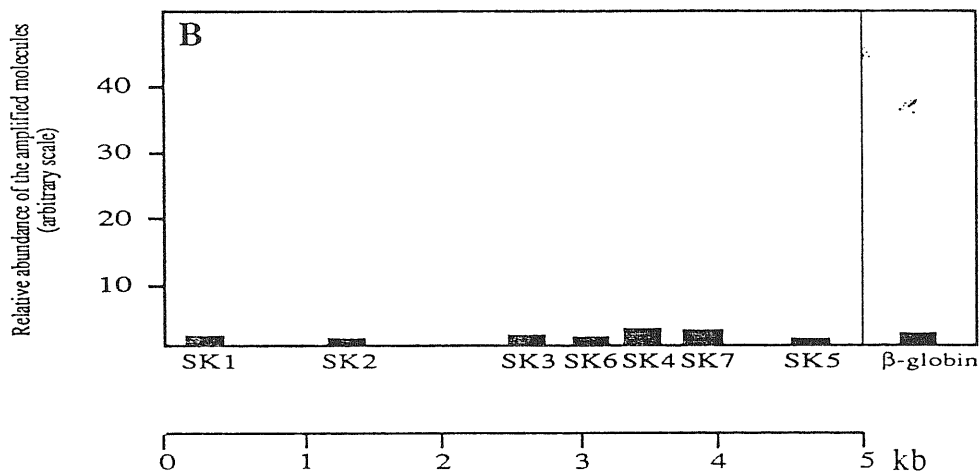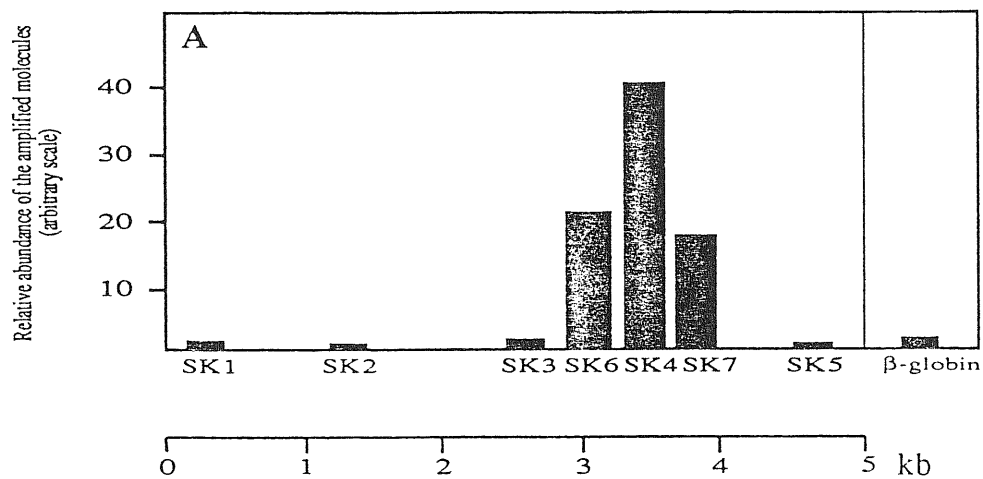
Figure 20. Graphical representation of quantitative PCR results on ssDNA from activated (A) and quiescent (B) peripheral blood lymphocytes from a normal donor. Primer set names and their relative distance in kilobases are represented on the X-axis while the relative abundance of the segment amplified by each primer set is shown on the Y-axis in terms of an arbitrary scale that represents the number of specific molecules amplified by a given primer set in a 10 µl aliquot of the sample DNA pool.

3.3.3. The SK4 *ori* in Detroit 6-7374 cell line

Since this new *ori* lies in the AAVS1 sequence which has been shown to undergo rearrangements upon AAV integration, we wondered whether this origin-activity could be seen also after the AAV genome had been integrated into the host. For this reason, an established human fibroblast cell line (D6-7374) with a latent AAV integrated in the AAVS1 was procured (Kotin et al., 1991). The presence of AAV on the q arm of chromosome 19 was ascertained by performing a Fluorescence In Situ Hybridization (FISH) test using the whole AAV genome as a probe (Figure 21) in accordance with what was already described for this cell line (Kotin et al., 1991). Newly synthesized DNA from this cell line was then isolated and purified according to the modified protocol described earlier (Figure 9B), and submitted to the quantitative PCR test. The results are shown in Figure 22A. Amplification could be obtained with all the primer sets (SK1-7) except for the set SK3. All the primer set amplification products obtained were quite low in their abundance and at identical levels. Primer set SK3 did not amplify the D6-7374 template though it could amplify the respective competitor molecule as well as the template from a fibroblast cell line without AAV viz. IMR90 (Figure 22B) indicating that some rearrangements in the template sequence had taken place in these cells. The detection of *SK4 ori* activity in IMR 90 cells (Figure 22B) serves also as an important positive control since this cell line represents fibroblasts that do not contain latent AAV.
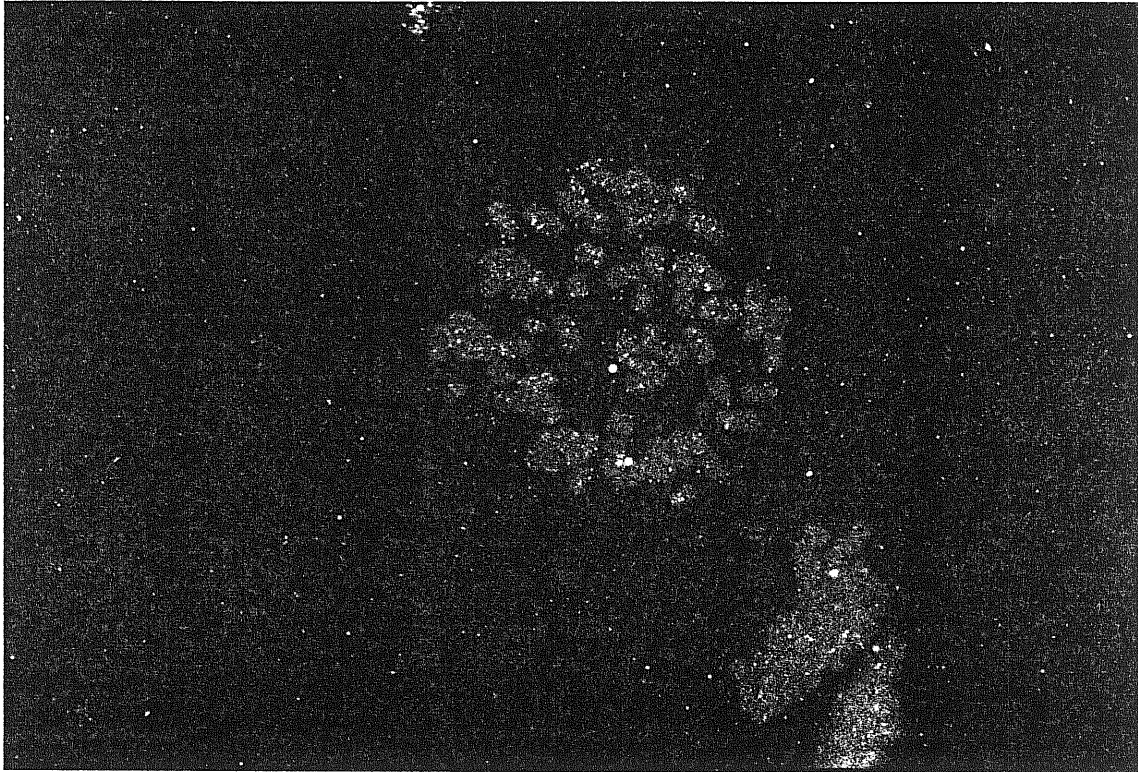
Figure 21. Flourescence In Situ Hybridization (FISH) results on metaphase chromosomal spreads of D6 7374 cells using whole AAV genome as a probe. Chromosome 19 showing the expected signal has an altered morphology that is inherent to the D6 7374 cell line (Kotin et al., 1991).
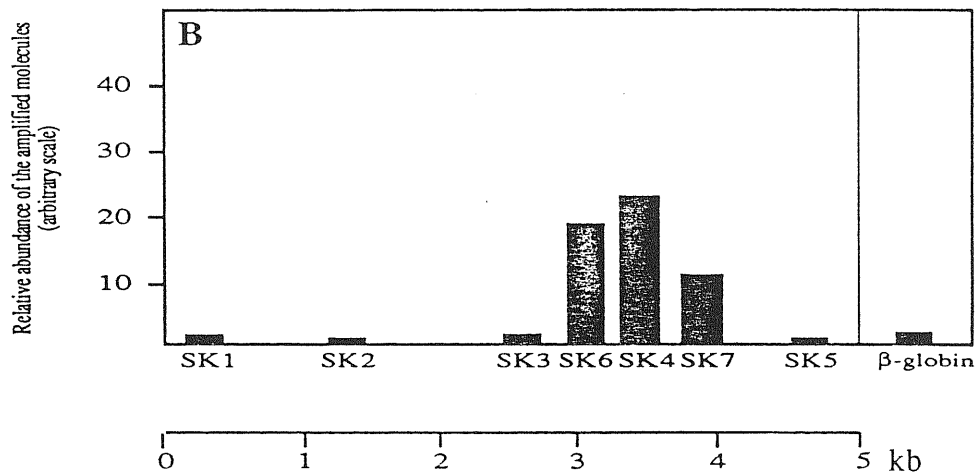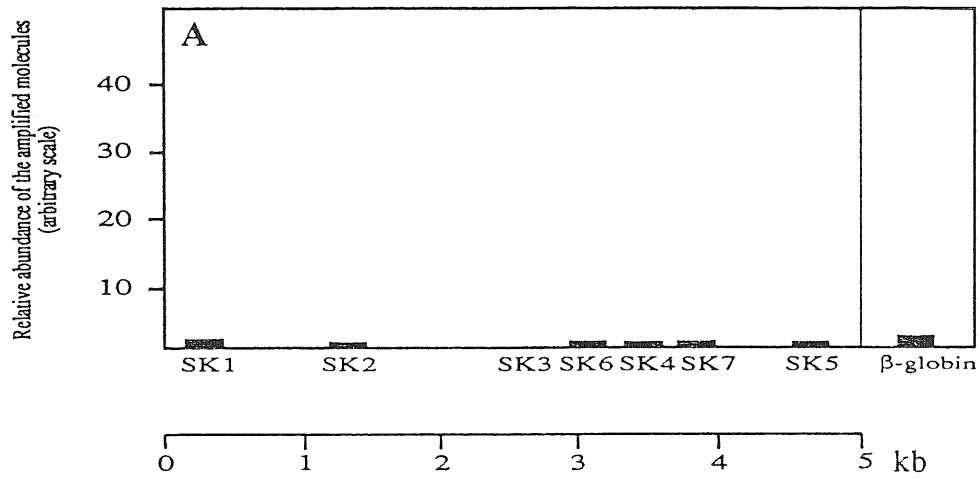
Figure 22. Graphical representation of quantitative PCR results on ssDNA from D6 7374 (A) and IMR 90 (B) cell lines. Primer set names and their relative distance in kilobases are represented on the X-axis while the relative abundance of the segment amplified by each primer set is shown on the Y-axis in terms of an arbitrary scale that represents the number of specific molecules amplified by a given primer set in a 10 µl aliquot of the sample DNA pool. SK3 primer set failed to function in the D6 7374 cell line.

3.4. DISCUSSION

The novel *ori*-mapping method based on the use of competitive PCR to map markers on nascent DNA fragments from growing cells, has been simplified further during the course of the work performed in this thesis. In this part of the report, the use of a "common" competitor molecule for different sets of primers was introduced in order to make the quantification-procedure more convenient since the use of such a molecule in competitive PCR experiments directly yields the relative abundance of the template without the need to know its absolute amounts. The use of such a competitor, that is quite different from the cellular template in its size as well as in the nucleotide sequence, does not affect its amplifiability in a PCR. This is shown by the identical results obtained in quantification experiments done on the lamin B2 *ori* using either separate competitor molecules for each primer set (Figure 12) or the common competitor molecule cLAM (Figure 17). This indicates that as long as the length of the competitor molecule is not too different from that of the target genomic DNA and it displays the required primer-binding sites, there is no difference in the amplification efficiency between the two species.

A novel human *ori*, the SK4 *ori*, is localized in growing primary cells as well as in established cell lines (peripheral blood lymphocytes and IMR90 fibroblasts, respectively) within the AAVS1 region on chromosome 19, that contains the preferred site for AAV integration. The absence of SK4 *ori* activity in quiescent

52

peripheral blood lymphocytes is an obvious but important control, that not only confirmed the presence of the SK4 *ori* but also reinforced the validity and high efficiency of our *ori*-mapping method. The region containing the SK4 *ori* has been delineated to a 0.8 kb fragment (between primer sets SK6 and SK7). The SK4 *ori* lies upstream of a chromosome 19-specific minisatellite (Figure 19). The evidence of a nearby transcription event represented by the partial cDNA clone identified at around 2 kb upstream from the SK4 site (Kotin et al., 1992), once again gives weight to the long founded notion of a close relationship between transcription and replication activities (see DePamphilis, 1993). Of interest is also the observation that the SK4 *ori* lies very close to and downstream of the AAVS1 site (Figure 18; site 3) shown to take part in cell-cell recombination upon AAV integration into the host cell. The activity of this *ori* in D6-7374 cells, where AAV is stably integrated at the preferred site, was absent. Though the exact nature of the rearrangements occurring upon AAV integration is yet unknown, this observation could indicate that the viral integration could interfere with *ori*-function. The D6-7374 cells could thus represent the only case known till date where an *ori* activity is abolished by a naturally occurring sequence rearrangement in DNA; though an analogous yet different case is already represented by the β-like globin locus in hemoglobin Lepore DNA where the naturally occurring deletion of DNA produces a genetic *knock-out* of a replicator (Kitsberg et al., 1993).

Another remarkable observation is the difference in the extent of "signal" detected in the case of the two human origins described, viz. the lamin B2 and the SK4 *oris*. While the B48 marker that detects the lamin B2 *ori* shows a 6-10 fold enrichment (Figure 11) in its relative abundance (signal) over that of the background (far off markers like B13); the SK4 "signal" is stronger, going from a 20-fold excess in the case of IMR90 (Figure 22B) cells to 40-fold excess in the case of activated peripheral blood lymphocytes (Figure 20A). This observation tends to indicate that there could exist differences in the frequency with which a particular *ori* is used in a cell.

Though hardly anything is apparently common between the lamin B2 and the AAVS1 sequences, the latter being extraordinarily GC rich, a few small stretches of a few nucleotides in length are visibly similar in both the regions. Interestingly, they lie 5' to the SK4 segment and represent hypersensitive sites in the lamin B2 *ori* region lying near the *ori*-specific footprint (the OBP) obtained by in-vivo footprinting studies in the region (Dimitrova et al., 1996). Obviously, further studies are necessary in order to gain some insight into the significance, if any, of these similarities between the two *oris*.

In conclusion, the lamin B2 and SK4 *oris* once again confirm the notion of the presence of fixed *ori*-sites in metazoan genomes, both *oris* being localized in less than 1 kb stretches. In the case of the SK4 *ori*, however, the exact nature of the rearrangements

following AAV integration in the region needs to be ascertained in order to confirm the non-functionality of the *ori* due to simple sequence rearrangements in DNA and not to any large translocations. Such studies, together with in-vivo footprinting to look for the trans-acting factor(s) involved in SK4 *ori* function, are now in progress.

Acknowledgments

REFERENCES

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

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

Aslanidis, C., Jansen. G., Amemiya, C., Shutler, G., Mahadevan. M.. Tsilfidis, C., Chen, C.. Alleman, J., Wormskamp, N. G. M., Vooijs. M., Buxton, J., Johnson. K.. Smeets, H. J. M., Lennon, G. G., Carrano. A. V., Korneluk, R. G.. Wieringa, B. and de Jong, P. J.: *Cloning of the essential myotonic dystrophy region and mapping of the putative defect.* Nature (1992). 355, 548-551.

Beaton, A., Palumbo, P. and Berns, K. I.: *Expression from the adeno-associated virus p5 and p19 promoters is negatively regulated in trans by rep protein.* J. Virol. (1989), 63, 4450-4454.

Beccera, S. P., Koczot. F., Fabisch, P. and Rose, J. A.: *Synthesis of adeno-associated virus structural proteins requires both alternative splicing and alternative initiations from a single transcript.* J. Virol. (1988), 62, 2745-2754.

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

Bell, S. P.: *Eukaryotic replicators and associated protein complexes.* Curr. Opin. Genet. Dev. (1995), 5, 162-167.

Benard, M., Lagnel, C., Pallotta, D. and Pierron, G.: *Mapping of a replication origin within the promoter region of two unlinked.*

abundantly transcribed actin genes of Physarium polycephalum. Mol. Cell. Biol. (1996), 16, 968-976.

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

Berns, K. I. and Linden, R. M.: *The cryptic life style of adeno-associated virus.* Bioessays (1995), 17, 237-245.

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

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

Bird, A.: *CpG islands as gene markers in the vertebrate nucleus.* Trends Genet. (1987), 3, 342-347.

Blacklow, N. R., Hoggan, M. D., Kapikian, A. Z., Austin, J. B. and Rowe, W. P.: *Epidemiology of adeno-associated virus infection in a nursery population.* Am. J. Epidemiol. (1968), 8, 368-378.

Brewer, B. J. and Fangman, W. L.: *Mapping replication origins in yeast chromosomes.* Bioessays (1991), 13, 317-322.

Brown, E. H., Iqbal, M. A., Stuart, S., Hatton, K. S., Valinsky, J. and Schildkraut, C. L.: *Rate of replication of the murine immunoglobulin heavy-chain locus; evidence that the region is part of a single replicon.* Mol. Cell. Biol. (1987), 7, 450-457.

Bullock, P., Champoux, J. J. and Botchan, M.: *Association of crossover points with topoisomerase I cleavage sites: a model for non homologous recombination.* Science (1985), 230, 954-958.

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

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

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

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

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

Carroll, S. M., DeRose, M. L., Kolman, J. L., Nonet, G. H., Kelly, R. E. and Wahl, G. M.: *Localization of a bidirectional DNA replication origin in the native locus and in episomally amplified murine adenosine deaminase loci.* Mol. Cell. Biol. (1993), 13, 2971-2981.

Chan, C. S. and Tye, B. K.: *Autonomously replicating sequences in Saccharomyces cerevisiae.* Proc. Natl. Acad. Sci. USA (1980), 77, 6329-6333.

60

Chejanovsky, N. and Carter, B. J.: *Replication of a human parvovirus nonsense mutant in mammalian cells containing an inducible amber suppressor.* Virology (1989), 171, 239-247.

Cheung, A. K., Hoggan, M. D., Hauswirth, W. W. and Berns, K. I.: *Integration of the adeno-associated virus genome into cellular DNA in latently infected human Detroit 6 cells.* J. Virol. (1980), 33, 739-748.

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

Das, H. K., Jackson, C. L., Miller, D. A., Leff, T. and Breslow, J. L.: *The human apolipoprotein C-II gene sequence contains a novel chromosome 19-specific minisatellite in its third intron.* J. Biol. Chem. (1987), 262, 4787-4793.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Giraud, C., Wincovour, E. and Berns, K. I.: *Site-specific integration by adeno-associated virus is directed by a cellular DNA sequence*. Proc. Natl. Acad. Sci. USA (1994), 91, 10039-10043.

Green, M. R. and Roeder, R. G.: *Definition of a novel promoter for the major adeno-associated virus mRNA*. Cell (1980), 1, 231-242.

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

Hamlin, J. L. and Dijkwel, P. A.: *On the nature of replication origins of higher eukaryotes*. Curr. Opin. Genet. Dev. (1995), 5, 153-161.

Handa, H., Shiroki, K. and Shimojo, H.: *Establishment and characterization of KB cell lines latently infected with adeno-associated virus type I*. Virology (1977), 82, 84-92.

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

Hay, R. T. and DePamphilis, M. L.: *Initiation of SV40 DNA replication in vivo: location and structure of 5' ends of DNA synthesized in the ori region*. Cell (1982), 28, 767-779.

Hay, R. T., Hendrickson, E. A. and DePamphilis, M. L.: *Sequence specificity for the initiation of RNA-primed simian virus 40 DNA synthesis in vivo*. J. Mol. Biol. (1984), 175, 131-157.

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

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

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

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

Hermonat, P. L., Labow, M. A., Wright, R., Berns, K. I. and Muzyczka, N.: *Genetics of adeno-associated virus: isolation and preliminary characterization of adeno-associated virus type 2 mutants*. J. Virol. (1984), 51, 329-339.

Hoggan, M. D., Thomas, G. F., Thomas, F. B. and Johnson, F. B.: *Continuous 'carriage 'of adenovirus associated virus genome in cell cultures in the absence of helper adenovirus.* Proceedings of the Fourth Lepetit Colloquium (1972), pp. 243-249. Elsevier/North Holland Publishing Co., Amsterdam.

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

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

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

Im, D. S. and Muzyczka, N.: *The AAV origin binding protein Rep 68 is an ATP-dependent site-specific endonuclease with DNA helicase activity.* Cell (1990), 61, 447-457.

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

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

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

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

Konopka, A. K.: *Compilation of DNA strand exchange sites for non-homologous recombination in somatic cells.* Nucleic Acids Res. (1988), 16, 1739-1758.

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

Korneluk, R. G., MacLeod, H. L., McKeithan, T. W., Brooks, J. D. and MacKenzie, A. E.: *A chromosome 19 clone from a translocation*

breakpoint shows close linkage and linkage disequilibrium with myotonic dystrophy. Genomics (1989), 4, 146-151.

Kotin, R. M. and Berns, K. I.: *Organization of adeno-associated virus DNA in latently infected Detroit 6 cells.* Virology (1989), 170, 460-467.

Kotin, R. M., Siniscalco, M., Samulski, R. J., Zhu, X. D., Hunter, L., Laughlin, C. A., McLaughlin, S., Muzyczka, N., Rocchi, M. and Berns, K. I.: *Site-specific integration by adeno-associated virus.* Proc. Natl. Acad. Sci. USA (1990), 87, 2211-2215.

Kotin, R. M., Menninger, J. C., Ward, D. C. and Berns, K. I.: *Mapping and direct visualization of a region-specific viral DNA integration site on chromosome 19q13-ter.* Genomics (1991), 10, 831-834.

Kotin, R. M., Linden, R. M. and Berns, K. I.: *Characterization of a preferred site on human chromosome 19q for integration of adeno-associated virus DNA by non-homologous recombination.* EMBO J. (1992), 11, 5071-5078.

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

Labow, M. A. and Berns, K. I.: *The adeno-associated virus rep gene inhibits replication of an adeno-associated virus/simian virus 40 hybrid genome in COS-7 cells.* J.Virol. (1988), 62, 1705-1712.

Labow, M. A., Hermonat, P. L. and Berns, K. I.: *Positive and negative autoregulation of the adeno-associated virus type 2 genome.* J. Virol. (1986), 60, 251-258.

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

66

Leffak, M. and James, C. D.: *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. (1989), 9, 586-593.

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

Levine, A. J., Kang, H. S. and Billheimer, F. E.: *DNA replication in SV40-infected cells. Analysis of replicating SV40 DNA.* J. Mol. Biol. (1970), 50, 549-568.

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

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

Lichter, P. Chang Tang, C.-J., Call, K. Hermanson, G., Evans, G. A., Housman, D. and Ward, D. C.: *High resolution mapping of human chromosome 11 by in situ hybridization with cosmid clones.* Science (1990), 247, 64-69.

Linden, R. M., Wincour, E. and Berns K. I.: *The recombination signals for adeno-associated virus site-specific integration.* Proc. Natl. Acad. Sci. USA (1996), 93, 7966-7972.

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

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

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

Little, R. D. and Schildkraut, C. L.:*Initiation of latent DNA replication in the Epstein-Barr virus genome can occur at sites other than the genetically defined origin.* Mol. Cell. Biol. (1995), 15, 2893-2903.

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

Mariani, B. D. and Schimke, R. T.: *Gene amplification in a single cell cycle in chinese hamster ovary cells.* J. Biol. Chem. (1984), 259, 1901-1910.

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

McKeithan, T. W., Rowley, J. D., Shows, T. B. and Diaz, M. O.: *Cloning of the chromosom translocation breakpoint junction of the t(4;19) in chronic lymphocytic leukemia.* Proc. Natl. Acad. Sci. USA (1987), 84, 9257-9260.

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

Mishra, L. and Rose, J. A.: *Adeno-associated virus DNA replication is induced by genes that are essential for HSV-1 DNA synthesis.* Virology (1990), 179, 632-639.

Montminy, M. R., Sevarino, K. A., Wagner, J. A., Mandel, A. and Goodman, R. H.: *Identification of a cyclic-AMP-response element within the rat somatostatin gene.* Proc. Natl. Acad. Sci. USA (1986), 83, 6682-6686.

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

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

Newlon, C. S. and Theis, J. F.: *The structure and function of yeast autonomously replicating sequences.* Curr. Opinions in Genet. Dev. (1993), 3, 752-758.

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

Razin, S. V., Kekelidze, M. G., Lukhanidin, E. M., Scherrer, K. and Georgiev, G. P.: *Replication origins are attached to the nuclear skeleton.* Nucleic Acids Res. (1986), 14, 8189-8207.

Richardson, W. D. and Westphal, H.: *A cascade of adenovirus early functions is required for expression of adeno-associated virus.* Cell (1981), 27, 133-141.

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

69

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

Samulski, R. J., Zhu, X., XIao, X., Brook, J. D., Housman, D. E., Epstein, N. and Hunter, L. A.: *Targeted integration of adeno-associated virus (AAV) into human chromosome 19*. EMBO J. (1991), 10, 3941-3950 (Published erratum appears in EMBO J. (1992) 11, 1228)

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

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

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

Siegl, G., Bates, R. C., Berns, K. I., Carter, B. J., Kelly, D. C., Kurstak, E. and Tattersall, P.: *Characteristics and taxonomy of Parvoviridae*. Intervirol. (1985), 23, 61-66.

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

Srivastava, A., Lusby, E. W. and Berns, K. I.: *Nucleotide sequence and organization of the adeno-associated virus 2 genome*. J. Virol.

(1983), 45, 555-564.

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

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

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

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

Tratschin, J. D., Miller, I. L. and Carter, B. J.: *Genetic analysis of adeno-associated virus: properties of deletion mutants constructed in vitro and evidence for an adeno-associated virus replication function.* J. Virol. (1984), 51, 611-619.

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

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

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

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

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

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

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

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

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

Weindler, F. W. and Heilbronn, R.: *A subset of herpes simplex virus replication genes provides helper functions for productive adeno-associated virus replication.* J. Virol. (1991), 65, 2476-2483.

Weitzman, M. D., Kyosto, S. R. M., Kotin, R. M. and Qwens, R. A.: *Adeno-associated virus (AAV) Rep proteins mediate complex formation between AAV DNA and its integration site in human DNA.* Proc. Natl. Acad. Sci. USA (1994), 91, 5808-5812.

West, M. H., Trempe, J. P., Tratschin, J. D. and Carter, B. J.: *Gene expression in adeno-associated virus vectors: the effects of chimeric mRNA structure, helper virus, and adenovirus VA1 RNA.* Virology (1987), 160, 38-47.

Yakinoglu, A. O., Heilbronn, R., Burkle, A., Schlehofer, J. R. and zur Hausen, H.: *DNA amplification of adeno-associated virus as a response to cellular genotoxic stress.* Cancer Res. (1988), 48, 3123-3129.

Yakobson, B., Koch, T. and Winocour, E.: *Replication of adeno-associated virus in synchronized cells without the addition of a helper virus.* J. Virol. (1987), 61, 972-981.

Yang, Q., Kadam, A. and Trempe, J. P.: *Mutational analysis of the adeno-associated virus rep gene.* J. Virol. (1992), 66, 6058-6069.

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

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

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