HUMAN ORIGINS OF DNA REPLICATION:
Identification, Analysis and Application

by
TORSTEN NIELSEN

Department of Medicine
Division of Experimental Medicine
McGill University
Montreal, Canada

October, 1995

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements of the degree of Doctor of Philosophy

ISBN 0-612-19757-3

Canada

ABSTRACT

While replication origins, *cis*-acting sequences directing the initiation of DNA synthesis, have been well-characterized in many model organisms, the multiple sequence and protein components present at the chromosomal origins of higher eukaryotic organisms have not yet been fully defined. Genetic assays that identify origin function in cloned DNA fragments would provide a useful approach for the isolation and analysis of mammalian DNA replication origins.

In this thesis, (1) cloned fragments from a known mammalian origin, the oriβ of the hamster 3' *DHFR* region, are demonstrated to replicate autonomously, both following transfection into human cells, and when used as templates in an *in vitro* replication system based on human cell extracts; (2) larger scale versions of these two assay methodologies are used to isolate over 40 novel putative origins of DNA replication from anticruciform purified human genomic DNA libraries; (3) transfection and *in vitro* autonomous replication assays are applied to demonstrate the potential origin function of a mitochondrial DNA sequence implicated in the insertional mutagenesis of a human genomic locus; (4) an origin mapping strategy based on the *in vitro* assay is used to provide evidence for the existence of a replication origin in a cloned and sequenced portion of the human 15q11q13 chromosomal subdomain, a region associated with allele-specific replication timing, genomic imprinting, and genetic disease; and (5) some of these autonomously replicating origins are cloned into a selectable YAC vector and are shown to permit the long term episomal maintenance, in human cells, of the transfected plasmid constructs.

These results consistently demonstrate that short mammalian genomic DNA fragments can replicate autonomously, supporting the applicability of the replicon model in humans, and could be extended to the search for an origin core consensus element, to the investigation of higher order organization and temporal control of human DNA replication origins, and to the construction of a complete human artificial chromosome.

RÉSUMÉ

Quoique les origines de réplication, des séquences en *cis* qui dirigent la réplication de l'ADN, soient bien caracterisées chez plusieurs organismes modèles, les composants multiples en sequence d'ADN et en protéines presents aux origines chromosomiques des organismes eucaryotes supérieurs ne sont pas encore complètement définis. Des tests génetiques identifiant les origines de réplication dans des fragments d'ADN recombinant procureraient une méthode utile pour isoler et analyser les origines de réplication des mammifères.

Dans cette thèse, (**1**) il est demontré que des fragments recombinant d'une origine connue de mammifère, oriβ, provenant de la région en 3' du gène *DHFR* de hamster, se répliquent de façon autonome après transfection dans des cellules humaines ou lorsqu' utilisés comme substrat dans un système de réplication *in vitro* utilisant des extraits de cellules humaines; (**2**) des versions à grande échelle de ces deux méthodologies sont utilisées pour isoler plus de 40 nouvelles potentielles origines de réplication de l'ADN à partir de banques d'ADN génomiques purifées avec des anticorps anticruciformes; (**3**) la fonction d'origine d'une séquence mitochondriale impliquée dans la mutagénese insertionelle d'un locus génomique humain est demontrée par des essais de réplication autonome par transfection et *in vitro*; (**4**) une stratégie de cartographie d'origines basée sur l'essai *in vitro* procure l'évidence de l'existence d'une origine de réplication dans une portion clonée et séquencée du sous-domaine chromosomique humain 15q11q13, région associée au minutage de la réplication d'une allèle spécifique, à l'empreinte génetique, ainsi qu'aux maladies génétiques; et (**5**) il est montré que quelques unes des origines de réplication autonomes, clonées dans un vecteur YAC selectable, permettent la maintenance à long terme de ces plasmides recombinants dans les cellules humaines transfectées.

Ces resultants systématiquement demontrent que des courts fragments d'ADN génomique de mammifère peuvent se répliquer de façon autonome, supportant l'applicabilité du modèle du réplicon chez l'humain, et peuvent s'étendre à la recherche du consensus de l'élément essentiel des origines, à l'examen du control temporel et de l'organization supérieure des origines de réplication d'ADN humaines, ainsi qu'à la construction d'un chromosome artificiel humain complet.

TABLE OF CONTENTS

PREFACE

This thesis is presented in manuscript-based form, under the terms listed by the Faculty of Graduate Studies and Research in their Guidelines for Thesis Preparation:

> Candidates have the option of including, as part of the thesis, the text of one or more papers submitted or to be submitted for publication, or the clearly-duplicated text of one or more published papers. These texts must be bound as an integral part of the thesis.
>
> If this option is chosen, connecting texts that provide logical bridges between the different papers are mandatory. The thesis must be written in such a way that it is more than a mere collection of manuscripts; in other words, results of a series of papers must be integrated.
>
> The thesis must still conform to all other requirements of the "Guidelines for Thesis Preparation". The thesis must include: A Table of Contents, and Abstract in English and French, an introduction which clearly states the rationale and objectives of the study, a comprehensive review of the literature, a final conclusion and summary, and a thorough bibliography or reference list.
>
> Additional material must be provided where appropriate (e.g. in appendices) and in sufficient detail to allow a clear and precise judgement to be made of the importance and originality of the research reported in the thesis.
>
> In the case of manuscripts co-authored by the candidate and others, the candidate is required to make an explicit statement in the thesis as to who contributed to such work and to what extent. Supervisors must attest to the accuracy of such statements at the doctoral oral defense. Since the task of the examiners is made more difficult in these cases, it is in the candidate's interest to make perfectly clear the responsibilities of all the authors of the co-authored papers. Under no circumstances can a co-author of any component of such a thesis serve as an examiner for that thesis.

Chapter 1 consists of a detailed review of the literature concerning origins of DNA replication. Chapter 2 presents a short summary of the rationale and objectives of the work described in subsequent chapters. Chapters 3, 4 and 5 are reprints of published articles, while Chapters 6 and 7 contain the complete text of articles submitted for publication. Chapter 8 contains the final summary discussion. A short statement, placed in the linking text before those chapters containing a published paper, contains the details of each article's publication and includes a statement of contribution, describing the extent of my role in these multi-authored papers. Addendums to chapters 4 and 6 provide

relevant additional data which was not included in the published or submitted manuscript. Each chapter contains its own separate reference list.

An acknowledgements section is present in each of chapters 3 through 7, detailing the assistance provided by others who were not co-authors. I would like to express my gratitude for the funding I have received from the Medical Research Council of Canada, the McGill MD/PhD program, the McGill Faculty of Medicine, and the Défi Corporatif Canderel. I would also like to thank my supervisor, Prof. Gerald B. Price, and my advisor, Prof. Harry L. Goldsmith, for their constant and unwavering support, from my days as a bemused final-year undergraduate at the University of British Columbia, until my return this month to clinical medicine here at McGill. My thanks also go out to Dr. Price for editing this thesis, and to both him and Prof. Maria Zannis-Hadjopoulos for advice in experimental priorities, planning and analysis. The French translation of the Abstract was prepared with the assistance of Richard Pelletier and Prof. Maria Zannis-Hadjopoulos.

Finally, I would like to thank the students and staff, past and present, of the McGill Cancer Centre, particularly my co-workers in the Price and Zannis-Hadjopoulos laboratories, for so graciously putting up with me over the past four years, and of course my wife Karen, and my family back in North Vancouver, and those friends I met along the way who helped me become who I am today.

Torsten Nielsen
August, 1995.

CLAIMS TO ORIGINALITY

The following results, presented in this thesis and constituting my own work, are original:

1) Demonstration of the reproducible and significant autonomous replication of short (< 5 kb) DNA fragments derived from the hamster 3' *DHFR*-associated origin of bidirectional replication, oriβ.

2) The development of a method for the isolation of putative human origins of DNA replication by mass transfection or mass *in vitro* assay of plasmids subcloned from anticruciform immunoaffinity-purified human genomic DNA libraries.

3) The demonstration of autonomous replication activity in a DNA fragment created by the insertion of mitochondrial DNA into human genomic DNA, and mapping of the initiation site for its replication *in vitro*.

4) Analysis of the statistical significance of matches to origin-associated consensus elements that are observed in a 13,022 bp sequenced portion of the imprinted human chromosome 15q11-q13 region, and demonstration that plasmid p82 replicates autonomously, initiating *in vitro* replication within its 1012 bp genomic insert cloned from this region.

5) The cloning of short (< 5 kb) DNA fragments containing mammalian origin sequences into a circular yeast artificial chromosome vector, and the demonstration, by fluctuation assay and PCR analysis of cesium chloride/ethidium bromide ultracentrifugation fractions, that such constructs can be maintained in human cells in a purely episomal state.

CHAPTER ONE

INTRODUCTION

OVERVIEW

DNA replication may be the fundamental process of life. From a completely reductionist perspective, the apparent goal of life itself may be simply the propagation of DNA, the closest thing to an immortal constant across generations of cells and independent organisms.

To a large extent, the physiology of the cell is geared towards the preparation, execution, and resolution of DNA replication. The control of replication is careful and precise; multiple pathways for signal transduction and cellular growth regulatory checkpoints converge upon the DNA to permit replication only when the internal and external environment is appropriate. Intricate biochemical processes are generally regulated at an early step in their pathway, and DNA replication is no different: the key control point seems to be at initiation. The *cis*-acting DNA element defining the position and providing control over initiation is the replication origin, the focus of this dissertation. It is hoped that increasing our knowledge of the nature and function of mammalian and particularly human origins of DNA replication will contribute to a better understanding of human cellular physiology, including not only replication and its direct regulation, but also the related processes of DNA transcription, repair, and recombination. Ultimately, this line of investigation may also help to answer questions about the causes of and treatments for human genetic and neoplastic diseases.

REPLICATION ORIGINS IN MODEL SYSTEMS

The human genome, with a haploid size of three billion base pairs, is the product of about as many billion years of evolution. Not surprisingly, investigators have found the mechanisms by which human DNA is replicated to be extremely complex; consequently, most of our current knowledge of DNA replication has come from the study of model organisms, both prokaryotic (bacteria, bacteriophages, and plasmids) and eukaryotic (viruses, yeasts and protists, and mammalian cultured cells).

The two key phases of the replication process are initiation and elongation, and the latter is the better characterized, in large part due to the development of soluble purified *in vitro* replication systems, such as those based on plasmids containing the *E. coli* chromosomal origin of DNA replication *oriC* (Fuller et al., 1981) or the mammalian SV40 viral origin and transactivating T-antigen protein (Li and Kelly, 1984). Several general principles have emerged from work dissecting the replication process in these and other model systems (Kornberg and Baker, 1992). DNA replication is semiconservative,

yielding double strand products containing one original template DNA strand paired with one antiparallel newly synthesized product strand. The basic enzymatic activities required are the *helicase*, which unwinds the double strand template and defines the head of the "growing fork," a *topoisomerase*, to relieve supercoiling stress which builds up ahead of the growing fork, *SSB's* (single-stranded binding proteins), which stabilize the unwound, single stranded template DNA, and the *polymerase*, which actively adds new nucleotides to the growing nascent strand which are complementary to those of the template. New bases can only be added by polymerases to free 3'-OH ends of the growing chain (in part, to facilitate associated *exonuclease* proofreading functions which help ensure the fidelity of the process), and this requirement for 5' - 3' replication polarity means that the two unwound template strands must be treated somewhat differently. The *leading strand*, which is unwound 3' - 5', can act directly as a template for 5' - 3' antiparallel elongation of the nascent strand in the same direction as growing fork progression. However, the *lagging strand* must be periodically reinitiated by *primase* activity, laying down RNA primers for polymerase extension as new tracts of template strand DNA are unwound. DNA polymerase extension of these primers produces Okazaki fragments, which are variable in size but generally about 2000 bp long in prokaryotes and 200 bp long in eukaryotes, and must eventually be connected by *DNA ligase* activity after all RNA has been replaced with DNA.

Kornberg and Baker (1992) assert that the key principles of replication with regard to the initiation step are that control over replication takes place at initiation, and that the initiation sites are specific DNA sequences termed origins. It is important to realize that the term "origin" can refer either to a *cis*-acting control sequence (sometimes termed a "replicator" or a "genetic origin") directing the beginning of the replication process, or to the actual initiation site (or "functional origin") where the template unwinding begins and the first new bases are laid down (DePamphilis, 1993a; Stillman, 1994). In the model systems characterized so far, the control elements and the initiation site coexist within a small functional unit a few hundred base pairs in size.

Model organisms may have important differences in their mechanisms of replication. Perhaps primary among these is the number of origins present in the genome. A single origin is sufficient for all known prokaryotic systems, as well as for most eukaryotic viruses, but each of the much larger eukaryotic chromosomes requires multiple initiation events, if the genome is to be replicated within a reasonable time (considering that replication forks progress at the rate of about 100 kb per minute in prokaryotes and 2 kb per minute in eukaryotes) (Kornberg and Baker, 1992). Some of the polymerases and other enzymes have fundamental differences in prokaryotic versus eukaryotic systems. The

controls over the replication process exerted at the origin level must also be grossly different in viruses and phages in comparison with their host cells, since viral genomes in lytic growth will divide as quickly as possible, whereas prokaryotic and eukaryotic cells are under strict growth control related to external signals and to their current position in the cell cycle. Another major contrast among model systems relates to whether they replicate in a bidirectional or a unidirectional fashion -- does a single initiation event result in two growing forks progressing in opposite directions from the origin (probably the more common case overall) or just one growing fork? Finally, the mechanism of the initial priming event for leading and lagging strand synthesis can vary dramatically. While RNA primers laid down by primase activity is typical in most systems, RNA polymerase is employed by ColE1 plasmids and M13 phage, and retroviruses hijack tRNA to act as a primer. Perhaps even more unusual are the adenoviruses (King and van der Vliet, 1994) and PRD1 phage (Savilahti and Bamford, 1993), linear DNA viruses which use a protein to prime initial strand synthesis. DNA primers resulting from a nicking event on a circular molecule (or linear DNA with hairpin termini) can also be employed to begin the new chain, resulting in the "rolling circle" type replication seen in the pT181 plasmid (Wang et al., 1993) and the ΦX174 and Ff phages, among others.

Several model systems have been especially instructive for understanding the structure and function of replication origins, and these include both prokaryotic and eukaryotic examples.

Prokaryotes

Bacteria, phage, and prokaryotic plasmids replicate their genome following a single initiation event, which includes several steps: identification of a starting point, local denaturation of the double helix, loading of helicase activity to permit further unwinding, and docking of primase and polymerase complexes to commence the elongation phase (Marians, 1992). The most common strategy, exemplified by *E. coli* and λ phage, involves the cooperative binding of an initiator protein multimer at a specific genomic site. This initiator complex alters the local DNA structure, resulting in the denaturation of a nearby A-T rich region, and binds and loads DNA helicases to commence bidirectional replication. An alternate strategy, employed by colE1-type plasmids and by T4 and T7 phage, uses specific cleavage of an RNA transcript as a primer for DNA synthesis, taking advantage of the DNA denaturation inherent in the process of transcription to facilitate docking of replicative helicase and primase activities, creating a single fork for unidirectional replication.

Escherichia coli

A single, specific origin element, *oriC*, allows the gram (-) enteric bacterium *E. coli* to replicate its entire 4.7 x 10^6 bp genome. The minimal size of *oriC* is 245 bp; substitutions are not tolerated in fully half of these nucleotide positions (Benbow et al., 1992), and spacer DNA sequences separating functional components are of fixed length. One side of the *oriC* element contains four separate TTATMCAMA (9-mer) "dnaA boxes," arranged as two inverted repeats. The other end of the locus contains a cluster of three A-T rich 13-mers.

Origin activation begins with the binding of dnaA protein monomers to the 9-mer boxes (Schaper and Messer, 1995), followed immediately by a highly cooperative cascade of dnaA oligomerization, with the double stranded chromosomal DNA of much of the *oriC* region left wrapped around the outside of a protein complex containing thirty dnaA subunits (Kornberg and Baker, 1992). The stresses on the DNA structure imposed by complex formation promote melting of the duplex at the A-T rich 13-mers, regions with fewer interstrand hydrogen bonds, creating an "open complex." The dnaB helicase protein, itself a homohexamer, is ordinarily sequestered as a soluble complex with the dnaC protein, but the affinity of dnaB for the combination of dnaA and the denatured 13-mers is sufficient to unload dnaB from its interaction with dnaC and dock this helicase to the open complex, forming a "prepriming complex." The helicase action of dnaB then begins to unwind the template further, with the assistance of single-stranded binding protein to stabilize thermodynamically the denatured DNA, and of DNA gyrase to relieve supercoil buildup ahead of the helicases. Two dnaB homohexamers define the heads of two growing forks, moving away from each other. The primase peptides then recognize, at the initiation site, the dnaA protein, dnaB protein, and single stranded denatured DNA, allowing RNA primers to be synthesized on the leading strand of each fork. Helicase procession then uncovers short DNA hairpins, formed from denatured inverted repeats, that can act as *pas*, or primosome assembly sites for the priming of lagging strand synthesis (Kornberg and Baker, 1992). Elongation of these primers requires loading of the DNA polymerase III holoenzyme complex, itself composed of ten subunits.

Regulation of the initiation process at *oriC* can occur at one of several steps (Marians, 1992). The dnaA protein is the key "initiator" protein in *E. coli*, and its transcription is under tight autoregulatory control, such that its concentration is dependent on overall cell size. Many other genomic binding sites for dnaA are known to exist, which may serve to sequester the protein, preventing it from reaching a critical mass for

oligomerization and initiation until a suitable excess has been synthesized. Furthermore, the dnaA protein is active only when conjugated to ATP, and hydrolysis of ATP during initiation may help prevent reinitiation. A second and perhaps more important check on premature reinitiation comes through the interaction of *oriC* with the cell membrane. The *oriC* region contains thirteen GATC *dam*-methylase recognition sites, which are left in hemimethylated form following replication of the genome. The hemimethylated GATC cluster is tightly bound to the cell membrane, sequestering *oriC* from interaction with dnaA until after the relatively slow methylase activity can return these sites to fully methylated form.

Regulation of initiation also takes place at the second step, formation of the open complex (Marians, 1992). While the melting of the duplex at the A-T rich 13-mers is favoured by the dnaA oligomerization and the close relation of three A-T rich tracts, melting also requires negative superhelical tension in the DNA. Local transcriptional activity, in the flanking *gidA* and *mioC* loci a few hundred base pairs away, helps establish a DNA domain with the appropriate superhelical state, and transcription-dependent structural changes may be transmitted into *oriC*. Structural changes conducive to denaturation are also facilitated by binding of the HU and Fis proteins, which induce DNA bending; HU has a strong binding affinity for DNA cruciforms (Bonnefoy et al., 1994). Finally, the Kornberg laboratory has discovered a novel protein, IciA, which binds the 13-mer sequences and blocks denaturation: a negative regulator of DNA replication at the level of open complex formation (Hwang and Kornberg, 1990).

In summary, the *E. coli* origin of DNA replication contains a reiterated binding site for an initiator protein, inverted repeat elements, an A-T rich region of reduced helical stability, close relation to regions of active transcription, and an interaction with the bacterial cell membrane. The *oriC* element is well conserved across the enterobacteriaceae family; indeed, the basic design of *oriC* is maintained in most gram (-) bacteria (Kornberg and Baker, 1992).

λ phage

The initiation of DNA replication in the *E. coli* bacteriophage λ has been very well characterized and is similar to that of its host. The λ phage genome is 48 kb in length, and while this phage is capable of integrative lysogeny, replicating passively with the host chromosome, it initiates its own lytic replication while in free circular form. The replication origin region, oriλ, is about 2500 bp in size and includes one of the phage's two major promoters, P_R, as well as a 200 bp core origin region containing multiple 19 bp inverted

repeat "λO boxes" (arranged as a series of four direct repeats), flanking a 40 bp A-T rich region, which itself borders on a 28 bp inverted repeat (Kornberg and Baker, 1992). An intrinsically bent oligo·dA sequence is also present. The phage replaces dnaA with the λO protein, and dnaC with λP, but otherwise relies on the host replication apparatus.

λO, the initiator protein, recognizes the 19-mers, and binds cooperatively, forming an "O-some" complex; 95 bp of oriλ DNA is wrapped around 8 λO monomers (Marians, 1992). An open complex is then formed by unwinding of the neighbouring A-T rich sequence; again, this is favoured by negative supercoiling of the genome, and is dependent upon transcriptional activity from the nearby P_R promoter. While P_R directs transcription into and through oriλ, plasmids based on oriλ maintain replication when P_R is deleted and a different promoter is introduced anywhere within 100 bp of oriλ, regardless of the new promoter's transcriptional orientation (Kornberg and Baker, 1992), suggesting that structural changes induced by transcription are vital to origin function, rather than transcription *per se*. The cI λ transcriptional repressor blocks initiation, and for λ phage, so too does the HU protein (which stimulates open complex formation in the host), perhaps because oriλ is intrinsically bent without HU action (Marians, 1992). After formation of the open complex, the dnaB helicase is loaded through the action of λP, which recognizes the combination of the O-some and the denatured A-T rich box (Stillman, 1994). λP captures dnaB from its host complex with dnaC by virtue of a very high binding affinity, but at the same time inhibits dnaB's helicase action and does not dissociate spontaneously from the prepriming complex. Removal of λP requires the action of three host heat shock chaperonin-type proteins: dnaJ, dnaK and GrpE. The free dnaB helicase then becomes active, unwinding the oriλ region further; primase and polymerase host proteins are in turn recruited to the growing forks. Stillman (1994) notes that the actual initiation site, where the first new bases are laid down, is a matter of debate and may include a broad zone of possibilities on either side of oriλ. Furthermore, up to 25% of initiations create only a unidirectional growing fork, but since a unidirectional fork can be switched to a rolling-circle replication mechanism, this represents a strategy for the phage to continue lytic replication without requiring further initiation events (Kornberg and Baker, 1992).

To summarize, the origin region of bacteriophage λ includes binding sites for an initiator protein, and inverted repeat, A-T rich, and bent DNA elements. Activation is dependent on the influence of transcription, and on action of heat shock proteins to modify the prepriming complex.

Many other bacteriophages and prokaryotic plasmids ensure their own replication by a similar strategy: replacing dnaA with their own initiator protein (Kornberg and Baker, 1992). In general, their replication origins display a structure including inverted repeat

sequences and multiple copies of an initiator protein binding site, bordering on an A-T rich region. In an interesting twist, the P1 phage (Chattoraj et al., 1988) and the R6K plasmid (Mukherjee et al., 1988) each have two clusters of initiator protein binding sites located about one kilobase apart, which are induced to form a DNA loop during formation of an initiator protein oligomerized complex.

ColE1

Most typical plasmid vectors commonly employed in molecular biology laboratories, including pBR322, contain a replication origin originally derived from the ColE1 *E. coli* plasmid, and employ an initiation strategy which differs significantly from *oriC* and oriλ (Kornberg and Baker, 1992). The ColE1 origin sequence contains a promoter 555 bp upstream of a "cleavage sequence," followed by two dnaA boxes 100 bp downstream, and a *pas* 50 bp further on. RNA polymerase transcription begins at the promoter and reads into the centre of the origin region, transiently creating an RNA:DNA hybrid together with a displacement loop (R- or D-loop) of single stranded DNA (Marians, 1992). Inverted repeat sequences transcribed into the RNA make its tail form a particular secondary structure of stem-loops, which is recognized by a specific RNase. Cleavage of the RNA within the origin provides primers for DNA polymerase I leading strand elongation, and DNA synthesis partially stabilizes the displacement loop. Two dnaA boxes just downstream of this cleavage site bind dnaA, and the combination of dnaA and the single-stranded DNA of the displacement loop attracts the dnaB helicase, which then heads a unidirectional growing fork. DNA polymerase I on the leading strand is replaced by the polymerase III holoenzyme; meanwhile, helicase action uncovers the *pas* sequence for Okazaki fragment priming and lagging strand synthesis is begun.

The frequency of initiation is controlled by an antisense RNA system: a second promoter in opposite orientation produces a transcript through the origin region, which can form an RNA:RNA duplex with the initiating RNA which would otherwise go on to adopt the secondary structure required for RNase recognition and cleavage (Kornberg and Baker, 1992).

The ColE1-type strategy, using RNA polymerase to synthesize a primer for DNA synthesis, is also used by the T-even and T-odd phage groups, the best studied of which are the T4 and T7 bacteriophage.

Eukaryotes

Replication in eukaryotes has consistently proven to be more complicated than in prokaryotes. At the level of replication enzymology, eukaryotes employ multiple subtypes of ligases (Lindahl and Barnes, 1992), DNA polymerases, RNA polymerases, and topoisomerases (Kornberg and Baker, 1992) and elongation is thought to employ a huge multiprotein apparatus including polymerase, primase, ligase, helicase and topoisomerase activities together with processive factors and proteins mediating entry and exit of others from the complex (Wu et al., 1994). Eukaryotic DNA differs in subtle ways from that of prokaryotes, for example by the presence of methylcytosine instead of methyladenine, and more frequent inverted repeat and A-T rich elements (Krysan et al., 1993). Large stretches of eukaryotic DNA may be transcriptionally inactive, and the coding sequences themselves usually contain introns (Lambowitz and Belfort, 1993). Importantly, eukaryotic DNA is packaged by histones, and further organized into tertiary structures to the level of large chromosomal subdomains, which may differ greatly in their overall transcriptional activity and degree of final DNA condensation. The long linear chromosomes contain multiple replication origins spaced out along their length, and overall replication is under tight cell cycle control.

While all these complexities may make the analysis of human DNA replication a daunting task, problems can be simplified through the use of convenient eukaryotic model systems. Many such systems have been studied, and each offers certain advantages; for example, human adenovirus is easy to grow and its overall size is very manageable; *Physarum polycephalum* forms a plasmodium in which all nuclei replicate in synchrony, facilitating detection of origin function (Benard et al., 1992), and many fungal systems exist which may represent relatively early evolutionary steps beyond the prokaryote (Peng et al., 1993). However, I will concentrate on four specific model systems which have been especially well characterized and may prove relevant to particular aspects of human genomic DNA replication: the human mitochondrion, Simian Virus 40 (SV40), Epstein-Barr Virus (EBV), and the budding yeast, *Saccharomyces cerevisiae*.

Mitochondria

Although the numbers vary widely according to cell type, a typical human cell may possess on the order of 500 mitochondria, each with 8 copies of the 16,569 bp human mtDNA (mitochondrial DNA) genome, accounting for roughly 1% of total cellular DNA (Kornberg and Baker, 1992). The complete human mtDNA sequence has been determined,

and it contains two rRNA genes, 22 tRNA genes, and coding sequence for thirteen polypeptides involved in oxidative phosphorylation, including, for example, three subunits of the cytochrome c oxidase complex (Anderson et al., 1981). Importantly, all enzymes required for mitochondrial replication are encoded in the nucleus, translated in the cytoplasm, and transported into the mitochondrion. mtDNA contains two replication origins, but this is somewhat misleading; each origin directs the replication of only one of the two mtDNA strands -- termed the H (heavy) and L (light) strands based on differential purine content (Clayton, 1992). O_H, the heavy strand replication origin, is located in the D-loop region of mtDNA, the only noncoding portion of the very compact mitochondrial genome, which contains both transcriptional promoters (an H and L strand promoter lie in the D-loop, pointing in opposite directions). The light strand replication origin O_L lies in a tRNA gene cluster some 6000 bp away.

In the case of the human mitochondrion, replication is primed by cleavage of an mRNA transcript, reminiscent of the ColE1 origin. Initiation takes place at the L strand promoter, which uses the L strand as a template to synthesize a new H strand as RNA. Initiation requires the nuclear-encoded mitochondrial transcription factor 1 (MTF1), which induces DNA bending and wraps the promoter around itself (Ghivizzani et al., 1994), and mitochondrial RNA polymerase proteins, acting on a mitochondrial genome containing about 100 negative supercoils (Clayton, 1992). Most such events lead to complete polycistronic mRNA transcription, but occasionally the transcript is cleaved at O_H by specific, sequential RNase ribozyme action to create a primer for DNA replication, recognized by γ-polymerase.

However, 95% of such DNA extensions are terminated about 500 bases downstream, and do not go on to replicate the mitochondrial genome (Madsen et al., 1993). Premature termination seems to be under control of specific, regulated DNA binding proteins which recognize a site just upstream of an approximately 80 bp termination zone. The result is a 570 - 655 bp aborted DNA strand with an RNA primer, which remains hybridized to the mtDNA L strand, leading to a stable displacement loop of H strand mitochondrial genomic DNA; hence, the term D-loop is used to describe this portion of mtDNA (Clayton, 1992). Under electron microscopy, the vast majority of mtDNA genomes appear as supercoiled covalently closed circles with a visible D-loop. In this state, the DNA is free to reinitiate transcription at the L strand promoter, and the transcript could subsequently be switched to another replication attempt by RNase action at O_H.

In cases where initiation is not terminated, replication of the H-strand by γ-polymerase continues unabated around the 16 kb genome, displacing the original H-strand as it progresses at the very slow rate of about four nucleotides per second (one order of

magnitude slower than eukaryotic genomic DNA, two orders slower than prokaryotic replication).

The O_L portion of mtDNA is not replicated until the H-strand has progressed 10 kb around the circle. Once the original H strand is displaced at O_L, an inverted repeat sequence is free to form a stem-loop structure which is recognized by mtDNA primase (Kornberg and Baker, 1992). An RNA L-strand replication primer is then synthesized and extended by γ-polymerase in a process reminiscent of lagging strand synthesis in prokaryotes, except that in this case the *pas* is 10 kb downstream of the leading strand origin.

Control points for replication (as opposed to transcription) occur at the steps of H-strand RNase cleavage and premature termination. Overall mitochondrial growth is kept proportional to cell growth, and thus mitochondrial numbers are relatively constant through cell divisions; however, there is no cell cycle limitation on the timing of mtDNA replication, and the choice of which mtDNA genomic copies will replicate seems to be a stochastic process, since some genomes replicate twice in a cell cycle, others not at all (Kornberg and Baker, 1992). Mitochondrial division is stimulated with the oxidative stress brought on by even mild exercise (Madsen et al., 1993), but the manner in which the cell influences the mtDNA initiation apparatus has yet to be worked out, although it is known that the level of the key H-strand RNase is proportional to cellular growth rate.

The endosymbiont hypothesis holds that mitochondria evolved from captured prokaryotic organisms, and evidence for this includes the prokaryotic protein synthesis apparatus (rRNA and tRNA that work in an 80S ribosome) and polycistronic nature of the mRNA transcripts of the mitchondrion. The transcription-primed mechanism of replication resembles ColE1-type prokaryotic origins much more than eukaryotic genomic DNA replication. In addition, DNA repair systems in the organelle are very poor; large deletions of mtDNA are common in the 10 kb interorigin region, and the RNA primers for replication are not replaced by DNA and become permanent parts of the genome. Thus, human mitochondria, while interesting in their own right, may not be a very good model for replication processes in the nucleus of the cell, where the very enzymes used for mtDNA replication are encoded.

In summary, human mitochondrial replication uses RNA priming requiring promoter and ribonuclease activity, in a region which contains induced DNA bends, a triple-stranded displacement loop structure, and extensive negative supercoiling. An inverted repeat is key to light strand replication, at least.

Other vertebrates seem to use the same general origin design; the inverted repeat structure of O_L is conserved in other species, even if primary sequence is not (Anderson et al., 1981). Lower eukaryotes have larger, more complex mtDNA, but at least in the case of

Neurospora crassa, the structure of origin *cis*-components, based on promoters and inverted repeats, seems to be maintained (Almasan and Mishra, 1990). *Drosophila melanogaster*, on the other hand, requires a large A-T rich component in its mitochondrial origin (Lewis et al., 1994). *S. cerevisiae* has a comparatively huge 75 kb mitochondrial genome with multiple replication origins, but has yet to be thoroughly characterized (Clayton, 1992).

SV40

Simian Virus 40 (SV40) is a eukaryotic virus for which primate cells are the natural host; it has a simple double-stranded DNA circular genome 5.4 kilobases in size. This virus makes a particularly good model for human DNA replication because, following infection, it is transported into the nucleus where it replicates strictly in S phase (Fotedar and Roberts, 1992). The viral DNA is packaged by host nucleosomes, and all replication machinery is supplied by the host, except for one viral protein: the SV40 large T antigen (T-Ag). T-Ag possesses the multifunctional attributes of initiator protein, helicase, and transcription factor, and interacts not only with the viral DNA, but also with both the elongation apparatus and several proteins involved in cell cycle control.

The SV40 replication origin lies within the 450 bp noncoding viral control region, also containing the early and late viral gene promoters and a transcriptional enhancer element. The indispensable minimal origin core is 64 bp in size, and includes 4 copies of the T-Ag recognition site GAGGC arranged as a 27 bp GC-rich inverted repeat element, flanked on the early gene side by a second inverted repeat, and on the late gene side by a 17 bp AT-rich element (Benbow et al., 1992). *In vivo*, SV40 is negatively supercoiled, and an inverted repeat in the origin region has been shown, by electron microscopy, to be able to adopt a DNA cruciform alternate secondary structure (Hsu, 1985). The AT-rich region has an intrinsic static bend (Schuller et al., 1994), and its absolute sequence is not as important as its overall structure and AT content (Galli et al., 1993).

Initiation is achieved through the binding of two SV40 large T antigen homohexamers to their recognition sites in the origin core, which induces a structural destabilization in the flanking inverted repeat and AT-rich sequences (Benbow et al., 1992). The T-Ag helicase activity then proceeds to unwind the DNA bidirectionally from the origin (Ramsperger and Stahl, 1995), and through its interaction with the polymerase α-primase complex, loads the enzymes necessary to lay down the first new bases.

The efficiency of this initiation process is heavily influenced by auxiliary control elements which flank the origin core: aux1, a site where T-Ag binds as a transcription

factor, and aux2, which is recognized by host transcription factors (DePamphilis, 1993a). Each alone stimulates initiation three to sixfold, but combined deletion of a neighbouring enhancer region and both aux elements reduces replication by a factor of at least 100 (Guo et al., 1989) suggesting a synergistic effect. While in theory, eukaryotic transcription factors could function in initiation by providing primers, displacing nucleosomes, or assisting in the loading of the initiator protein, evidence from the SV40 system is most consistent with a model whereby the regional binding of transcription factors influences the DNA structure at the origin, facilitating its unwinding by T-Ag -- similar to the influence of transcription on *oriC* and oriλ.

Because only a single viral protein is required to replicate SV40 in human cells, it has proven a particularly useful model for *in vivo* and *in vitro* analysis of human DNA replication, except for the initiator and helicase roles (which are filled by T-Ag). After T-Ag loads polymerase α-primase, a 10 bp RNA primer is laid down, followed by a 34 bp initial DNA sequence; subsequently, the polymerase δ holoenzyme complex takes over, a process performed once on each leading strand, but repeated, on the lagging strand, approximately every 200 bases (Waga and Stillman, 1994). Elongation continues at about 50 bp per second, terminating when the replication forks eventually collide. Typically, 100,000 virus genomes will be made (Kornberg and Baker, 1992); replication is shut down when high concentrations of T-Ag allows its binding to lower-affinity sites which switch off replication and early gene transcription in favour of late gene transcription (viral capsid proteins).

The host proteins required for viral replication are only active in the S phase of the cell cycle, so the virus forces its host into active cycling by using T-Ag to bind and sequester the Rb and p53 tumor suppressors (Zhu et al., 1991). Phosphorylation of T-Ag by host cdk proteins is necessary to activate the helicase function of T-Ag, but not any of its other DNA or protein binding functions (Moarefi et al., 1993).

Plasmids carrying the SV40 origin core and auxiliary sequences replicate as episomes in human or monkey cells, if T-Ag is provided in *cis* or *trans*. The efficiency of initiation at the SV40 origin is influenced, in these engineered contexts, by the nature of the flanking DNA sequence; for example, human *Alu* repeat elements increase origin function (Saëgusa et al., 1993), while some pBR322 sequences inhibit initiation (Lusky and Botchan, 1981).

Thus, the key features of the SV40 origin are the presence of a short reiterated initiator protein binding site, inverted repeats capable of forming cruciform secondary structures, an AT-rich component which includes bent DNA, and the influence of neighbouring flanking DNA regions (especially sequences binding transcription factors) on

origin activity. These components function in the nucleus, as packaged DNA, under the influence of cell cycle control. However, the relevance of SV40-origin based systems to human genomic replication is challenged by the requirement for a viral initiator protein, and the nature of SV40 as a short, circular element with a single replication origin that can be activated more than once per cell cycle.

Polyomavirus (Py), another commonly-studied model system, is the rodent equivalent of SV40. It utilizes the same basic origin structure and function, based on the Py large T-Ag; initiation is if anything even more greatly influenced by flanking auxiliary sequences (Tang et al., 1987), while the Py T-Ag does not have the same high affinity for host p53 as does SV40 (Manfredi and Prives, 1993). The Py origin has also been used to make murine cell episomal vectors (Gassman et al., 1995).

Epstein-Barr Virus (EBV)

Also known as Human Herpesvirus 4 (HHV-4), EBV is the causative agent of infectious mononucleosis, and can infect human lymphocytes in a latent state, persisting in the host cell nucleus as a 172 kb covalently-closed, circular supercoiled episomal element (Griffin et al., 1981; Frappier and O'Donnell, 1991).

EBV has two replication origins, oriL for lytic replication, and oriP for latent persistence. oriL requires the virally-encoded EBV DNA polymerase for activation, and uses a rolling-circle type of replication to amplify the viral genome during its acute infectious spread (Yates and Guan, 1991; Kornberg and Baker, 1992).

However, it is oriP which is of greater interest as a model for human DNA replication, because it is activated in concordance with its host: once and only once per cell cycle. oriP spans 1700 bp, and requires only one virally-encoded protein for activation, Epstein-Barr virus Nuclear Antigen 1, or EBNA-1 (Yates and Guan, 1991). oriP has a bipartite structure, with two functional components about one kilobase apart, termed the dyad symmetry element (DS) and the family of repeats (FoR) (Frappier and O'Donnell, 1991). The DS element consists of a 65 bp inverted repeat, with four associated binding sites for EBNA-1, each of which is an 18 bp imperfect inverted repeat with the sequence GGATAGCATATACTACCC (Williams and Kowalski, 1993). The FoR is composed of 21 repeats of a 30 bp sequence (each including an EBNA-1 site), and has enhancer activity (Frappier and O'Donnell, 1991). Interestingly, the spacing and orientation of the FoR relative to DS is not crucial to oriP function. In addition, the entire oriP region is stably associated with the nuclear matrix (Jankelevich et al., 1992).

In the case of oriP, DNA structure may be particularly important for origin activation. Williams and Kowalski (1993) used S1 nuclease hypersensitivity analysis of oriP-containing plasmids to demonstrate that, even in the absence of EBNA-1, secondary structures are formed at key regions. Specifically, under conditions of negative supercoiling, the DS element (the actual initiation site) can unwind to form single-stranded DNA, with the inverted repeats forming hairpins, while the FoR extrudes DNA cruciforms at a subset of EBNA-1 binding sequences. Both can thus be considered as DNA unwinding elements (DUE), with a function similar to that of AT-rich regions in other origins. EBNA-1 binds first to some of its sites at the FoR, and eventually reaches a critical binding concentration and forms a cooperative homocomplex (reminiscent of initiator binding at *oriC* and oriλ). Only at this point does binding occur at the DS EBNA-1 sites, and a DNA loop with the FoR and DS connected by an EBNA-1 complex is formed (Frappier and O'Donnell, 1991). The DS element provides the initiation site (Niller et al., 1995), but mapping of initiation events, by two-dimensional gel electrophoresis, in one EBV-infected cell line suggests that a second, delocalized initiation zone may be used in rare cases (Little and Schildkraut, 1995). Ensuing steps in activation remain unknown, but it is likely to be host proteins which then act on the unwound DS element.

Replication of EBV is bidirectional, but one fork is blocked at the FoR, meaning that most of the genome is replicated by the other fork, with termination occurring at the FoR (Kornberg and Baker, 1992).

Thus, the FoR element is necessary for initiator binding, but also binds transcription factors, and functions as a replication terminator. All EBV-based plasmids require the FoR, but many mammalian sequences greater than 12 kb in size can apparently substitute for the DS (Heinzel et al., 1991). Bacterial DNA fragments, which have fewer inverted repeats and AT-rich elements, cannot replace DS function unless they are considerably larger (Krysan et al., 1993). A known human replication origin, cloned beside the FoR, is initiated properly at the correct site (Virta-Pearlman et al., 1993). Many details of the initiation process remain to be worked out, but already Sun et al. (1994) have used the EBV oriP to design human episomal vectors, capable of propagating inserts in the 100 kb size range, although only following transfection into EBNA-1 producing human cell lines.

The EBV model is particularly interesting because oriP-containing plasmids have been shown to replicate in S phase, once per cell cycle, and to segregate evenly at mitosis (Yates and Guan, 1991). EBNA-1 and the FoR are key elements in this control (Krysan et al., 1989), but overexpression of EBNA-1 or the presence of extra cloned oriP sequences does not lead to amplification (Yates and Guan, 1991). Hence, host cell proteins (which may serve similar roles in chromosomal replication) must play an important role in limiting

replication at oriP and directing segregation of the two product molecules; such *trans*-acting factors are the subject of active investigation (Kirchmaier and Sugden, 1995).

While our picture of the mechanism of replication initiation in EBV is not yet complete, it is clear that activation at oriP again requires specific initiator protein binding sites. Inverted repeats are present and form secondary structures which may facilitate origin activation, and sites of reduced helical stability exist where unwinding actually commences. An element with transcriptional enhancer activity is a component of the origin. Effects of higher order DNA structure may be involved, including formation of DNA loops, and association with underlying nuclear substructures via matrix attachment regions (MAR).

Bovine Papillomavirus (BPV), like EBV, replicates once per cell cycle during latent infection and is maintained as an episomal circular element (Kornberg and Baker, 1992). The latent origin does not require any stably bent DNA (Schuller et al., 1994), but includes an AT-rich DUE, inverted repeat elements, and a transcriptional promoter and enhancer. Auxiliary sequence effects seem to be especially important, since the viral E2 enhancer protein is indispensable for replication. A T-Ag like BPV protein, E1, is the only other viral factor required, functioning both as an initiator and as a helicase.

Saccharomyces cerevisiae

The budding yeast *S. cerevisiae* has proven to be a convenient model for molecular biologists. A unicellular eukaryotic organism, this yeast packages its small genome (1.5×10^7 bp, 0.5% of human size) into typical nucleosome-based chromosomes within its nucleus. *S. cerevisiae* has 16 chromosomes, ranging in size from 250 kb to 1.5 Mb, which contain little repetitive DNA (Brewer and Fangman, 1991). A complete cell cycle during log phase growth requires about two hours. Importantly, as a non-viral eukaryotic replication model, *S. cerevisiae* must coordinately regulate the firing of multiple replication origins. Interorigin spacing is estimated at 50 - 100 kb, meaning that approximately 200 origins must be activated during each S phase (Brewer et al., 1993), and replication, as in the human, is under the influence of both cell cycle control and the constraints of higher chromatin structure.

Natural *S. cerevisiae* plasmids exist, which take advantage of host proteins to replicate in early S phase. The 2μ plasmid converts a single initiation event into an amplified copy number through the use of a recombinase acting upon large inverted repeat elements to create a shift to rolling-circle replication; copy number is stable at approximately 100 per cell, with a MAR assisting in partitioning at mitosis (Kornberg and Baker, 1992).

Initiation events, however, remain limited to once per origin per cell cycle (Kornberg and Baker, 1992).

The foundation of the relatively advanced state of knowledge of yeast replication origins is the ARS (autonomous replicating sequence) assay, the observation that genomic digest fragments cloned into prokaryotic vectors will sometimes function as yeast replication origins. These ARS plasmids transform yeast at a high frequency, and replicate autonomously, maintained *in vivo* as non-integrated episomal genetic elements (Stinchcomb et al., 1979). While their mitotic instability (caused by an inaccurate partition at mitosis) means that these constructs are lost without selective pressure, and can integrate into the genome during long-term culture, ARS plasmids have proven invaluable in identifying and defining replication origins in yeast (Held and Heintz, 1992).

Sequences capable of functioning as ARS are present roughly every 20 kb in the yeast genome (Fangman and Brewer, 1991), meaning that there are two to five times more ARS than active replication origins. When the neutral-neutral and neutral-alkaline two dimensional gel electrophoresis techniques for mapping initiation sites in yeast became available, it was found that not all ARS elements function as chromosomal origins in their native context (Dubey et al., 1991). However, to date all *bona fide* yeast chromosomal origins function as ARS when cloned into plasmids. Thus, ARS assays are extremely sensitive, but not completely indicative of origin function in any given chromosomal context or cellular state. It may be that some of the constraints imposed by chromosomal structure on origin function are relaxed in the plasmid context (Held and Heintz, 1992).

The relative simplicity of the ARS assay, particularly as a screen to identify new potential replication origins, has meant that this approach has been invaluable in the definition of the structure of functional yeast origins. Deletion analysis has shown that ARS need only be 100 - 120 bp in size, with an 11 bp short core sequence WTTTATRTTTW (A domain) in common among all ARS, but with indispensable flanking sequences (B domains) which may differ greatly (Fangman and Brewer, 1991). Relatively inefficient ARS, conferring reduced plasmid stability, often contain one or two mismatches to the A domain consensus, and their activity can be increased by cloning in other weak ARS elements, or unrelated sequences containing transcription factors, or inverted repeat or bent DNA structures (Kipling and Kearsey, 1990; Hyman and Garcia-Garcia, 1993). In fact, many heterologous DNA fragments, from prokaryotes, organelles, or other organisms, can have ARS activity if they contain a near perfect match to the consensus, in a context of transcriptionally-active or AT-rich flanking DNA (Kipling and Kearsey, 1990; Delouya and Nobrega, 1991). However, "strong" ARS elements are those consistently

activated during every S phase; cloning in extra ARS or other stimulatory sequences cannot enhance plasmid stability any further in these cases (Hyman and Garcia-Garcia, 1993).

Mutational analysis of ARS elements has shown that the conserved AT-rich core consensus element, WTTTATRTTTW, is the only portion where point mutation can knock out origin function (Umek et al., 1989), but divergent flanking B regions 3' to the T-rich strand, often containing transcription factor binding sites, potential DUE, or bent DNA, cannot be deleted (Held and Heintz, 1992; Newlon and Theis, 1993). Marahrens and Stillman (1992) performed an especially thorough analysis of linker substitution mutants across ARS1, a yeast origin from chromosome IV, and proved that the A element was absolutely required, while mutation of any one of three separate B sequences (B1, B2, B3) slightly decreased ARS plasmid stability, and if all three were mutated, origin activity was completely lost. B3 is a known transcription factor binding site (for ABF1); other flanking regions have been postulated to be rich in DUE, bent DNA, nucleosome free regions, attachments to the chromosome scaffold, or extra ARS matches (Diffley and Stillman, 1990). Marahrens and Stillman's data suggest that flanking transcription factor binding sites are important, but extra ARS consensus matches are not; Huang and Kowalski (1993), in a mutational analysis of ARS305 on chromosome III claim instead that DUE activity is the only vital function required in the flanking sequence. Regardless, it seems that when two ARS consensus matches are available, the flanking context makes all the difference as to which origin gets activated (Marahrens and Stillman, 1994).

The determination of the ARS core consensus allowed the application of DNA footprinting to the search for the first eukaryotic genomic initiator protein, although initial attempts served to isolate associated transcription factors and SSB's instead (Li and Alberts, 1992). A breakthrough in the field came in 1992 with Bell and Stillman's isolation of a protein complex, which in a highly purified state bound specifically (in an ATP-dependent fashion) to double stranded DNA at the ARS A element consensus sequence. Christened the ORC, or origin recognition complex, it is composed of 6 subunits ranging in size from 50 to 120 kD. Mutated ARS which confer poor plasmid stability correlate with poor ORC binding. Diffley and Cocker (1992) showed by "nucleotide resolution genomic footprinting" that the same ORC footprint is visible *in vivo*. This ORC complex binds and protects the A element, in an internucleosomal space, and apparently wraps the DNA around itself, keeping it under torsional strain but in double-stranded form (Bell and Stillman, 1992; Diffley and Cocker, 1992). An extension of the footprinting work suggests that the B1 region, 10 - 20 bp away from A, may also be important for ARS recognition (Rao and Stillman, 1995).

Subsequent evidence that ORC mutants arrest as a large, budded cell with a single nucleus helped confirm that ORC functioned in yeast DNA replication (Newlon, 1993), but its possible activities remain many: binding and defining the origin, loading replication factors, directing chromatin assembly and nucleosomal distribution, localizing itself to the nucleus, and interacting with cell cycle control proteins (Bell et al., 1993b). Several of the polypeptides comprising ORC have been isolated (Bell et al., 1993a; Foss et al., 1993; Micklem et al., 1993), but the amino acid sequences deduced largely have been novel, giving few functional clues (Newlon, 1993). ORC2 has, in addition to a weak homology to human topoisomerase I, many potential phosphorylation sites (Foss et al, 1993), and genetic evidence suggests that it may be involved in cell cycle control of origin activation through interaction with CDC46, as a checkpoint preventing premature entry into DNA synthesis (Bell et al., 1993b). ORC6 also interacts with cell cycle proteins, including both CDC46 (MCM5) and CDC6 (Li and Herskowitz, 1993). The sequence of ORC5 suggests that it may have a direct role in binding ARS element DNA (Loo et al., 1995).

The observation that ORC is bound to the ARS site throughout the yeast cell cycle (Diffley and Cocker, 1992) served to beg the question: what activates ORC? An important *in vivo* footprinting study (Diffley et al., 1994) showed that while the A (recognized by ORC) and B3 (recognized by ABF1) elements remain protected in the chromatin context, an extra footprint, appearing in the 20 bp B1 - B2 region, is present strictly between early telophase and very late G1. This footprint is not seen in G0 cells. Interestingly, this pattern of binding exactly fits the model for a "licensing factor," a protein required for activation of initiation which can only access the genome in M phase, and is irreversibly inactivated by replication; consequently, such a protein would serve to limit initiation events to once per cell cycle. The identity and function of the protein responsible for the B1-B2 footprint has yet to be determined, but Stillman (1994) speculated that the yeast "licensing factor" may turn out to be CDC46 (MCM5), and it may function like the λP protein, as a vital part of the "prereplication complex" which must be removed during initiation. Genetic evidence predicts that CDC6 is another component of the prereplication complex (Liang et al., 1995).

The interaction between replication origins and transcription remains an active topic of research. On one hand, it is known that the B elements in ARS sequences are frequently transcription factor binding sites; in fact, Marahrens and Stillman (1992) showed that ARS B elements can be removed but full origin activity maintained if replaced by any of several other defined transcription factor binding sequences. On the other hand, active transcription through the ARS element impairs origin function (Snyder et al., 1988), possibly by displacing ORC. Most ARS-associated transcription factors are at best moderately potent;

engineering a strong promoter beside an ARS decreases replication function if transcription is induced through the origin region, and in *S. cerevisiae* until recently no *in vivo* origin had been found within a transcribed sequence (Tanaka et al., 1994).

Interestingly, in loci controlling expression of yeast mating factors, an ARS element serves as a transcriptional silencer (Rivier and Rine, 1992), and this action requires ORC protein binding (Foss et al., 1993). ORC binds to ARS elements at all four yeast mating loci, but only at one of these sites can any replication origin activity be demonstrated (Bell et al., 1993a). Thus, the role of ORC in transcriptional silencing may be independent of its role in replication, and perhaps even specific to the mating loci (Newlon, 1993).

Overall, the part transcription factors play in yeast origin activation remains enigmatic. They may well function, as in other systems, to alter DNA structure or potentiate DNA unwinding, and while ORC recognition of the ARS may not require transcription factor assistance (Bell and Stillman, 1992), transcription factors may alternatively help to load licensing factor or other proteins associated with activation of the ORC (Marahrens and Stillman, 1992). Considering the proven role of ORC in transcriptional silencing, origin arrangement may help to direct formation of large-scale chromosomal transcriptional domains (Bell et al., 1993a).

There are many more ARS consensus matches present in the yeast genome than there are origins utilized during S phase (Stillman, 1994; Murakami et al., 1995). The subset of potential origins which will be used for chromosomal replication in the next cell cycle can be determined in late S phase, when some newly replicated ARS sequences are bound by limiting quantities of ORC (Rowley et al., 1995), in late M or in the G0 - G1 commitment, when ORC is recognized by licensing factor, or in S phase, when bound origins are actually activated (Diffley et al., 1994). From late S to early M, the ORC complex is bound to a subset of ARS. A prereplication complex forms on ORC in late M with the addition of CDC6 and MCM proteins, licensing those origins that will be used for replication. Action of other cell cycle factors is needed to activate some or all of these licensed origins during S.

Within S phase, chromosomal origins can be activated at different times: a particular ARS sequence may be activated at the beginning of S, or alternatively at some point later on in S phase (Brewer et al., 1993). Chromosomal context can influence the timing of origin firing within S, or even serve to shut down particular origins (Brewer and Fangman, 1991). As discussed previously, cryptic ARS which are non-functional *in situ* can work as efficient origins in the context of an ARS plasmid assay. An ARS cloned into a plasmid may replicate early in S phase, but the same ARS engineered into a position within 5 kb of a telomeric chromosomal end will now replicate late in S phase (Ferguson and Fangman,

1992), suggesting that the subtelomeric chromosomal context promotes late replication. However, non-telomerically located late replicating ARS do exist, and sometimes retain their late replicating phenotype after cloning into an ARS plasmid (Brewer et al., 1993); thus, some aspects of timing control appear to be conferred by small, cloneable elements. Another example of higher level ARS regulation comes from the observation that if two ARS are cloned into one plasmid, within 10 kb of each other, only one, stochastically-chosen ARS is activated on each molecule in each S phase (Brewer and Fangman, 1993). Nevertheless, in the chromosomal context, cryptic ARS are not activated following deletion of the nearest functioning ARS origin.

The sequencing of entire yeast chromosomes will soon yield important information about how ARS sequences function as origins in their chromosomal context. While, as of this writing, the complete sequence of several of the shorter yeast chromosomes has been published (Oliver, 1995), a concurrent survey of functional ARS components has been undertaken only in the case of chromosome VI (Murakami et al., 1995). This 270 kb chromosome contains eight ARS which function as origins, distributed in relatively AT-rich regions that can be 5', 3' or within active transcriptional units; none are located within 13 kb of the telomeres.

Origins of DNA replication, together with telomeres and centromeres, constitute the three necessary *cis*-active elements required for a functional eukaryotic chromosome. Advanced understanding of these components in the *Saccharomyces cerevisiae* system has permitted the successful assembly of yeast artificial chromosomes (YACs), which have since shown their utility in many aspects of molecular biology.

The strategy employed in the construction of a YAC is instructive for those attempting to build other artificial chromosome systems. The story begins in 1979 with the initial identification of short, cloneable *S. cerevisiae* origins of DNA replication, designated as ARS (Stinchcomb et al., 1979). Circular plasmids containing ARS sequences (YAC-R) are capable of replication within yeast and can be maintained indefinitely if the plasmid carries a selectable marker; however, upon removal of selective pressure, ARS plasmids are rapidly lost (unless a recombination event has integrated the plasmid into a host chromosome) because they lack a centromere and do not partition accurately at mitosis. While most ARS elements can confer origin function on a YAC, some are considerably less efficient than others (Hieter et al., 1985). Szostak and Blackburn (1982) found that easily-cloned *Tetrahymena* rDNA 0.7kb telomeres are functional in yeast, and added them to (circular) YAC-R constructs to create linear ARS plasmids (YAC-RT). Such short, linear, acentromeric constructs were no more stable than their circular counterparts (Table I), so the next step was the addition of cloned yeast centromere elements (CEN) to form short,

complete YACs (Dani and Zakian, 1983), maintained at a copy number of 1; these plasmids had a maximum stability of 98% per generation without selection. It was up to Murray and Szostak (1983) to complete the construction of a stable YAC with their discovery that, particularly for linear artificial chromosomes, size itself is an important parameter for stability, and only larger (> 50 kb) constructs can exceed 99% stability per generation. These large YAC constructs require the presence of multiple origin elements for maximum stability, as they are too long to replicate efficiently from a single origin (Dershowitz and Newlon, 1993). The relative stability of these YAC constructs is detailed in Table I. While YACs were built from the ground up by piecing together cloned components, deletion derivatives of native yeast chromsomes were able to confirm many of the conclusions about YAC stability (Surosky et al., 1986). Refinements to the YAC design, such as the addition of markers selectable in other systems (Traver et al., 1989) and tricks for amplifying copy number (Smith et al., 1990) have since been made, but the basic origin, centromere and telomere elements used remain the same.

Table I: Stability of Yeast Artificial Chromosome Constructs [a]

CONSTRUCT	COMPONENTS	FORM	SIZE	STABILITY [b]	COPY #
YAC-R	ARS	Circular	Any	66%	50
YAC-RT	ARS, TEL	Linear	Any	84%	50
YAC-RTC	ARS, TEL, CEN	Linear	10 kb	90%	15
YAC-RC	ARS, CEN	Circular	10 - 50 kb	98%	1
YAC-RTC	ARS, TEL, CEN	Linear	50 kb	99%	1
YAC-RTC	ARS, TEL, CEN	Linear	137 kb	99.85%	1
Yeast Chromosome III		Circularized	350 kb	99.7%	1
Yeast Chromosome III		Linear Wild Type	350 kb	99.996%	1

a. Data from Murray and Szostak, 1983; Hieter et al., 1985; Surosky et al., 1986.

b. Stability = fraction of mitotic events which result in the construct being faithfully transmitted to both daughter cells.

YAC technology has proven to be valuable for molecular biologists, its best known application being as a cloning vector (Burke et al., 1987) with the capacity to carry 1 Mb inserts in which repetitive and certain "uncloneable" elements are much more stable than in prokaryotic systems (Schlessinger, 1990). YACs have hastened genome mapping projects

and have been used to generate complete genomic libraries and chromosome contig maps for lower eukaryotes (Kuspa et al., 1989), and with improvements in yeast strains and computer analysis, mice (Chartier et al., 1992) and humans (Bellanne-Chantelot et al., 1992). YACs can carry intact large genomic regions and complete transcriptional units, and with the addition of mammalian-selectable markers, YACs are being used in large-region complementation studies following fusion of yeast spheroplasts with mammalian cells (Traver et al., 1989). The potential application of YACs in gene therapy and genetic engineering is already being explored; transgenic mice have been made with complete transcriptional unit transgenes (Forget, 1993). For the purposes of genetic engineering, YACs can be further manipulated by yeast homologous recombination, but also face a disadvantageous requirement for integration to achieve a stable transfection of non-yeast hosts.

Knowledge of replication origins in yeast is relatively advanced, due to the availability of ARS plasmid assays to study the genetic elements which define an origin, and convenient mapping techniques such as two dimensional gel methodology which localize initiation sites *in vivo* (Fangman and Brewer, 1991). To summarize the lessons learned from this model system, *S. cerevisiae* represents a eukaryotic organism with hundreds of fixed origins of replication spaced across the lengths of its chromosomes, where the *cis*-active defining origin sequence, the ARS element, coincides with the initiation sites for DNA synthesis (Stillman, 1994). Origins may differ in their efficiency of activation (the chance of a particular ARS being used during one S phase, a function of its binding to ORC and ORC-activating cofactors) and, independently, in their timing of activation within S phase. However, timing of activation may be more a function of the ease of the unwinding step which follows formation of a preinitiation complex (Diller and Raghuraman, 1994). Yeast origins are relatively simple structures. Instead of each being composed of elaborate sequences dictating cell cycle control, limitations in frequency of origin firing, and temporal regulation within S phase, many of these properties are conferred by *trans*-activating factors like ORC (Li and Alberts, 1992), or by the higher order chromosomal architecture of eukaryotes. Degree of chromosomal condensation, origin position relative to other functional chromosomal components, and transcriptional domain arrangement are examples of higher order structural parameters which may influence origin function (Fangman and Brewer, 1991). Indeed, origins may fulfill a concurrent role in transcriptional regulation (Rivier and Rine, 1992). Such higher order effects are still a subject of active research. Other current foci of research in *S. cerevisiae* replication include the development of an *in vitro* replication system to help in the definition of the enzymology of initiation and replication, and the quest to define the function of ORC

components and to find homologs in higher eukaryotes (Bell and Stillman, 1992). Of course, successful strategies in yeast are also being adapted for the study of more complicated, multicellular eukaryotic organisms.

Studies of origin biology in other fungi are not as advanced as in *S. cerevisiae*, but the ARS methodology has worked in *Neurospora*, *Aspergillus*, and *Schizosaccharomyces pombé*. In the fission yeast *S. pombé*, evolutionarily very distant from the budding yeasts, a similar AT-rich ARS consensus has been derived, which is generally found in the context of unusually AT-rich sequences with potential DUE function (Zhu et al., 1994).

ORIGIN STRUCTURE IN HIGHER EUKARYOTES

Studies of simple eukaryotic organisms, including viruses, yeast, protozoans, and slime molds, have revealed a common general origin structure, where the *cis*-active genetic origin control elements exist in a 50 - 1000 bp region encompassing an initiator-binding core and modulating auxiliary elements, very closely related to a DNA unwinding element, wherein lies a specific initiation site (DePamphilis, 1993a). Higher eukaryotes, truly multicellular or "metazoan" organisms, have more complex genomes, typically containing on the order of 6 x 10^9 base pairs spread over 20 - 50 chromosomes (Hamlin et al., 1994), incorporating vast amounts of non-coding DNA including repetitive sequences (which make up 35% of the human genome, for example), numerous introns, and large intergenic spacers. A complete cell cycle requires 16 - 30 hours (including 6 - 9 hours in S phase), with an M phase where the nucleus is entirely disassembled and then reassembled.

Three lines of investigation have suggested that origin function in metazoans is significantly different from that in other organisms, and in fact does not require specific origin DNA sequences, a phenomenon not entirely without precedent, since SV40 T-Ag mutants show low efficiency nonspecific replication initiation events (Umek et al., 1989). First, any exogenous DNA injected into *Xenopus* eggs will be replicated semiconservatively in the early amphibian embryo, with equal efficiency, once per cell cycle. No specific origin sequences are required, and the initiation sites are apparently situated at random (Mahbubani et al., 1992). Secondly, when two-dimensional electrophoretic (2DGE) techniques, used so successfully for mapping initiation sites in yeast, were applied to mammalian origin regions such as the hamster dihydrofolate reductase (*DHFR*) locus, results indicated the presence of a very broad "initiation zone," 30 - 50 kb in size, in which replication intermediates associated with initiation events could be found (Vaughn et al., 1990). Finally, an attempt to apply autonomous replication assays to human cells, using mainly EBV-based vectors containing the FoR portion of oriP together

with cloned human genomic fragments, suggested that any sufficiently long human DNA fragment (>12 kb) possessed the capacity for autonomous replication and therefore had potential origin activity, allowing (as mapped by neutral-neutral 2DGE) initiation at randomly placed sites on the plasmid construct (Krysan and Calos, 1991).

However, the interpretation of these results has been challenged. The completely relaxed origin requirements seen in fertilized *Xenopus* oocytes may be a function of the presence of an extremely high concentration of replication initiators (Umek et al., 1989), and while *Drosophila* embryos behave similarly, neither mouse embryos, nor *Xenopus* cells beyond the cleavage embryo stage maintain the capacity for replication of any exogenous DNA (DePamphilis, 1993b). Another important point is that the original observations were carried out using naked DNA substrates. Subsequently, it has been shown that histone-packaged DNA is initiated more efficiently by *Xenopus* egg extracts (Sanchez et al., 1992), although still at random positions, and more importantly that intact foreign nuclear DNA added to *Xenopus* eggs is replicated starting at its specific chromosomal origin sequences (Gilbert et al., 1993), implying a role for nuclear chromatin organization in origin specificity.

2DGE techniques rely on assumptions about the structure of fragmented replication intermediates, which are present only at very low frequencies when probing for a unique site in a large genome. The key structure, the "replication bubble," could be damaged during the relatively extensive steps required for mammalian genomic DNA purification (DePamphilis, 1993b), leading to artifacts which might confuse the interpretation (Hamlin et al., 1994). Alternatively, if a relatively large region of DNA is unwound in the origin region, unusual structures could be formed which challenge the assumptions required for interpretation of neutral-neutral 2D gels (Benbow et al., 1992).

Several groups have shown that specific, short human DNA sequences can confer the capacity for autonomous replication onto standard prokaryotic vectors transfected into human cells (Frappier and Zannis-Hadjopoulos, 1987; Ariga et al., 1989; McWhinney and Leffak, 1990; Landry and Zannis-Hadjopoulos, 1991), while random genomic fragments lack such replication activity (Bell et al., 1991). Results have been reported (Virta-Pearlman et al., 1993) that contrast completely with Krysan and Calos' (1991) observations supporting non-specific origins; using the same EBV-based vector, Virta-Pearlman and coworkers (1993) found that specific origin sequences were necessary for autonomous replication, and directed initiation events at the same fixed site on the plasmid as is utilized in the origin's chromosomal genomic context, as mapped by a sensitive nascent strand PCR technique. Also, very large episomal circular elements in a mammalian cell line have been shown by Carroll et al. (1993) to replicate synchronously in early S phase, utilizing a

specific fixed origin mapped by the earliest labeled fragment and Okazaki fragment strand switching techniques.

In fact, as Coverley and Laskey (1994) state, the "evidence is now overwhelming that eukaryotic DNA replication normally initiates at specific sites." This evidence includes results from electron microscopy, strand extrusion, earliest labeled fragment, nascent strand PCR, Okazaki fragment strand switching, imbalanced DNA synthesis, 2DGE, and fluorescent in-situ hybridization mapping methodologies, applied to chromosomal genomic loci in *Tetrahymena*, *Physarum*, fly, chicken, mouse, hamster, monkey and human cell systems (Umek et al, 1989; Razin et al., 1990; Benard et al., 1992; Gilbert et al., 1993; Kitsberg et al., 1993b, Berberich et al., 1995). The presence of properly spaced, specific replication origins in the metazoan genome, controlled by specific initiators in turn regulated by cell cycle controls, ensures an efficient completion of S phase, and may help coordinate replication and transcription (DePamphilis, 1993b; Coverley and Laskey, 1994).

Efforts have been made to reconcile the cases of apparent non-specific initiations in higher eukaryotes with the results obtained suggesting that specific origins are required. Most mapping techniques in fact identify the position of the initiation site, as distinct from an initiator-binding, *cis*-acting origin control element (Stillman, 1994), and if there is more than one potential start site near this origin, some techniques like neutral-neutral 2DGE might show a large initiation zone, where other mapping methods localize only the most common initiation site (Gilbert et al., 1993). Several models have been proposed to explain the apparently divergent results. The "Jesuit" model suggests that, as was the case for yeast ARS, there are many more possible origins present in the genome than are required for replication, and a stochastic process, where the joint probabilities of initiator binding, DNA unwinding, and initiation of new base synthesis are influenced by the chromosomal and transcriptional context (which may be relaxed in some assay systems) to yield a final probability of any individual origin firing in any one S phase (Gilbert et al., 1993). In the "strand separation" model, a very large unwound region is opened, an unstable initiation zone is defined by chromatin loop domains, and duplex melting occurs at multiple "microbubble" sites, followed by unwinding over a broad region and initiation at multiple possible sites within this initiation zone (Benbow et al., 1992; Hamlin et al., 1994). However, there is no good evidence as yet for the presence of the kilobase-sized single-stranded regions predicted by this model. A "unidirectional bubble" model suggests that predominant initiation sites exist, but local low efficiency secondary sites may be induced, perhaps by structural stresses on the DNA near a primary origin, to unwind and initiate leading strand priming and synthesis (DePamphilis, 1993a). The related "reformation" model (Coverley and Laskey, 1994) allows for multiple initiations by polymerase

α-primase in the origin region, which are eventually spooled through a polymerase δ holoenzyme complex "replication factory," that replaces the potentially inaccurate first bases laid down.

In fact, many of these models are actually concerned with the events which immediately follow activation of the origin. The question remains: what controls origin activation in higher eukaryotes? Extrapolating from the comparatively well-characterized simpler prokaryotic and eukaryotic systems already described, and adding what is known from the current handful of higher eukaryotic origin regions which have been positively identified and characterized, a functional human replication origin is postulated to be designed like a multicomponent "promoter of DNA synthesis" (Held and Heintz, 1992; DePamphilis, 1993a). Its core would be an AT-rich DUE activated by an as-yet uncharacterized initiator protein recognizing an as-yet uncharacterized binding sequence, which may include unusual DNA secondary structures whose presence may be influenced by the local chromatin state. Auxiliary sequences near the core, binding transcription factors or enhancers, modulate the function of the initiator and DUE either directly, or through alterations in DNA or chromatin structure. The underlying nuclear matrix, a cytoskeletal structure, may serve to ground the initiation and replication apparatus and to define chromosomal replication domains, while the higher order chromosomal context influences the choice of which potential replication origins will be activated.

The DNA Unwinding Element (DUE)

DUE are regions of reduced helical stability where the DNA duplex can be unwound relatively easily, a key step in the process of replication initiation. The thermodynamic cost of duplex melting is partially offset by subsequent association of SSB's (like the human RP-A protein), and is favoured by negative supercoiling, but opposed by the association of duplex DNA with polyamines, histones, and other double strand DNA binding proteins (Kornberg and Baker, 1992). DUE generally coincide with AT-rich regions, since the two hydrogen bonds present per AT base pair are intrinsically easier to break than the three bonds present in GC pairs. However, base stacking interactions are also a factor in the energy cost of duplex melting; consequently, the DUE activity of a sequence is more than a simple function of AT content (DePamphilis, 1993a). A computer program, "Thermodyn," has been developed by Kowalski and colleagues (Natale et al., 1992), which can calculate the free energy for DNA strand separation as a function of sequence. DUE sites predicted using this program have shown excellent correlation with the accepted standard method for DUE identification, P1 nuclease hypersensitivity mapping: P1 endonuclease functions at

neutral pH to cut negatively supercoiled single stranded DNA (Williams and Kowalski, 1993).

Potential DUE sites are not only found in model system origins, but are also found in mammalian origin-enriched libraries, autonomously replicating sequences, and known *in vivo* chromosomal origins. Human (Tribioli et al., 1987), monkey (Rao et al., 1990), and mouse (Dimitrova et al., 1993) origin-enriched libraries formed by cloning the short, newly synthesized nascent strand DNA produced at the beginning of S phase have been found by three independent groups to contain asymmetric clusters of AT-rich sequences, suggesting that AT-rich potential DUE are consistently associated with the DNA at or near active mammalian replication origins. Deletion analysis of human (Wu et al., 1993a) and monkey (Todd et al., 1995) sequences displaying autonomous replication activity after transfection into human cells has shown that AT-rich tracts are part of indispensable 200 - 400 bp minimal origin core regions. AT-rich DUE have been found in the chromosomal replication origins associated with the human c-*myc* and hamster *DHFR* (Caddle et al., 1990), rhodopsin (Gale et al., 1992), and *RPS14* (Tasheva and Roufa, 1994) loci. In the case of c-*myc*, Berberich et al. (1995) have shown that the thermodynamic cost of unwinding 100 bp of the origin-associated DUE is 101 kcal/mol, similar to the cost of unwinding at typical yeast ARS elements, and less than the energy required for unwinding at the SV40 origin.

Thus, it appears that areas of reduced duplex stability, generally including a stretch of tens or hundreds of bases of DNA which can be as much as 75% AT-rich, are common features of all replication origins, including those of the human. DUE sites in model systems are generally found immediately beside an initiator binding site; DePamphilis (1993a) points out that this may mean that only one unidirectional replication fork can be active immediately following initiation, processing away from the initiator complex, at least until the initiator complex can collapse and allow the other fork to replicate through the origin core.

It should be noted that AT-rich sequences may have other functions, besides acting as a DUE. The reduced helical stability of such regions facilitates the kinetics of switching to alternate secondary DNA structures, such as DNA cruciforms, Z-DNA, or triple-stranded DNA (Caddle et al., 1990). Bent DNA sequences are commonly AT-rich (Krajewski and Razin, 1992), as are the consensus elements for nuclear matrix attachment regions, and chromosome scaffold attachment and topoisomerase II binding sites (Tribioli et al., 1987). Furthermore, just as the yeast ARS core initiator binding element WTTTATRTTTW is AT-rich (Gale et al., 1992), mammalian AT-rich sequences may also contain the binding site for an initiator protein (Galli et al., 1992).

Initiator Protein

The initiator protein is the central *trans*-acting factor operating at a replication origin. Examples of such proteins from replication model systems include *E. coli* dnaA, λO, SV40 large T antigen, EBNA-1 of EBV, and ORC of *S. cerevisiae*. Perhaps the latter can be expected to be the closest homolog of a higher eukaryotic initiator. ORC is a rare protein in the cell, and specifically footprints the minimal origin core, although it is responsible for only one of the several footprints found in the origin region (Bell and Stillman, 1992). Higher eukaryotes are expected to employ a similar origin-binding initiator protein, perhaps functioning, as in some model systems, through a cooperative oligomerization or multifunctional complex to wrap the origin DNA around itself and alter local DNA structure, facilitating duplex unwinding (Stillman, 1994). Such a protein might remain bound throughout the cell cycle, waiting for a signal to start its work, under the influence of cell cycle controls including a licensing factor which limits initiation to once per S phase (Diffley et al., 1994).

Several strategies have been employed in the search for a mammalian initiator protein. In the absence of any generally-accepted minimal origin consensus element equivalent to the yeast ARS, some groups have investigated the proteins binding in the neighbourhood of known replication origins or binding to early-replicated nascent strand DNA, while others have looked at factors which bind DNA structures associated with origins, or alternatively at proteins which are bound by certain replication inhibitors.

Bergemann and Johnson (1992) examined a known human origin mapped to a position just 5' of the c-*myc* gene. A 28 kD HeLa protein, christened the PUR factor, footprinted the single-stranded form of a purine-rich region of bent DNA containing the consensus element GGNNGAGGGAGARRRR, also found near other origin and promoter regions. The PUR factor may serve to induce a DNA bend which may alter local DNA structure, but a causative role in initiating unwinding is difficult to envisage for a protein which recognizes single-stranded DNA. PUR may function in replication, or its role may instead concern transcription or recombination.

Iguchi-Ariga and coworkers suggested in 1987 that since anti-c-*myc* antibodies inhibit cellular replication, and c-*myc* binding sites exist in short autonomously replicating mouse fragments, that c-*myc* is a candidate mammalian initiator protein. The c-*myc* protein recognizes a TCTCTTA motif present in a 21 bp minimal autonomously replicating sequence derived from the chromosomal origin region 2 kb upstream of the c-*myc* gene (Ariga et al., 1989). A c-*myc* binding site has also been found in the lamin B2 origin, located in a transcriptionally active region on the short arm of human chromosome 19

(Biamonti et al., 1992), and in an origin mapped in the promoter of human heat shock protein 70 (Taira et al., 1994). The role of c-*myc* in cell physiology has now been convincingly shown to be that of a carefully regulated transcription factor component (Ayer et al., 1993), and thus any function it may have in replication is most likely as an auxiliary factor. The Ariga group has subsequently identified other protein binding activities which specifically bind to their 21 bp minimal autonomously replicating sequence. The 50 kD MSSP-1 protein recognizes an AT-rich WCTWWT motif, has homology to RNA binding factors, and is expressed in G1 and S phases in most human tissues except the brain (Negishi et al., 1994); MSSP-1 may interact with transcription factors like c-*myc*, but its primary function in initiation now seems likely to be as an upstream factor involved in cell cycle control (Takai et al., 1994). A closely-related but separately-encoded 52 kD protein, MSSP-2, binds the same AT-rich sequence from the 5' c-*myc* origin region, and is also hypothesized to be involved in cell cycle control over transcription or replication.

The well-characterized hamster *DHFR* associated 3' origin has also been searched for *trans*-activating factors (Hamlin et al., 1994), and a 60 kD protein, RIP60, has been found which recognizes a sequence 800 bp away from the most likely initiation site, near a region of bent DNA. RIP60 binds cooperatively to ATT repeats, oligomerizes, and establishes looped-out supercoiled chromatin domains (in a fashion similar to EBNA-1) which may contribute to duplex unwinding (Mastrangelo et al., 1993). However, an *in vivo* physiological role for RIP60 in initiation has yet to be demonstrated.

Because there remain few examples of completely sequenced, known mammalian origins with a small, defined critical region, another approach that has been taken is the isolation and cloning of short nascent strands encompassing initiation sites and their flanking regions, which are then probed for proteins which consistently bind this origin-enriched DNA. Using such a strategy, Dimitrova et al. (1993) have identified a 63 kD/65 kD protein doublet bound to replication initiation sequences, while Ruiz et al. (1995) have identified a 150 kD protein which specifically recognizes the minimal region required for the autonomous replication of the nascent strand clone *ors*8, and which copurifies with several replication elongation factors in a HeLa cell fraction capable of supporting initiation and autonomous replication *in vitro*.

The novel approach employed by Pearson et al. (1994a) involves the search for factors which bind specific DNA structures associated with replication origins. An engineered, stable DNA cruciform (Nobile and Martin, 1986) served to identify a novel 66 kD cruciform-specific binding protein (CBP); given the frequent association of inverted repeat elements with replication origins, CBP becomes an immediate candidate for a potential human initiator protein (Boulikas, 1995).

A completely different tactic for the identification of a mammalian initiator involves the study of replication inhibitors. The tyrosine analog mimosine causes a characteristic slow-stop inhibition of replication at the *DHFR* origin, suggestive of a block in the initiation step (Hamlin et al., 1993). This drug can be crosslinked to a 50 kD protein, termed MBP or p50, with a pI of 7.0 (although it may undergo optional post-translational modifications). Protein levels and the mimosine binding activity of p50 are constant throughout the cell cycle (Mosca et al., 1995). It is not yet clear where this protein may function in the initiation of replication, and recent data (Gilbert et al., 1995) showing that mimosine is actually an inhibitor of elongation rather than initiation suggests that p50 probably does not function as an initator after all.

A protease inhibitor, aprotinin, inhibits initiation and replication of isolated nuclei in a semi-*in vitro* system (Coffman et al., 1993a). An aprotinin-binding protein (ADR) greater than 90 kD in size has been identified which stimulates polymerase α- and δ-dependent DNA synthesis; the authors suggest that ADR is vital for initiation *in vitro*, but its exact role in replication has yet to be determined (Coffman et al., 1993b).

In spite of intense efforts, there remains as yet no definitive, proven mammalian initiator protein. The lack, in mammals, of an identified, minimal origin consensus sequence has hampered this work, but more mammalian origins are being identified, cell cycle controls which may interact with initiators are being sorted out, and work is progressing rapidly on some candidate initiator proteins.

DNA Cruciforms

Inverted repeat (palindromic) DNA elements have been implicated in the structure of single-strand replication start sites such as *E. coli pas* sequences, the mitochondrial oriL, parvoviruses, and the G4, Ff, M13, and ΦX174 phages (Kornberg and Baker, 1992; Wang et al., 1993), as well as in replication origins of the ColE1, pSC101, pT181, F and R prokaryotic plasmids, λ phage, *E. coli*, SV40, Herpes Simplex Virus, EBV, BPV, and the yeast 2μ and other fungal plasmids (Almasan and Mishra, 1990; Kornberg and Baker, 1992). Overall, inverted repeat sequences (IR) are quite common in eukaryotes, where they are hypothesized to be formed by polymerase "backtracking" after hitting a block to replication (Cohen et al., 1994). An estimated $2x10^6$ inverted repeat elements are present in the human genome, representing 6% of total DNA. Prokaryotes, where palindromic DNA elements are notoriously difficult to maintain on plasmids (Frappier et al., 1989), typically contain far fewer IR (Kornberg and Baker, 1992).

The DNA cruciform is an alternate secondary structure, similar to the Holliday junction, which can be formed at inverted repeat elements by the extrusion of hairpin loops from both strands. Hydrogen bonds are maintained at the 4-way junction at the base of the stems, as well as on each extruded B-helix stem, but not at a minimum 3 base loop present at the tip of each newly extruded cruciform arm. Cruciform formation is favoured by negative supercoiling, because the intrastrand twists of the cruciform which replace the interstrand twists of the linear duplex remove one supercoil per 10 base pairs (Kornberg and Baker, 1992). Importantly, cruciform extrusion is also favoured by the presence of flanking AT-rich sequence tracts (Bowater et al., 1991), and may facilitate action of neighbouring DUE. Cruciforms are not bound by nucleosomes (Pearson et al., 1994a), and are not as stable as the parent linear duplex because of the thermodynamic cost of the single strand loops, and steric limitations induced at the base. In spite of the kinetic and thermodynamic barriers against adoption of this secondary structure, analysis with specific nucleases (such as T7 endonuclease I) suggests that cruciforms do form transiently at inverted repeat sequences in genomic DNA. In theory, even small inverted repeats can form transiently stable cruciforms, with stems containing four or more base pairs and loops with less than 20 unpaired bases; such potential cruciforms can be identified by computer analysis using the Genetics Computer Group STEMLOOP program (Williams and Kowalski, 1993).

Cruciforms have been demonstrated to form at the replication origins of the pT181 prokaryotic plasmid (Wang et al., 1993), SV40 (Hsu, 1985), and EBV (Williams and Kowalski, 1993). In mammals, early replicated nascent strands and autonomously replicating sequences have been demonstrated to be highly enriched for inverted repeat elements (Iguchi-Ariga et al., 1987; Tribioli et al., 1987; Rao et al., 1990), and the published hamster *DHFR* origin sequence contains a 46 bp poly (AT) repeat (Caddle et al., 1990) capable of forming many alternative cruciform structures.

In order to study the association of cruciforms with replication origins, a tool was needed to probe for cruciform formation *in vivo*. A breakthrough in the field came with the development of monoclonal antibodies directed against an engineered stable cruciform antigen which had been formed by heteroduplexing two DNA strands with differing central IR sequences (Frappier et al., 1987). These anti-cruciform monoclonal antibodies recognize, specifically, the cruciform DNA structure and not double stranded linear DNA, by binding to the four-way junction at the base of the cruciform stem loop in a sequence independent fashion (Frappier et al., 1989). Using this antibody, several lines of evidence have been found to suggest the association of DNA cruciforms with human origins of DNA replication. First, introducing anti-cruciform monoclonal antibodies into permeabilized cells

enhances S phase nucleotide incorporation by a factor of two to seven; genes (like c-*myc)* close to known replication origins are amplified, suggesting that the antibody might bind and stabilize otherwise transient cruciform structures and thereby allow multiple rounds of initiation to proceed from each associated origin (Zannis-Hadjopoulos et al., 1988). Secondly, immunofluorescent labeling of human cells with the anti-cruciform monoclonal antibody demonstrates that cruciforms are limited to S phase, and in fact peak at the very beginning of S (Ward et al., 1990). There are on the order of 10^5 cruciforms per nucleus, matching the estimated number of active human replication origins, and the observed biphasic peaks of antibody binding within S immediately precede bursts in nucleotide incorporation. Finally, anti-cruciform immunoaffinity purification of genomic DNA fragments has created libraries not only enriched for previously-identified origin sequences, but also highly enriched for autonomously replicating sequences (Bell et al., 1991).

Inverted repeat elements can act as protein binding sites (Tribioli et al., 1987); for example, the estrogen, glucocorticoid, and thyroxine steroid-type receptors function as dimers and recognize inverted repeat DNA as their "response element" (Mader et al., 1993). However, the unique structural determinants present at the DNA cruciform present a particularly attractive site for protein recognition (Pearson et al., 1994a). A number of proteins involved in DNA recombination recognize and bind the cruciform motif, as do HMG1, an abundant mammalian DNA-binding protein which helps in establishing higher order DNA structure (Paull and Johnson, 1995), *Sry*, the mammalian sex-determining protein (Harley and Goodfellow, 1994), the *E. coli* HU protein, which facilitates *oriC* unwinding *in vivo* (Bonnefoy et al., 1994), and the pT181 initiator protein, RepC (Wang et al., 1993).

Recently, using a stable heteroduplex cruciform as a probe for cellular proteins binding to a cruciform motif, a novel human protein, designated CBP, has been identified (Pearson et al., 1994a). CBP binding is structure- instead of sequence-specific, recognizing existing cruciforms, but not inducing cruciforms from linear IR DNA. Cellular levels of this protein peak in late G1. Binding of the 66 kD CBP affects the regional DNA structure, increasing the binding of some other proteins to neighbouring linear DNA regions. Thus, the properties of the CBP make it an attractive candidate for a potential human initiator protein.

Although cruciforms are not, apparently, closely associated with the nuclear matrix (Ward et al., 1991), at least one author (Boulikas, 1995) has presented a model of the mammalian replication origin with a cruciform as the core, initiator-binding element, activating a neighbouring DUE in conjunction with transcription factors to initiate replication.

DNA direct repeats could also alter their structure to produce single-stranded regions for replication priming, according to another model (Wanka, 1991), but only at a much higher energetic cost.

Other Secondary Structures Associated with Origin Activity

Several other potential DNA secondary structures have been associated with replication origins, including bent DNA, triple-stranded DNA, and Z-DNA.

Bends in the longitudinal axis of the double helix can occur in any 50 bp or longer sequence, containing in one strand runs of three to six adenines repeated with a ten to eleven base pair periodicity, which causes a narrowing of the minor groove at the same place in consecutive helical turns (Kornberg and Baker, 1992). Such a bend could facilitate binding of proteins like topoisomerases, or alternatively DNA bending could itself be induced by the binding of other proteins such as the PUR factor (Bergemann and Johnson, 1992). "Antibent" sequences, oligo d·A with a six to eight base periodicity, resist protein-induced bending and may function as DUE when placed next to a bent sequence. The presence of DNA bending is assayed using two dimensional electrophoresis, room temperature agarose followed by conformation-sensitive 4°C acrylamide gels (Caddle et al., 1990; Todd et al., 1995). Bent sequences have been found near the λ and SV40 origins, in ARS elements replicating in yeast (Nakajima et al., 1993), in the minimal deleted nascent strand *ors*8 monkey autonomously replicating sequence (Todd et al., 1995), and in the regions of the hamster *DHFR* (Caddle et al., 1990), hamster *RPS14* (Tasheva and Roufa, 1994), human β-globin (Wadakiyama and Kiyama, 1995), and human c-*myc* (Bergemann and Johnson, 1992) replication origins. At this point, however, no role for bent DNA has been proven in the mechanism of origin activation, and at least in the case of BPV, bends near the origin are functionally dispensable (Schuller et al., 1994).

Triple-stranded DNA is an alternate conformation for long homopurine : homopyrimidine sequences. A denatured portion of the polypyrimidine tract loops back and forms Hoogsteen hydrogen bonds with the purine rich strand of an A-form duplex (which has pre-existing Watson-Crick hydrogen bonds with the upstream portion of the homopyrimidine sequence). The reversed strand polarity cancels out one negative supercoil per eleven bases of triplex DNA, but the backbone distortions and single-stranded leftover homopurine DNA make this a very high energy state. Triplex DNA is only formed *in vitro* in plasmids with extensive supercoiling, at an acidic pH that protonates cytosine, a prerequisite for Hoogsteen hydrogen bonds. Caddle et al. (1990) have shown that the DHFR origin has a potential triplex that can form *in vitro* at pH 5.2, and suggest that if such

a structure could form transiently *in vivo*, it would create a single-stranded region for priming of DNA synthesis. Polypyrimidine tracts have also been reported in the human c-*myc* and hamster rhodopsin origins (Benbow et al., 1992), and act as preferred start sites for mammalian polymerase α-primase *in vitro* (Suzuki et al., 1993). There is no evidence, as of yet, that triple-strand structures form *in vivo* at mammalian origins under physiological conditions.

Z-DNA, a left handed helical structure less stable than the right handed B-helix, removes one negative supercoil per six base pairs, but the B-Z transition region between helical types is thermodynamically expensive (Kornberg and Baker, 1992). Z-DNA can be formed at physiological pH at a $(GC)_5(AC)_{18}$ tract near the *DHFR* origin (Bianchi et al., 1990), but has not been found as a regular feature at other replication origins.

Transcription Factors

Transcription and replication are broadly similar, both involving the unwinding of the genomic DNA duplex and the processive addition of new nucleotide monomers to a growing complementary nucleic acid strand. However, these two fundamental cellular processes are physiologicially coupled in more ways than one; not only are RNA primers used to begin each nascent DNA strand, but also the *cis*-acting auxiliary components which enhance core origin function have turned out to be transcription factor binding sites (DePamphilis, 1993a). In fact, the replication origin can really be looked on as a "promoter" of DNA replication.

Eukaryotes use three separate RNA polymerases: RNAP I for rRNA transcription, RNAP II for mRNA, and RNAP III for tRNA. Each requires its own series of transcription factors to recognize a promoter site, since the association of an RNA polymerase with its template is intentionally weak, to facilitate processivity. Typical genes coding for mRNA have several promoter components: a CA site at position +1, and upstream of the start site a "TATA" box with consensus TATAWAW around position -25 (recognized by TFIID), several other upstream activating sequences like the CP-1/CAAT box and Sp1 GC box, and often at some distance from this promoter region, a transcriptional "enhancer" which binds often powerful, tissue-specific factors like AP-1, NF-κB, or steroid receptors. The minimal core promoter binds the common TFII components which interact directly with RNAP II, while upstream sequences, acting in a more gene- or tissue- specific fashion, are hypothesized to function in one of several ways: by loading the TFII proteins, by displacing histones, by altering regional chromatin

structure to facilitate transcription, or by forming DNA loops which place critical regions in closer apposition.

How might transcription factor binding be related to the function of origins of DNA replication? DePamphilis (1993a) has proposed several models. First, RNA transcripts promoted by origin-associated transcription factors can serve as primers for DNA replication. While this mechanism is utilized in prokaryotic systems like the T-odd series of phages and ColE1 plasmids, the only documented case of eukaryotic transcription-primed replication occurs at the mitochondrial O_H, which builds on a transcript initiated at the mitochondrial light-strand promoter. Alternatively, transcription factors can function by directing the loading of the initiator complex (as is the role for host NF-I in adenovirus replication or viral E2 in BPV initiations) or other components of the replication apparatus, such as the mammalian SSB RP-A, which can be recruited to viral origins by the E2 (Li et al., 1993), p53, VP16 or GAL4 transcription factors (Coverley and Laskey, 1994). For this reason, DePamphilis (1993a) has presented a model of a eukaryotic replication origin showing possible direct interactions between auxiliary-sequence-binding transcription factors and both the initiator complex and replication apparatus. However, in a third model these interactions may not be based on the loading of important proteins, but instead on activation of the bound initiator, such that transcription factors facilitate the later steps of initiation, duplex unwinding and the addition of the first new bases. A final proposed mechanism for the activation of origins by transcription factors suggests that they function in preventing chromatin-structure mediated repression of origin function, by displacing histones or counteracting chromatin hypercondensation. This model explains that distant enhancers can facilitate origin activation at the chromosome domain level, and is generally consistent with the observation that euchromatin replicates much earlier in S than does heterochromatin.

These models make predictions about the organization of auxiliary binding sites (DePamphilis, 1988). If transcription factors act by direct interaction with the initiator, they would be expected to lie in a fixed orientation close to the origin core, but if their action is through facilitating unwinding, they may bind near the DUE rather than the initiator. Alterations to chromatin structure could be possible even with binding sites several kilobases away.

Unfortunately, most of the studies probing the role of particular transcription factors in origin function have been carried out in viral systems, which may be poor models for what goes on in the mammalian genome (Biamonti et al., 1992), and which have sometimes given contradictory results. Guo et al. (1989) have shown that in SV40, the host AP-1, Sp1, and NF-I transcription factors act at the auxiliary sites by facilitating origin

unwinding, and not by priming active transcription, assisting in the loading of T-Ag, or displacing histones. In BPV, the situation seems different; evidence suggests that the E2 enhancer displaces histones to increase the specific interaction of the E1 initiator with the origin (Li et al., 1993). Different mechanisms are suggested, for example, when SV40 is studied *in vivo* versus *in vitro* (Coverley and Laskey, 1994). Clearly, the role that transcription factors do play at the origin is still a matter of active debate. It is agreed, however, that the ability of transcription factors to stimulate replication does not correlate quantitatively with their ability to stimulate transcription (Bennett et al., 1989; DePamphilis, 1993a). From studies in yeast, it is known that relatively "weak" transcription factors are often the ones present at origins, including negative regulators which decrease or silence transcription, and that the function of a factor is to some extent interchangeable with that of other transcription factors (Marahrens and Stillman, 1992).

A consistent observation, in mammals, is that transcriptionally active euchromatic DNA replicates at the beginning of S phase. All of the mammalian chromosomal replication origins isolated and sequenced to date have come from transcriptionally active regions of the genome. In repetitive ribosomal DNA regions, replication initiates in the intergenic spacer between rDNA genes. The functional components of the c-*myc* associated origin which confer autonomous replication capacity onto plasmids are contained within a 2.4 kb genomic fragment located immediately 5' to exon 1, containing all key portions of the c-*myc* promoter domain (McWhinney and Leffak, 1988). An origin activated at the beginning of S has been mapped at high resolution by Biamonti et al. (1992) to a position on the short arm of human chromosome 19 (19p13.3) centered on a 600 bp spacer separating the active nuclear lamin B2 gene from a second, unidentified transcript. Recently, it has become clear that in humans, origins may reside on actively transcribed DNA, because human cDNA clones are enriched for autonomously replicating sequences (Wu et al., 1993a). Chromosomal *in vivo* origin activity for one such sequence has been demonstrated by nascent strand PCR (Wu et al., 1993b), and mapped to chromosome 6, region q22-qter (Shihab-El-Deen et al., 1993). Two groups have used the nascent strand PCR technique to map an origin to the murine immunoglobulin heavy chain intronic enhancer (Eμ), between the J and Cμ regions of this important locus (Ariizumi et al., 1993; Iguchi-Ariga et al., 1993). Interestingly, the Eμ origin is only able to confer autonomous replication onto plasmids transfected into B lymphocytes, and is used during S phase at least ten times more often in B cells than in fibroblasts; nonlymphoid cells which do not have a transcriptionally-active IgH gene replicate this locus from a downstream initiation site (Ariizumi et al., 1993).

Timing of replication within S phase is not controlled by the Eμ enhancer and origin (Ariizumi et al., 1993); Stillman (1994) notes that in the case of the human β-globin locus, studies have shown that the Locus Control Region, which is not a functional component of the β-globin-associated replication origin, not only activates transcription in the globin gene cluster but simultaneously controls timing of origin activation within S, suggesting that timing of activation may be a function of higher order structure, while usage of a replication origin is determined at a local level.

While most authors have assumed that it is transcription which modulates replication, it may well turn out to be the other way around in some cases. Coverley and Laskey (1994) suggest that replication of a gene leaves it in a transient state, before chromatin packing and condensation is reestablished, highly conducive to active transcription, and thus replication may serve to activate genes otherwise repressed by chromatin structure. In the case of Eμ, the enhancer may function first by activating its associated replication origin, which then leads to a secondary change in chromatin structure necessary for VDJ recombination and active transcription (Ariizumi et al., 1993). Boulikas (1995) has suggested that all transcription enhancers function as "replication enhancers," and that the resulting active origin is mechanistically required for transcriptional enhancement. Another model suggesting that replication origins control transcription in neighbouring loci was proposed by Smithies (1982), who hypothesized that the direction of replication through a gene could "open" that gene for transcription if the replication fork proceeds in the same direction as transcription, or "close" the gene if it is replicated in the reverse polarity from its coding. For example, the c-*myc* and β-globin origins are located 5' to their genes and support this model, while *DHFR* does not. As an aside, a consequence of antitranscriptional replication may be a "fork collision," and if this stalls or terminates replication, the result may be an apparently unidirectional or imbalanced bidirectional replication process from the responsible origin (Ariizumi et al., 1993).

All things considered, the mechanism of the observed association between replication and transcription is still a subject of intense study. The role that transcription factors play in origin biology is not fully understood, and may vary at different origins. The biological reasons for the interconnection of these fundamental processes remain to be worked out even at the basic level of cause-and-effect.

Known mammalian replication origins display many different auxiliary transcription factor binding elements. The murine IgH intronic enhancer Eμ contains sites for thirteen different transcription factors, including Oct-1, 2, and 3, a series of NFμ elements, USF, and others (Staudt and Lenardo, 1991); a minimal autonomously replicating sequence derived from Eμ contains a TNATTTGCAT motif recognizing Oct-1 and NF-III (Iguchi-

Ariga et al., 1993). Oct-1 is a relatively common finding in origins; a near match to the consensus binding site is associated with the minimal *ors*8 autonomously replicating sequence (Todd et al., 1995), and two perfect matches, together with an AP-1 site, are found in the 3' *DHFR* origin (DePamphilis, 1993a). The hamster rhodopsin origin, in contrast, lacks Oct-1, but has AP-1 and NF-III, as well as a CP-1/CAAT box (Gale et al., 1992). The hamster RPS14 origin is similar to rhodopsin, with NF-I instead of NF-III, and additional sites for p53 and Sp1. The 5' c-*myc* origin contains basic promoter elements like the TFIID TATA box, and sites for NF-I as well as for other proteins like *myc*, PUR and MSSP which may be either initiators or transcription factors (Negishi et al., 1994; Berberich et al., 1995). The 19p13.3 lamin B2 origin also contains a site for the *myc* protein, as well as for Sp1 and suspected enhancers (Tribioli et al., 1987; Biamonti et al., 1992).

Consensus Sequences Present at Mammalian Replication Origins

The many transcription factors and potential initiator proteins associated with origin regions illustrate an important problem in the field, the definition of what functional *cis*-acting DNA consensus sequences are present at higher eukaryotic replication origins.

Computer methods are available which identify common elements in different sequence fragments, by using a weight matrix technique to define the probability of each base in a sequence being conserved across a group of aligned fragments (Staden, 1984). Such methodology has proven useful in identifying consensus sequences for splice junctions, ribosome binding sites, and conserved promoter elements, where the RNA or cDNA positions relative to genomic DNA provide an obvious landmark for the alignment of different fragments.

Short consensus elements which function as protein binding sites have also been identified precisely and more directly through DNA footprinting techniques, which can be enzymatic or chemical and probe interactions *in vitro* or even *in vivo* (Rein et al., 1994). Similar information on the DNA sequences required for factor recognition can be obtained by combining bandshift assays or southwestern blots with the synthesis of selective oligonucleotide mutants.

Several sequence elements thought to be associated with replication origins have been defined with the above techniques. AT-rich ten base pair scaffold attachment region (SAR) consensus sequences, which may mediate association of DNA with the chromosome scaffold (termed by others the nuclear matrix), were identified by testing labeled fragments for binding to an immobilized *Drosophila* scaffold preparation (Gasser and Laemmli,

1986). When later compared with a number of sequences found to attach to the nuclear matrix in other organisms, only a very short perfect common consensus (ATTA or ATTTA) could be derived, which is also found in homeobox domains and in the Oct-1 transcription factor binding site (Boulikas, 1992). The pyrimidine-rich primase start site consensus was determined by applying an *in vitro* primase assay to synthetic oligonucleotide substrates (Suzuki et al., 1993). The recognition of ATT repeats by RIP60 was established by DNase I protection assay (Mastrangelo et al., 1993), while the HeLa PUR factor consensus GGNNGAGGGAGARRR was determined by bandshift assay and computer analysis of GenBank sequences (Bergemann and Johnson, 1992). Following deletion studies that established a minimal length autonomously replicating sequence, southwestern blotting and DNase I protection assays were used to establish binding sites for the *myc* protein and MSSP factors (Ariga et al., 1989; Negishi et al., 1994).

When the complete sequence of an origin region has been obtained, it is a relatively simple matter to look for matches to these consensus elements, any of dozens of known transcription factors, or to repetitive DNA elements or other homologous sequences in GenBank. As a result, the lamin B2 origin has been matched to the topoisomerase II recognition site, the Sp1 transcription factor, and the *Alu* repetitive element (Tribioli et al, 1987). Nascent strand *ors* clones are enriched for SAR, yeast ARS, the β-globin control element CACCC, and the B1 intron (Rao et al., 1990). The hamster rhodopsin origin region contains SAR, ARS, *Alu*, B1, and transcription factor sites like AP-1, Oct-1, and CP-1/CCAAT (Gale et al., 1992), while the mouse 3' adenosine deaminase origin has PUR and RIP60 sites, in addition to NF-1, Oct-1, p53, ARS, and matrix attachment matches (Virta-Pearlman et al., 1993). The minimal form of the 343 origin carries SAR and yeast ARS matches (Wu et al., 1993a).

Few consensus elements have shown up consistently, but when a match is found, a second question must be asked which is, unfortunately, all too frequently ignored: what is the statistical significance of having found this match? When no functional role in origin activation has been definitively assigned to a particular binding site, it is especially important to understand the probability of a match occurring strictly by random chance in a sequence of the same length, if one wishes to imply that the presence of a consensus element is likely to be biologically important. Assuming that the elements to be searched for are defined *a priori*, the significance of a match can be determined by finding the expected number of times a particular base sequence will occur by chance in a random DNA sequence of the same size and base composition as the analyzed region (Rao et al., 1990). A 95% confidence interval can be assigned to this value, and the case for the functional relevance of a match is made much stronger if it is present significantly more often than

would be expected by chance. Because genomic sequences are not random, the resulting values should have taken into account the overall AT-richness of a sequence (Rao et al., 1990).

The real goal of origin sequence analysis is the identification of a common binding site for an as-yet undetermined mammalian initiator protein, which could then be used to isolate that protein, the strategy that proved successful in yeast. However, as Rao et al. (1990) state, "an enormous difference exists between looking for a previously established consensus in any stretch of DNA and deducing a new consensus from the matching of two [or more] stretches of DNA." The latter problem is far more difficult, and especially so in the case of a search for origin components, where the first problem in aligning two sequences is knowing which complementary orientation to choose (when key transcription-related consensus elements were determined, the difference between the "coding strand" and "template strand" was obvious). Multiple sequence alignment algorithms (Higgins and Sharp, 1989) could separate strands into two groups for comparison, but only if the consensus among them is large relative to their length. Secondly, it is unclear where sequences for comparison should be aligned, in the absence of any landmark for exact origin position. This problem can approach intractability when dealing with large sequences, but is manageable where minimal origin elements can be defined. Third, if the key origin components are structure- instead of sequence-specific, analysis for common base sequences will fail. Fourth, there is the possibility that origins may fall into a number of different subfamilies which have different core elements, much as they appear able to use any of a number of different transcription factors. Finally, the consensus may turn out to be a short, degenerate element like the yeast ARS (Kipling and Kearsey, 1990), which in the absence of functional data can only be recognized with confidence after computer analysis of a very large number of sequences. Because DNA is composed of only four bases, all sequences comparisions have a default expected homology of 25%.

Consensus mitochondrial origins have been relatively easy to determine, because there is a defined strand polarity, transcriptional start site, and primer cleavage site (Almasan and Mishra, 1990). Identification of the yeast ARS consensus WTTTATRTTTW was facilitated by the availability of a large number of minimal origin sequences defined by ARS plasmid assay. One group (Dobbs et al., 1994) has attempted a thorough computer analysis of relations between published sequences of six higher eukaryotic replication origins, those associated with *Tetrahymena* rDNA, *Drosophila* chorion gene, chicken histone H5, hamster *DHFR* and rhodopsin, and human c-*myc*. They find common elements include potential DUE, pyrimidine-rich primase start sites, SAR, and various transcription factors, and derive a degenerate consensus sequence present in each:

WAWTTDDWWWDHWGWHMAWTT. While they show that this particular sequence is present more commonly than would be expected by chance in a random sequence, that claim is only relevant when the consensus is defined *a priori*. The more complicated question of what the chances are of any 21 bp consensus of equivalent degeneracy falling out of any comparison of six random sequences of the same size (i.e. the significance of the fact that they found a consensus) is left unaddressed. Also, the identified sequence could be an associated functional element, like a DUE or SAR, rather than an initiator binding site.

Ultimately, biological proof of the relevance of any identified consensus element is required. In the search for a mammalian initiator-binding element, it would be helpful if a large number of relatively short origin sequences from a single organism could be identified. Until then, the best way to predict the presence of origins in a sequence may be through the search for clusters of origin-associated elements (DUE's, potential cruciforms, transcription factors, SAR's).

Nuclear Matrix

The nuclear matrix, an entity first recognized in the late 1950s, is the intermediate-filament-like cytoskeletal meshwork remaining in a nucleus after the digestion of all chromatin (Georgiev et al., 1991). Through periodic attachments to this insoluble proteinaceous matrix, nuclear DNA is organized into 10 - 100 kb topologically-constrained loop domains. A related structure, the chromosome scaffold, may represent a collapsed form of the matrix present during mitosis.

A "matrix fraction" of subcellular material can be obtained by high salt or detergent extraction of nuclease-treated, washed isolated nuclei, and contains essentially all cellular DNA and RNA polymerase activity, as well as primases, topoisomerases, DNA methylases, and transcription factors (Mah et al., 1993); thus, the processes of transcription and replication appear to occur in close association with the nuclear matrix, perhaps to allow for management of loop topology. Thorough microscopic examinations have shown that human DNA replication occurs in distinct "replication factories" on the nuclear matrix (Hozak et al., 1993), and the exogenous DNA of invading viruses also assembles onto the matrix while replicating (Jankelevich et al., 1992; Chaly and Chen, 1993).

In prokaryotes, replication origins are grounded by fixation to the plasma membrane, and there is abundant evidence that eukaryotic origins are similarly grounded, in this case through attachment to the nuclear matrix (Razin et al., 1990). The earliest replicated DNA in the cell, the extrudable nascent strands, is bound by the nuclear matrix,

and the little DNA remaining in the matrix fraction is highly enriched for autonomously replicating sequences, to the point where it has been estimated that about half of all chromosomal matrix attachment regions include replication origins (Georgiev et al., 1991). Matrix attachment may be a key characteristic which defines a sequence as an *in vivo* origin, since in the otherwise relaxed *Xenopus* oocyte system, only nuclear matrix-packaged DNA has been found to initiate correctly at specific origin sequences (Gilbert et al., 1993).

Since the earliest replicated DNA stays fixed to the matrix, it is postulated that there is a constitutive attachment of origins to the nuclear matrix (Razin et al., 1990). Each chromosome may be attached to the nuclear membrane at one point, and to the nuclear matrix at many other points, including all of its replication origins (Georgiev et al., 1991). A model of a fixed replisome remaining attached to the nuclear matrix as the DNA is spooled through it and replicated has direct microscopic support (Hozak et al., 1993); hundreds of replication factories are visible during S phase, each bound to the matrix and containing around 20 growing forks, initiated at 10 matrix-bound origins. Neighbouring matrix attachment regions on the chromosomal DNA are modeled to be neighbours on the nuclear matrix, meaning that neighbouring chromosomal origins may end up clustered together in a single region of the matrix which will later become a replication factory; each neighbouring origin pair subtends a chromatin loop domain, and the group of clustered origins on the matrix constitutes a larger chromosomal replication timing domain (Wanka, 1991). The loop domains subtended by origin pairs are from 10 - 100 kb in size, and are hypothesized to border the 5' and 3' regions of genes, thereby simultaneously defining a transcriptional domain (Boulikas, 1992).

How are matrix attachments mediated, and how is this reflected in the DNA sequence at the origin? Working in the *Drosophila* system, Gasser and Laemmli (1986) identified an A-rich (strict match = AATAAAYAAA, loose match = WADAWAYAWW) and a T-rich (strict match = TTWTWTTWTT, loose match = TWWTDTTWWW) "SAR" consensus in sequences bound to the chromosome scaffold (nuclear matrix). These sequences also frequently contained matches to the topoisomerase II consensus binding site GTNWAYATTNATNNR (core cleavage sequence = WAYATT); topo II is a major scaffold protein component and functions to control the topological state of DNA loops, with peak activity during G2/M chromosomal condensation. The only absolute "MAR" consensus among a larger number of matrix attachment regions from different higher eukaryotes is ATTA or ATTTA, a very common sequence (Boulikas, 1992). Inverted repeats are also present in matrix-bound DNA (Boulikas and Kong, 1993), but anti-cruciform antibody immunofluorescence does not detect cruciform formation at MAR (Ward et al., 1991).

Matrix attachment regions have been demonstrated to exist at the EBV oriP (Jankelevich et al., 1992) and at known higher eukaryotic replication origins, including the hamster *DHFR* origin (Dijkwel and Hamlin, 1988), mouse IgH intronic enhancer origin (Ariizumi et al., 1993) and the human 343 (Wu et al., 1993a) and β-globin (Boulikas, 1993) origins. SAR binding site matches have been found in the human lamin B2 origin (Tribioli et al., 1987) and in several human cDNA with autonomous replication activity (Wu et al., 1993a), but examples of mammalian autonomously replicating sequences which lack SAR matches or detectable matrix binding have also been found (Wu et al., 1993a; Mah et al., 1993).

In spite of the identification of short consensus elements associated with matrix attachment sites, the role of cytoskeletally-mediated higher order structures in origin function is difficult to investigate experimentally, as tools like DNA cloning, sequencing and footprinting are of limited use at that level. Nevertheless, as the DePamphilis group has stated (Gilbert et al., 1993), "site-specific initiation of DNA replication in metazoan cell chromosomes appears to be determined by nuclear organization as well as DNA sequence."

CHROMOSOMAL ORGANIZATION

The association of matrix attachment regions with replication origins serves as a reminder that human DNA is organized into higher order structures, up to the level of the visible metaphase chromosome, and that origins must fill a particular role in the biology of the complete chromosome.

Interphase chromosomal DNA is wrapped around core histones as a 10 nm fibre of nucleosomes, which are sequentially stacked into a 30 nm solenoid through interaction with linker histones such as H1, and then organized by high mobility group (HMG) proteins, the nuclear matrix (in part via its topoisomerase activity) and heterochromatin-associated proteins into packed loop domains and major chromosomal subdomains (Clarke, 1992). Taylor, in 1960, first noted that large chromosomal subdomains are replicated at defined, reproducible times within S phase, with the transcriptionally-active euchromatin regions replicating at the beginning of S, and the more condensed heterochromatin replicating towards the end of S. Modern techniques have allowed a more careful dissection of chromosomal subdomain replication in the human S phase (O'Keefe et al., 1992): 30 minutes into S, all visible replication occurs in euchromatin; at 2 hours, euchromatin and bordering heterochromatin replicates; at 5 hours, perinucleolar heterochromatin begins to replicate; at 7 hours, interconnecting electron-dense heterochromatin is active; and near the

end of the 9 hour S phase only a few large regions of heterochromatin still show ongoing replication. The degree of chromatin condensation is thus intimately associated with both replication and transcription. Moving an early-replicated origin into a region of heterochromatin converts it into a late-replicating state (Ferguson and Fangman, 1992); moving an expressed gene from euchromatin into heterochromatin turns it off (Brewer et al., 1993); and genes which are active and early-replicating in some cells are replicated late in S in other cell types which do not express the associated gene (Kitsberg et al., 1993b). In addition, drugs (5-azacytidine analogs) which inhibit chromosome condensation shift late-replicating heterochromatin to an early-replicating state, and can thereby activate expression of some genes (Haaf, 1995).

The arrangement and activation of replication origins along the length of a chromosome may thus be related to overall chromosomal organization and transcriptional activity. The basic unit of replication on the chromosome is the replicon, the region of DNA replicated from a single origin. Each chromosome consists of perhaps 1000 replicons, with their origins clustered into synchronously-activated groups of 20 or so. Origin arrangement may have a role in determining areas of chromosomal instability, including amplified regions, or sequences which may form extrachromosomal circular elements (Gaubatz and Flores, 1990). The association of origins with the higher order domain structure and transcriptional activity of their replicons may portend a role in the phenomenon of genomic imprinting. Cell cycle controls limit replication to S phase, but the mechanisms of temporal activation of origins within S are largely unknown, although it has been recognized that the number of activated origins is significantly higher in embryonic cells (which have a shorter S phase), and that the overall pattern of replication in S shows biphasic peaks in DNA synthesis (McAlear et al., 1989). Finally, origins function alongside centromeres and telomeres as the three necessary *cis*-acting components required for the replication and maintenance of a chromosome.

The Concept of the Replicon

The classic fibre autoradiography work of Huberman and Riggs (1968) showed that DNA replication in higher eukaryotes is initiated at a large number of discrete origins spread over the length of each chromosome. At least 90% of origins promote bidirectional replication, with two growing forks moving in opposite directions away from an origin, before eventually colliding with forks initiated at neighbouring origins; the region of the chromosome replicated from a single origin is termed a replicon (Razin et al., 1990). There are approximately 10^5 total replication origins in the human diploid genome, with an

average replicon size (and interorigin spacing) of about 100 kb. While prokaryotic polymerases lay down approximately 1500 bases per second, replication forks in eukaryotes progress at a relatively constant rate of 50 bases per second, an apparent handicap counterbalanced by the hundreds or thousands of origins utilized on each chromosome (Kornberg and Baker, 1992). The length of S phase is related directly to replicon size and inversely to origin number, but the overall length of S phase is longer than would be predicted considering only the number and size of replicons and the rate of fork progression, suggesting that delays in the activation of origin clusters also constitute an important parameter (Diffley and Stillman, 1990).

A vital concept is that "there are more potential replicators [origins] in the chromosome than are actually needed, perhaps to allow for selected use of different replicators during development" (Stillman, 1994). Cells of the early embryo employ approximately ten times as many origins during their abbreviated S phases, implying that there are a number of weak, cryptic or embryo-specific origins in the genome. The embryonic cell may activate weak origins by establishing very high levels of initiator protein, or alternatively may activate cryptic origins by synthesizing extra subtypes of initiator. Interestingly, transformed cells may also activate cryptic origins not utilized in their parent cell lineage.

Particular chromosomal origins can be efficient, activated in virtually every S phase, or inefficient. Chromosomal orgins can also be classified as temporally early (activated at the beginning of S phase) or late (activated hours after the start of S). Fibre autoradiographic studies of clonal cell populations (Amaldi et al., 1973) and *Drosophila* polytene chromosomes (Umek et al., 1989) suggest that for origins located in regions of similar chromosomal condensation, the efficiency and timing of activation is at least partly a stochastic process. The likelihood and timing of origin firing is a function of both the initiator binding affinity and the initiator activation efficiency at that particular origin (Diller and Raghuraman, 1994). The binding affinity is influenced by the exact *cis*-active binding sequence present and by hypothetical, possibly tissue-specific variations in cellular *trans*-acting initiators. Activation efficiency refers to the ease of unwinding at that origin, which may be a function of DUE sequence, chromatin condensation, transcriptional status, and accessory factors in flanking regions; the latter three parameters may again be tissue-specific.

The timing and choice of chromosomal origin activation is intimately connected with transcriptional status. For example, the β-globin associated origin initiates early in S in cells which express globin, but late in S in those which do not (Kitsberg et al., 1993b). A single gene may be replicated from different origins depending on whether it is transcriptionally

active or not, and the direction of replication through a transcribed gene tends to be the same as transcription (Razin et al., 1990). Studies on the Eμ murine IgH origin suggest that timing is not controlled by *cis*-acting elements present within 15 kb of flanking sequence (Ariizumi et al., 1993), leaving open the possibility that timing of origin initiation may be controlled by a long-range *cis* element, like a Locus Control Region enhancer (Kitsberg et al., 1993b), or by the context of higher order chromosomal structure (Diffley and Stillman, 1990).

What end could be served by regulating the choice and timing of replicon activation? Ariizumi et al. (1993) suggest that the activation of the IgH intronic enhancer could secondarily cause structural changes which allow access of transcription factors and recombination enzymes to the IgH locus. More generally, genes replicated early in S may have a competitive advantage in the binding of transcription factors or of structural proteins which maintain chromatin in an open state (Haaf, 1995). However, suggestions that early replication may cause gene activation are speculative at this point; it is just as likely to be the other way around, i.e. transcription creates an open chromatin state which permits early origin activation (Diller and Raghuraman, 1994).

On a local level, origins have been observed to fire in synchrony, with clusters of contiguous replicons activating almost simultaneously to replicate an entire chromosomal region (Diffley and Stillman, 1990). Consecutive origins on a chromosome have been hypothesized to be permanently attached in linear order to a structural backbone, the nuclear matrix (Wanka, 1991). This organization may serve to define replicon clusters and therefore chromosomal subdomains. The observations by Brewer et al. (1993) on the position of replicon termination sites support this model; where two early-replicating origins are chromosomal neighbours, they activate in synchrony as part of an origin cluster, and constant fork progression means that fork-collision-mediated termination will occur reproducibly at a very similar genetic position, whereas a late-replicating origin next to an early replicating one will be part of a different replication domain, activating at a different time. In this case the position of fork termination will end up closer to the late origin, and in a more variable position dependent on the eventual time of the late cluster's activation.

In conclusion, the arrangement of replicons along the length of human chromosomes defines perhaps 50 - 100 subdomains on each chromosome, generally replicated by clusters of 10 - 20 replicons each, 50 - 300 kb in size.

Imprinting

If replication origin biology is intimately associated with both the process of transcription and with higher order chromosomal domains, perhaps it is not all that surprising that recent work has suggested that origin function is altered in regions of genomic imprinting. Replication of a genomic locus can be visualized by fluorescent *in situ* hybridization (FISH) and mapped specifically to the paternal or maternal allele when a second probe for a heterozygous locus on the same chromosome is available. The timing within S phase of locus replication can be determined by synchronization-release, flow sorting for DNA content, and/or the simultaneous use of other probes for which timing patterns have already been established. Using these techniques, Kitsberg et al. (1993a) demonstrated that while typical alleles throughout the genome replicate synchronously in S phase (both allelic copies replicate at the same point within S, be it early or late in S), this rule breaks down for imprinted alleles. In the mouse genome, all four known imprinted genes are replicated earlier on the paternal copy than on the maternal. The same result has been shown in humans, where the 15q11-q13 imprinted genomic region displays allele-specific timing asynchrony at several loci (Knoll et al., 1994), where the paternal copy is replicated in mid-S, but the maternal only in late S. The first example of a maternal early/paternal late pattern of replication timing asynchrony was also identified in this region, in a GABA receptor gene cluster (LaSalle and Lalande, 1995).

The hallmark of imprinted genes, of course, is allele-specific transcription. Some kind of signal is imprinted onto an allele during gametogenesis, which is maintained following embryogenesis, such that the adult progeny express only the maternal or the paternal copy of that gene, although both are present. As an epigenetic phenomenon controlling the differential expression of allelic copies, imprinting is related to, for example, X chromosome inactivation (Lyon, 1993). A handful of imprinted loci have been identified to date: the *Snrpn/Znf127*, *H19/Igf2*, and *Igf2r* gene regions in mice, and the 15q11-q13, 11p15·5, and sometimes the 6q25-27 regions in the human (Nicholls, 1994). Imprinting has a role in several genetic diseases, including Prader-Willi Syndrome (PWS), Angelman syndrome (AS), Beckwith-Wiedemann syndrome and Wilms tumor, and possibly in chronic myelogenous leukemia, fragile-X syndrome (Nicholls, 1994), and some neurobehavioural disorders (Durcan and Goldman, 1993).

The human 15q11-q13 region is particularly interesting because it is associated with two distinct genetic syndromes, depending upon which allelic contribution is missing. PWS, characterized by hypotonia, failure to thrive, hyperphagia, mild retardation, and hypogonadism, occurs in one birth in 25,000 (Nicholls, 1993). Patients either carry a

deletion in the paternal 15q11-q13 region, or lack a paternal chromosome 15 entirely, i.e. maternal uniparental disomy (Knoll et al., 1993). Angelman syndrome is rarer, and is characterized by severe retardation, seizures, and malformations of the head; patients lack a maternal contribution from 15q11-q13, usually because of a deletion. Families with small deletions have permitted finer mapping studies, and it is now clear that "the imprinted AS and PWS loci are separate" (Nicholls, 1994). The critical PWS region has been localized to 400 kb (Lalande, 1994), including three transcripts which display paternal-specific expression: the small ribonucleoprotein-associated polypeptide N (*SNRPN*), and two other transcripts which lack an open reading frame, reminiscent of XIST, which controls X-inactivation (Brown et al., 1992). No maternal-specific transcripts, candidates for AS, have been identified in the region, but an identified deletion under 200 kb in size (Buxton et al., 1994) suggests that the AS critical region is distal to PWS, and proximal to the GABA receptor cluster (Reis et al., 1993; Sinnett et al., 1993).

However, in the case of imprinted genes, the critical regions carrying differentially-expressed coding sequence for the gene(s) responsible for the disease phenotype may not tell the whole story. A *cis*-acting, heritable imprinting control element may not be part of the critical gene *per se*, but may lie elsewhere in the same genomic region or chromosomal domain. As yet, no model for how imprinting is transferred, maintained, or imposed by *cis*-acting heritable elements has been proven, but perhaps the most popular mechanism invoked is that of differential methylation (Razin and Cedar, 1994). Methylation of upstream CpG islands may inhibit transcription factor or enhancer binding and prevent expression of genes dependent upon hypermethylated activating sequences. In support of this model, islands of differential methylation have been identified near all recognized imprinted loci, and recently it has been shown that methylation patterns spread over hundreds of kilobases near the PWS critical region are dependent on the methylation pattern in the *SNRPN* α exon (Sutcliffe et al., 1994). However, there are problems with CpG methylation as the definitive mechanism for imprinting; for example, sometimes it is hypomethylation which correlates with gene activation (which could be explained if transcriptional repressors have methylation-sensitive binding). More seriously, it is now apparent that gametic methylation patterns are not maintained in embryogenesis. An extensive or possibly complete demethylation takes place in the pre-implantation morula, followed by *de novo* methylation at implantation which establishes the adult pattern (Razin and Cedar, 1994). It is not yet known how the establishment of the adult methylation pattern is controlled by the paternal source of a chromosome, nor is there any direct proof that methylation itself causes the altered transcription at an imprinted locus.

Thus, it remains possible that epigenetic inheritance may be controlled by factors other than differential methylation. In *S. pombé*, an imprinting-like phenomenon controlling mating type interconversion is caused by inheritance of specific double strand breaks and a "cleavable" chromatin state, which is established at the time of chromosome replication (Klar and Bonaduce, 1993). In higher eukaryotes, the state of chromatin condensation has been put forward as a trait inherited in *cis* which could, hypothetically, confer the imprinted phenotype. The observation of replication timing asynchrony in imprinted regions suggests that differential function of replication origins may have a role in the establishment of imprinted expression patterns (Lalande, 1994), and some authors have suggested that it is replication timing which directs imprinted transcription and not the other way around (Coverley and Laskey, 1994). However, the relationship may again be complex, because regions with a single pattern of paternal early replication timing asynchrony sometimes contain not only a paternal-specific imprinted gene, but also non-imprinted genes, or even (in the case of murine *H19* and *Igf2r*) a maternal-specific imprinted gene (Nicholls, 1994). In fact, a combination of models may be involved, since DNA methyltransferase can localize to replication factories (Nicholls, 1994), and methylation directed from a single control element may affect large chromatin domains (Sutcliffe et al., 1994).

The nature of the observed relationship between allele-specific replication timing asynchrony and allele-specific (imprinted) transcription remains to be determined. Regardless of which is the cause and which the effect, investigations of these two related genetic phenomena will shed light on the nature of chromosomal replication origins, and perhaps even their role in genetic diseases. Identification of a replication origin in an imprinted, asynchronously-replicating genomic region would be a valuble first step.

Replication, Amplification, and Genetic Instability

While DNA replication is closely related to transcription, it is also associated with the two other fundamental processes of DNA biology, repair and recombination.

Even in the absence of mutagenic insults, DNA has limited stability *in vivo* (Lindahl, 1993). Each day, acid-catalyzed hydrolytic damage to the human genome creates 6000 apurinic sites, and deaminates 300 cytosine or 5-methylcytosine residues, which can lead to transition mutations. Oxidative damage to about 300 guanine residues can cause transversion mutations, and non-enzymatic methylation of 600 adenine residues creates potential blocks to replication. During replication, mismatches caused by polymerase base pairing errors occur at a frequency of 1 base in 10^6 with polymerase α, but only 1 in 10^8

with polymerase δ, which has an associated 3' - 5' exonuclease "proofreading" activity (Kornberg and Baker, 1992). Polymerase incorporation errors can also occur when replicating short polynucleotide repeats (perhaps by Okazaki fragment slipping), or after hitting a replication block (Cohen et al., 1994). Many types of damage can accumulate, including missing, modified, dimerized or mismatched bases, deletions, insertions, strand breaks, and crosslinks.

The cell protects its genome from damage through its structural design (oxidative metabolism is sequestered in the mitochondria, histones quench damaging chemical agents) and by employing DNA repair pathways (Lindahl, 1993). Specific enzyme systems repair the common types of damage; for example, AP endonucleases clip out hydrolyzed apurinic nucleotides, and the small DNA polymerase β, present at constant levels throughout the cell cycle, fills the gap, which is subsequently sealed by DNA ligase I. The excision repair pathway provides a backup system flexible enough to fix unusual mutagenic lesions. Deficiency in an enzyme responsible for postreplicative mismatch repair makes HNPCC patients susceptible to early-onset colon cancer (Cleaver, 1994), but most transformed cell lines are not deficient in mismatch repair, although they do show increased genetic instability, manifested by a propensity to accumulate genomic rearrangements (Boyer et al., 1993). Transcribed sequences are preferentially repaired, with lesions fixed much more quickly than in other genomic sites (Troelstra et al., 1992). While some repair genes apparently double as transcription factors (Schaeffer et al., 1993), a likely explanation for the observation of preferential repair is that the changes in superhelical density and chromatin condensation associated with transcription and early replication may facilitate DNA repair processes.

Genetic recombination may be similarly related to replicative phenomena. The four categories of recombinative events are homologous (between two identical sequences), site specific (requiring vital sequence signals at all cleavage sites), transposition (only one of the two sequences involved needs a specific motif), and illegitimate (all other events). In general, recombination is also more commonly associated with transcriptionally-active, early replicating sequences. For example, V(D)J immunoglobulin site-specific recombination, which uses its own particular recombinase enzymes, requires a transcriptionally-active substrate that may reflect an open chromatin, early-replicating state (Schatz et al., 1992). The intronic enhancer origin has been proposed to be the key to achieving this recombinative state at the IgH locus (Ariizumi et al., 1993), and could play a similar role in subsequent immunoglobulin class switch recombination (Harriman et al., 1993). In replicating mammalian cells, illegitimate recombination is the major process, and occurs particularly at bent DNA sequences which may be prone to the necessary strand

breaks. An example of illegitimate recombination is the surprisingly common insertion of mitochondrial DNA fragments into nuclear genomic DNA (Thorsness and Fox, 1990), which may have a role in carcinogenesis and aging (Shay and Werbin, 1992). The mechanism behind illegimate recombination *in vivo*, including whether replication origins contribute in any way, is largely unknown. It has been observed that in some cases cryptic replication origins can be activated by flanking sequence rearrangements, possibly by creating inverted repeat elements where there were none before (Leu and Hamlin, 1992). Finally, in some bacteriophages, replication and recombination are tightly interdependent (Kornberg and Baker, 1992).

Amplification represents a case of a genetic instability which could involve deranged origin function. The *DHFR* gene, c-*myc*, and the epidermal growth factor receptor (Dolf et al., 1991) are examples of genes amplified in cell lines to confer a growth advantage or drug resistance. The repeated region of DNA is termed the amplicon, and contains at least one replication origin; DePamphilis (1988) has suggested that it is those origins most tightly associated with auxilliary transcription factors and active genes which are most likely to amplify.

Amplified loci are found *in vivo* in two forms, as double minute chromosomes or as homogeneous staining regions. Double minutes are circular, acentromeric large episomes which can exist in variable copy number due to unequal segregations at mitosis. The amplified mouse adenosine deaminase double minute has been shown to employ the same 5' origin used for replication of the native chromosomal locus, under the same cell cycle control (Carroll et al., 1993). The mechanisms required for formation of extrachromosomal circular DNA remain enigmatic (Gaubatz and Flores, 1990), but the presence of a single, controlled origin confers the capacity for apparent amplification via unequal segregations under selective pressure. Homogeneous staining regions are large, ectopic chromosomal domains containing nothing but repeated amplicon copies. Ma et al. (1993) have presented evidence that unequal somatic recombination with sister chromatid exchanges is the major process at work in the formation of homogeneous staining regions, and is caused by defects in strand break repair rather than any origin hyperactivation.

Thus, while origins are present and active in amplicons, the key event in most amplifications seems to be a recombination rather than an origin hyperactivation, although examples of origin overactivation do exist in certain cases of programmed developmental amplifications, such as at the *Drosophila* chorion and *Tetrahymena* rDNA loci (Yu and Blackburn, 1990).

Mutations and large rearrangements occur most commonly in decondensed, early replicating chromosomal regions (Haaf, 1995). In this respect, origin activity may be

related to fragile sites which are more commonly involved in translocations and rearrangements, such as the sister chromatid exchanges which can lead to amplification.

Cell Cycle Control

Research on human cell cycle control has expanded rapidly into a large, multifaceted field of study. However, perhaps the most important single event to take place in a cell cycle is the initiation of DNA replication. How is origin firing regulated by cell cycle control factors?

Eukaryotic cells are in one of two states: G0 or active cycle. In the former state, cells contain about 1/3 less protein and RNA and require a number of physiological signals before they can return to an actively cycling state. A complete cycle in a higher eukaryotic cell requires some 16 - 30 hours. G1 phase can last any length of time (there is no G1 in cleavage embryos, while some cells can move from G1 to G0 and rest for an indeterminate period of time), although 12 hours is most typical; a key event near the end of G1 is the passage of a START point, where the cell commits to entering and completing the S, G2, and M phases. All genome replication takes place in S phase, which can last 6 to 12 hours -- the overall time spent in S is relatively constant for a given cell type (Hamlin et al., 1994). G2 phase, where sister chromatids remain paired and uncondensed, requires about 4 hours, during which a checkpoint must be passed to ensure that the cell is ready for the M phase of active mitosis, which requires 1 - 2 hours to complete. Levels of the cellular proteins known as cyclins are dependent on cell cycle position: cyclin A is present during S, the cyclin B family during M, and cyclins C, D, E, and F dominate in G1 (Dowdy et al., 1993). Cyclins drive posttranslational modification of select enzymes by activating cell cycle dependent kinases (cdk), and the control over origin function by cyclins and cdk's is suggested by the observation that protein kinase inhibitors can specifically inhibit replication initiation in human cells (Gekeler et al., 1993).

In *S. cerevisiae*, present evidence suggests that at least one component of the initiator protein, ORC6, has serine/threonine phosphorylation sites and is activated by yeast cdk factors like CDC6 and CDC46/MCM3 (Li and Herskowitz, 1993). Yeasts have been popular models for investigation of cell cycle control factors, but may not model all that well those controls required in multicellular eukaryotes (Heichman and Roberts, 1994). For example, yeast cell cycles are ten to twenty times shorter than in human cells, and the breakdown of the nuclear envelope in M phase is not observed (Coverley and Laskey, 1994).

In the 1970s, the cell cycle-imposed limitations on DNA replication were first investigated with a series of now classic cell fusion experiments (Diffley and Stillman, 1990; Heichman and Roberts, 1994). When a G1 cell is fused with an S phase cell, the G1 nucleus is pitched into an early S phase, suggesting the presence of *trans*-acting activators of DNA replication initiation in S cells -- but not in G2, M, or G1 cells. When a G2 cell is fused with an S cell, the G2 cell does not initiate DNA replication, implying that it has a *cis*-acting factor preventing replication, such as a change in nuclear organization or the lack of a bound "licensing factor" required for replication, but not a *trans*-acting inhibitor of replication, since the S nucleus continues its normal replication.

Only very recently has the nature of some of these factors begun to be clarified, and the picture is complex. Cdk2, the vertebrate homolog of yeast CDC28, is the cell cycle factor most closely associated with the initiation of DNA replication in higher eukaryotes (Diffley and Stillman, 1990), and is a required component for entry into S phase and for replication *in vitro* by *Xenopus* egg extracts (Yan and Newport, 1995). The 32 kd cdk2 protein, which complexes with cyclins A and E, is similar to but distinct from the p34^{cdc2} kinase, which instead binds cyclin B to form MPF (mitosis- or maturation-promoting factor), the key complex promoting entry into M phase from G2 (Coverley and Laskey, 1994). During G1, there is almost no cellular cyclin A, and cdk2 is located in the cytoplasm; however, on entry into S, high levels of cyclin A and cdk2 are found in the nucleus (Brenot-Bosc et al., 1995). This association of cdk2 with the S phase cyclin A is particularly intriguing: cdk activity is needed for the assembly of nuclear replication factories, and cyclin A colocalizes with active replicon clusters (Heichman and Roberts, 1994). The mechanism by which a cyclin A-cdk2 complex could promote initiation remains a matter of controversy; some groups suggest that cdk2 phosphorylation of a subunit of the human RP-A SSB activates this necessary replication protein (Brenot-Bosc et al., 1995), while others insist that cdk2 acts neither on the origin recognition step nor through RP-A, but instead by activating uncharacterized proteins needed for stable origin unwinding and the later switching step converting RNA priming to DNA polymerase extension (Yan and Newport, 1995). Cdk2 has also been hypothesized to act by activating helicase activity after association with cyclin E (Heichman and Roberts, 1994).

An alternative possible role for cdk2, or perhaps a complementary mechanism by which it indirectly induces replication, comes through its kinase action on a key tumor suppressor substrate, the retinoblastoma protein (Rb). Cdk2, when complexed with either cyclin A or E, hyperphosphorylates the Rb protein, inactivating its multifaceted growth suppressing functions (Dowdy et al., 1993). One result is that hyperphosphorylated Rb can no longer bind and sequester cyclin D, which is then free to complex with cdk4 to

activate the kinase and act as a START signal, although cdk4 can still be kept in check by the action of the recently discovered tumor suppressor p16 (Parry et al., 1995). Rb also acts by downregulating the E2F transcription factor (White, 1994) and by binding and sequestering the c-*abl* nuclear tyrosine kinase (Welch and Wang, 1993). Thus, cdk2-mediated deactivation of Rb frees up multiple pathways which promote progression from late G1 to the first initiation in S.

A strong candidate for a negative regulator of replication has emerged in $p21^{Cip1}$, which directly deactivates cdk2, and may thus knock out any direct activation of initiation by cyclin-cdk2 or indirect activation of S phase by cyclin-cdk2 phosphorylation of Rb (White, 1994). A second, independent function of $p21^{Cip1}$ is in blocking the activity of PCNA, the processive factor for the polymerase δ holoenzyme complex. In this manner, $p21^{Cip1}$ may act to regulate the switch between priming and elongation (Yan and Newport, 1995), or to check both initiation and elongation when S phase must be arrested because of DNA damage (Heichman and Roberts, 1994).

The tumor suppressor p53 is thought to serve as a major checkpoint in late G1, ensuring that S phase does not start (no initiation takes place) until the cell is ready and any outstanding DNA damage has been repaired (Hamlin et al., 1994). The observation that loss of p53 correlates with frequent formation of tetraploid and other aneuploid tumor cell subclones suggests that one specific manifestation of this function of p53 is as a checkpoint against overreplication, by ensuring that S phase cannot ensue if M has not been properly completed (Cross et al., 1995). Evidence now suggests that p53 blocks replication by acting as a transcription factor for $p21^{Cip1}$, which then blocks the cyclin/cdk complexes needed for induction of S phase (White, 1994). Should inappropriate replication nevertheless ensue, p53 activates the apoptosis pathway.

The licensing factor model for cell cycle control of origin function postulates the existence of a protein which is necessary for activation of the initiator (or deactivation of an initiator inhibitor) and which, by virtue of its presence in G1 but not G2, limits origin activation to the appropriate phase of the cell cycle. Such a factor would lack a nuclear localization signal and could only access the nucleus after the mitotic nuclear membrane disintegration, and then remain to "license" the DNA for replication before being consumed during S phase (Huberman, 1995). If the licensing factor becomes part of a pre-formed initiation complex, it might be destroyed during origin activation, limiting firing of that origin to once per cell cycle. In *S. cerevisiae*, the footprint of a suspected licensing factor is present at origins near the constitutive ORC and auxiliary transcription factor footprints, but only between early telophase and the end of G1, suggesting that a licensing factor does in fact contribute to a prereplication complex at origins (Diffley et al., 1994). The identity of

the yeast licensing factor is a subject of active research, with the MCM2, MCM3, and MCM5/CDC46 minichromosome maintenance proteins (mutants lose ARS plasmids rapidly) as postulated components, because they only enter the (non-disintegrating) yeast nucleus during M. *Xenopus* and human homologs of MCM3 have recently been discovered (Kubota et al., 1995) and seem to act as part of a larger multiprotein complex (Chong et al., 1995), which includes homologs of MCM2 and MCM5 (Madine et al., 1995). These proteins are only bound to chromatin during G1, dissociating during S phase and only reassociating with chromatin in late M. "Licensing factor" seems to be a large protein complex, and its subcomponents may play separate roles in conferring its properties. Another interesting observation is that a murine MCM3 homolog associates specifically with heterochromatic regions (Starborg et al., 1995). This raises the exciting possibility that the use of different licensing factors, like the yeast MCM series, may help determine the timing of origin activation within S (Diller and Raghuraman, 1994). Another group has speculated that mammalian "licensing factor" could include the unphosphorylated 34 kD subunit of RP-A. RPA34 binds newly replicated DNA in late G2, serves as a site for assembly of other RP-A components (thereby recruiting an SSB to the site of potential unwinding), and is phosphorylated by cyclin A-cdk2 to deactivate its DNA binding property following initiation (Cardoso et al., 1993). This model is attractive because it shows how cdk-mediated phosphorylation can activate a licensing factor at the beginning of S, but unfortunately this model also begs the question of how the G2 dephosphorylation of RPA34 is achieved, and leaves no role for nuclear membrane dissolution in the licensing process.

The cell cycle control proteins regulating entry into M phase from G2 bear some similarities to those involved in the G1/S transition, but have been better characterized. The MPF complex, composed of cyclin B and $p34^{cdc2}$, is structurally similar to the cyclin A-cdk2 complex, but is not found at replication foci (Cardoso et al., 1993). MPF directly regulates the three key steps needed for entry into M: chromosome condensation, by phosphorylation of histone H1, nuclear envelope breakdown, by phosphorylation of nuclear lamin A, and spindle formation, by phosphorylation of RMSA-1 (Yeo et al., 1994). Checkpoints limiting MPF function include the RCC-1 - $p25^{ran}$ system, which associates with growing forks and reports on the completion of replication (Coverley and Laskey, 1994), and the *wee*1 - cdc25 system, which downregulates MPF if the DNA is damaged at the end of S (Heald et al., 1993).

The elucidation of the complete pathways by which cell cycle control is imposed on origin activation has proven difficult, often because the specific substrates for cdk phosphorylation events are not easy to identify. As a result, it has not yet been possible to

discover novel direct origin transactivating factors by working forward from known cyclins and cdk's (Heichman and Roberts, 1994). However, exciting links have been made between suspected G1 and S phase progression factors and known oncogenes and tumor suppressors; indeed, cyclins A and D are now recognized to be protooncogenes (Dowdy et al., 1993). The most important gaps in our knowledge concerning the human proteins involved in initiation are likely to turn out to be cdk substrates, including a yet to be characterized human initiator, licensing factor, and DNA helicase.

Centromeres

Replication origins represent one of the three vital *cis*-acting structures required for the stable maintenance of a chromosome, the other two being centromeres and telomeres.

The centromere, the genetic locus directing mitotic and meiotic partition, lies at a metaphase chromosome's primary constriction; its function is position-independent on natural and artificial chromosomes.

In the case of the budding yeast, *S. cerevisiae*, the centromere has been cloned as a 220 bp CEN element (Mann and Davis, 1986), containing a short 125 bp core, composed of three defined consensus CDE subregions. Specific binding proteins directing chromatid and spindle interactions have been discovered through studies of temperature-sensitive mutants (Earnshaw and Tomkiel, 1992). In at least one case (CBF3, binding CDEIII), a centromere-binding protein has been shown to represent a phosphorylation-dependent complex under cell cycle control. The function of these very short, defined *S. cerevisiae* centromeres can be knocked out by transcription directed through the CEN locus (Smith et al., 1990).

Unfortunately, the relevance of budding yeast as a model for human centromeric function is questionable, as it is structurally very much simpler: while 2% of a human chromosome is taken up by the centromere, the value for budding yeast is only 0.02% (Bloom, 1993). This is the CEN value paradox; animals, plants, and some specific lower eukaryotes such as fission yeast (*S. pombé*), which have larger chromosomes than *S. cerevisiae* (5 - 6000 Mb DNA vs. 1 -5 Mb) and attach more than 1 microtubule per chromosome at mitosis (humans use 20 - 30), nevertheless seem to require up to two orders of magnitude more centromeric DNA *per microtubule*. The *cen* elements of the three *S. pombé* chromosomes are 60 - 100 kb in size, and contain B, K, and L repetitive sequences, with K the only one indispensible for function (Hahnenberger et al., 1989). Fission yeast may be the best available model, and is the model of choice for active research into the protein-binding and higher order folding properties of large centromeres (Steiner and

Clarke, 1994). The equivalent centromere locus in humans is estimated to span 0.5 - 10 Mb, mostly in the form of short α-sattelite repetitive elements organized into higher order repeats specific to each chromosome (Earnshaw and Tomkiel, 1992). The Y chromosome centromere may be the best defined of these, and includes, minimally, a 200 kb α-satellite array and 300 kb associated moderately repetitive DNA (Tyler-Smith et al., 1993). α-satellite DNA is indispensible for centromere function, although the actual repeated DNA sequence as well as the number and structure of repeats shows interchromosomal variation (Lin et al., 1993; Tyler-Smith et al., 1993). The primate α-satellite monomer is 171 bp in length and AT-rich; the only binding site identified to date is for the CENP-B protein, which may help define centromeric chromatin structure (Bloom, 1993). Other repetitive satellite centromere sequences are present on some chromosomes, with unknown function. No direct microtubule-binding activity has yet been demonstrated for any of the DNA elements of the centromere, although elaborate fibre folding structures have been proposed. Histones are required to maintain proper structure of the centromere, which is not distinguishable from other heterochromatin except during M phase, when the kinetochore becomes apparent and the degree of condensation of centomeric DNA is less than that of the rest of the metaphase chromosome (Earnshaw and Tomkiel, 1992). Multiple domains are formed within the metaphase centromere: a pairing domain involved in sister chromatid interactions, a central structural domain, and a kinetochore domain concerned with microtubule attachment (Rattner, 1991).

The kinetochore, an ultrastructural specialization thought to function in the tethering of microtubules to chromosomes, shows a trilaminar constitution under the electon microscope, consisting of electron dense outer and inner plates surrounding a middle space, with the whole structure nestled on the surface of the centromere (Rattner, 1991). An additional possible role for the kinetochore is in monitoring nondisjunction, delaying anaphase until all chromosomes have achieved proper spindle interaction (Bloom, 1993). Given the poor conservation of centromere architecture between budding yeast and the human, as well as the likely more complicated role required of the human kinetochore in the capture of multiple microtubules, there may be little similarity between yeast CBP proteins and their human counterparts. Nevertheless, several human centromere proteins have been isolated by various methods, and they fall into two major categories. INCENP (inner centromere proteins) localize to the sister chromatid interface and represent a novel group of cytoskeletal proteins which are "chromosomal passengers," detaching at metaphase to direct the cytokinetic cleavage (Earnshaw and Bernat, 1991). The CENP (centromere binding proteins) are associated with the inner plate of the kinetochore itself (no outer plate proteins have been isolated to date), and direct the interaction between kinetochore and centromere

heterochromatin (Earnshaw and Tomkiel, 1992). Included among this family are proteins directing higher order DNA structure (CENP-A and CENP-B), microtubule-based motors (CENP-E), and cell cycle control factors (RCC-1 is also known as CENP-D).

A large centromeric region spanning hundreds of kilobases must contain its own internal replication origins. In this regard, Mah et al. (1992) have shown that the *ors*12 nascent strand enriched autonomously replicating putative monkey origin contains both an α-satellite portion and a low copy sequence specific to monkey centromeres.

The nature of the human centromere as a vast, as yet undelineated chromosomal region has presented problems for those who wish to clone a human CEN element. Because a functional chromosome must always retain a centromere, one approach has been to create deleted chromosomal derivatives (by radiation, translocation, or directed insertion of telomere-containing DNA) which are by their nature enriched for centromere sequences that must remain (Miller et al., 1992; Tyler-Smith et al., 1993). A 14 kb sequence derived from sonicated human chromosomes, which reacts with autoimmune sera and contributes to dicentric chromosome formation on transfection into mice, generated some initial excitement, but has since been shown not to be a functional centromere (Hadlaczky et al., 1991).

Since all evidence suggests that human centromeres are huge in size, the appropriate cloning vector is an incomplete artificial chromosome containing origin and telomere components, which need a cloned centromere insert to gain complete chromosomal function. In the absence of a functioning centromere, an origin-containing plasmid does not segregate evenly, and progeny cells show either amplification of the plasmid to high copy number, or complete loss (Smith et al., 1990); however, a functioning centromere stabilizes the construct at a copy number of 1. Such a cloning strategy worked in *S. cerevisiae* (Murray and Szostak, 1983) and *S. pombé* (Hahnenberger et al., 1989) leading to the construction of the first complete artificial chromosomes in these systems.

Telomeres

Telomeres represent the third vital *cis*-acting chromosomal structure, serving to stabilize replication of linear chromosome ends against the progressive loss of the 3' overhanging strand terminus, which cannot be primed for replication by conventional means (Kornberg and Baker, 1992). In addition, telomeres are postulated to protect against exonuclease action, to block unwanted chromosome fusion events, and to act as a nuclear matrix attachment region (Runge and Zakian, 1993) for nuclear subcompartmentalization, perhaps directing chromosome arrangement in meiotic prophase. In many eukaryotes

(*Drosophila* is an exception), telomere structure is similar; a simple repeat sequence, with multiple nicks on one C-rich 5'-ended strand and a singly nicked G-rich 3'-ended strand (Szostak and Blackburn, 1982). The major repeat is T_2G_4 in *Tetrahymena*, TG_{1-3} in *S. cerevisiae*, and T_2AG_3 in mammals (Tyler-Smith, 1993). Long stretches of simple repetitive sequences, such as those present in telomeres, can adopt unusual secondary structures which may play a role in telomere function and recognition by specific proteins (Kang et al., 1995). Proximal to the telomere repeats are transcriptionally-silent 4 - 200 kb proterminal repetitive regions, which may play a role in synapsis or metaphase (Brown et al., 1990). Telomere function is partially conserved across species boundaries, such that *Tetrahymena* telomeres are recognized by yeast, and mammalian telomeres could be initially cloned in yeast by complementing the function of YACs with one missing telomere (Schlessinger, 1990). 0.5 kb human telomere sequences isolated by YAC assay have since been demonstrated to function as telomeres in humans, permitting stable truncation of chromosome ends (Farr et al., 1991). However, native telomeric structures *in vivo* are somewhat more complicated: typical chromosome ends carry a 7 - 10 kb array of T_2AG_3 repeats, but the length is variable and can be several kilobases longer, for example in sperm, or shorter, as in some tumor cell lines (de Lange et al., 1990).

Telomere length depends on the action of telomerase, a ribonucleoprotein reverse transcriptase which extends the T_2AG_3 repeat of the G-rich strand (Blackburn, 1992). There is no detectable telomerase activity in most somatic cells, and only very low telomerase in highly proliferating normal tissues like bone marrow; in contrast, telomerase activity can be very high in malignantly transformed tissues (Counter et al., 1995). Thus, a key event in immortalization and tumorigenesis may be the reactivation of telomerase, and telomere integrity may be a limiting factor driving cellular senescence.

While human and mouse T_2AG_3 telomeres are functionally indistinguishable (Barnett et al., 1993), *Tetrahymena* T_2G_4 telomeres seem not to work in mammalian cells (Shervington, 1993), which only tolerate T_2AG_3 telomeres. However, T_2AG_3 arrays as small as 500 - 800 bp in length can function in human cells whether nicked or intact, and even if interspersed with T_2AG_5 sequences or initially capped by up to 1 kb of non-telomeric DNA; the key to telomere function seems to be recognition by a specific telomeric repeat binding factor, TRF (Hanish et al., 1994). Budding yeast (but not fission yeast) may be unusually permissive for foreign telomeres because they are able to perform a gene conversion, perhaps mediated by a common telomere secondary structure, with endogenous chromosome ends to cap foreign telomeres with the yeast TG_{1-3} repeat (Wang and Zakian, 1990). When an artificial chromosome containing a functional origin and centromere, but

non-functional telomeres, is transfected, for example, into *S. pombé*, the linear construct is often found circularized in those cells which maintain it (Hahnenberger et al., 1989).

Experience from the construction of YACs suggests that while small circular constructs containing origin and centromere components are about as stable as linear constructs with an origin, centromere, and telomeres, the linear versions become the much more stable form when overall size exceeds 100 kb. Thus, human artificial chromosomes designed to act as cloning vectors for the study of large genomic regions are expected to require functional telomeres. It seems highly likely that the *Tetrahymena* telomeres used in YACs will have to be replaced with cloned mammalian telomeres if one hopes to generate a linear acentromeric human artificial chromosome, which could then be used to clone a large centromeric element.

MAPPING REPLICATION ORIGINS

Over a dozen commonly-used techniques are currently being employed by laboratories around the world for the localization of active origins of DNA replication. Each methodology has its own set of inherent assumptions, advantages and disadvantages, and capacity for mapping resolution which must be taken into account when interpreting the results. In fact, when multiple techniques are applied to the same replication origin, the apparent answers are not always in full agreement (Burhans et al., 1990).

Mapping techniques fall into two major categories: physical methods which localize the initiation site (functional origin) where the first new bases of the nascent daughter strands are laid down, and replication assays which determine the position of the *cis*-acting origin control element (genetic origin) directing initiation in its associated genomic region (DePamphilis, 1993a).

The various mapping methodologies currently available have been the subject of extensive and thorough reviews (Vassilev and DePamphilis, 1992; Falaschi et al., 1993; Hamlin et al., 1994). Consequently, this section of the dissertation will cover most methods only in brief, and emphasize those methods actually employed in the experiments to follow.

Physical Methods for Locating the Initiation Site

A simple, direct and obvious way to identify an initiation site of DNA replication is to add radioactive precursor and visualize the restriction fragments which incorporate the

label immediately following initiation. In practice, this is only applicable to *in vivo* origins activated at the very onset of S phase in synchronized cells, and specifically to high copy number origins (such as those present in amplified DNA regions) which yield obvious specific fragments above the background from low copy origins. Imperfect synchrony limits resolution to ten kilobases or so. This "earliest labeled fragment" method has been used to map the origins present in the CHOC400 hamster dihydrofolate reductase amplicon (Hamlin et al., 1994) and the B-1/50 mouse double minute adenosine deaminase amplicon (Carroll et al., 1993). While opportunities for determining an earliest labeled fragment *in vivo* are limited, cloned origins replicating autonomously *in vitro* are inherently synchronized, intensely radiolabeled, and present at a high copy number; consequently, determination of the restriction fragment predominantly labeled at early time points represents, in this system, a particularly convenient method for mapping an initiation site (Pearson et al., 1991).

A similar, direct method for mapping initiation sites uses the labeled, purified short nascent DNA strands as probes to screen a restriction fragment panel. While this strategy has also been applied to the *DHFR* locus, for example, it suffers the same limitations as the earliest labeled fragment method, and is further complicated when repetitive DNA sequences are involved.

However, a new technique which takes advantage of the incredible sensitivity and sequence specificity of PCR has overcome many of the limitations in nascent strand analysis. As a prerequisite, a suspected origin region must be sequenced and three or more unique primer pairs chosen which amplify distinct genomic sequences spaced across the region in question. Log phase cells are pulse-labeled with bromodeoxyuridine (BrdU) and total genomic DNA is prepared. Nascent strands are purified and size-fractionated by denaturing density gradient ultracentrifugation, and may be further purified from unreplicated DNA with anti-BrdU antibodies. The size range of the nascent fragments present in each fraction can be determined by alkaline agarose gel electrophoresis, and fractions are used as template in separate PCR reactions with each primer pair. The primer pair closest to the replication origin will successfully amplify the fraction containing the shortest nascent strands, and the technique thus determines the distance each primer pair lies away from their nearest replication origin. With three primer pairs separated by only a few kilobases, nascent strand PCR can map an initiation site to within a kilobase or so. Importantly, this method can be applied to any unique locus in unsynchronized, replicating cells. However, primers must be specifically and intentionally chosen in regions where origin activity has been suggested by other data, and the instability of long BrdU-containing sequences means that primers must be relatively close to an origin for the technique to work

at all. Nevertheless, since this technique was first applied to the mapping of the 5' human c-*myc* origin (Vassilev and Johnson, 1990), it has become a popular method for the high resolution confirmation of origin activity in, for example, the hamster *RPS14* (Tasheva and Roufa, 1994) and *DHFR* 3' regions (Burhans et al., 1990), the murine IgH enhancer (Ariizumi et al., 1993), the promoter of human heat shock protein 70 (Taira et al., 1994), and the human 343 origin from chromosome 6q (Wu et al., 1993b).

Since PCR amplification can sometimes give inconsistent results, because of variable sample quality, contamination, or other factors, at least one group (Falaschi et al., 1993) advocates the addition of specifically-designed competitor substrates to each nascent strand PCR reaction, to allow quantitative comparison between different fractions and primer pairs. For repetitive or viral loci, PCR detection is not required, and origin position can be mapped by direct probing and densitometry to measure the presence and abundance of short nascent strands (Yoon et al., 1995).

Instead of merely mapping the position of nascent DNA strands, it is possible to isolate and clone these short pieces of newly replicated DNA that must include replication origins. The "origin trap" technique involves *in vivo* psoralen or trioxsalen crosslinking of DNA every 1 to 2 kb, addition of BrdU and subsequent isolation and cloning of the short labeled uncrosslinked nascent fragments, which can only form in spaces containing an initiation site that lie between crosslinked positions of genomic DNA (Dimitrova et al., 1993). Alternatively, short nascent strands can be obtained from those origins activated at the beginning of S phase without having to use a potentially damaging crosslinking agent, by "strand extrusion." In this method, synchronized cells are released briefly into S, and their isolated, labeled DNA is heated to 50°C to denature and release the unstable nascent strands in the replication bubble; after cooling, the genomic DNA prefers to close back up with itself and the nascent strands are left to associate with their semiconservative complementary partner (Kaufmann et al., 1985). Using either method, origin-enriched libraries of nascent strand clones are obtained which can be probed to see if they include sequences of interest, or themselves be used as genomic probes to localize the native position of the identified initiation site. Because bulk genomic fragments can contaminate nascent strand isolates, origin activity should be confirmed using another technique.

The identification of the position of short nascent strands also underlies initiation site mapping by neutral-alkaline two dimensional gel electrophoresis (Dubey et al., 1991). Unlabeled total DNA from log phase cells is restriction digested and electrophoresed first in a standard agarose gel according to size, and then in a second dimension under alkaline denaturing conditions which release nascent strands (Hamlin et al., 1994). The gel is then blotted and probed with a series of defined genomic fragments; the probe which hybridizes

to the shortest nascent strands is closest to the end of the restriction fragment into which the replication fork first enters. In this manner, fork direction can be determined, pointing the way to the genomic region containing the initiation site. Neutral-alkaline 2DGE has yielded excellent results when used to map yeast origins, but has proven difficult to apply in higher eukaryotes with large genomes and therefore vanishingly low concentrations of any specific nascent strand sequence, and is further complicated by the presence of any nicked or damaged parental genomic DNA or of repetitive sequences in the probed regions.

A second 2DGE technique uses a high voltage, low temperature second dimension in the presence of ethidium bromide, which makes restriction fragment mobility highly dependent on molecular shape, allowing differentiation of bubble from Y-shaped replication intermediates, a novel method for mapping initiation sites (Brewer and Fangman, 1991). In fact, "neutral-neutral" 2DGE yields a great deal of information about the structures formed during replication of a genomic fragment for which a probe exists, and modifications of the technique also allow the direction of fork movement through a fragment to be determined. This method has become especially popular for the analysis of yeast origins and requires neither synchronization nor labeling, allowing mapping to within a few hundred base pairs resolution (Vassilev and DePamphilis, 1992). However, gel interpretation depends on assumptions about replication intermediate structures formed by undamaged DNA preparations, and about the bidirectionality of fork movement. Application of neutral-neutral 2DGE to higher eukaryotes has been hampered for some of the same reasons as for neutral-alkaline 2DGE, and observed hybridization patterns are more complicated than expected by theory. In mammals, 2DGE has suggested, in direct contrast with many other techniques, that initiation sites may be non-specific across a broad potential initiation zone (Vaughn et al., 1990; Krysan et al., 1993).

The key replication intermediate defining an origin, a "replication bubble," can also be visualized directly via electron microscopy. In cases where a visualizable landmark (for example, a restriction site in repetitive or plasmid DNA) is available, an initiation site can be mapped by direct measurement on an electron micrograph (Pearson et al., 1994b).

Mammalian origins mapped by any of the above methods can be assigned to a specific chromosomal position by using the defined sequence as an *in situ* hybridization probe (Shihab-El-Deen et al., 1993). Alternatively, other mapping techniques are available which examine replicon patterns on a larger, chromosomal scale.

One example is the classical technique of fibre autoradiography, where pulse-labeled DNA is spread over a slide and covered with a photoemulsion. This technique permits identification of replicon clusters, and differentiates bidirectional from unidirectional

origins, but takes months to complete, and does not assign origin position within the genome (Kornberg and Baker, 1992).

The direction of fork movement through a large restriction fragment can also be determined by "run-off replication" (Vassilev and DePamphilis, 1992). Isolated S phase nuclei contain replication bubbles of various sizes, which are extended in the presence of BrdU *in vitro*, under conditions where no new initiations take place. Fragments furthest away from origins will contain the most BrdU. Run-off replication is useful for mapping the predominant direction in which genes are replicated, for example, but permits only very poor resolution of initiation site position.

Large-scale analysis of the organization of replication timing domains has also become possible, with the recent application of fluorescent *in situ* hybridization techniques to S phase cells (Kitsberg et al., 1993a). FISH probes light up as a visible singlet at each locus, and as a doublet following replication of that locus, allowing immediate analysis of allelic synchrony in replication timing, when second, chromosome allele-specific probes allow differentiation of the maternal versus paternal copy. Where still other probes with known timing of replication within S are available, or where an initial FACS step separating S phase cells on the basis of completeness of their DNA synthesis is included, the timing of the replication of the chosen locus can be determined (LaSalle and Lalande, 1995).

Initiation sites in chromosomal replicons can also be mapped by the "imbalanced DNA synthesis" method (Vassilev and DePamphilis, 1992). *In vivo*, the protein synthesis inhibitor emetine blocks lagging, but not leading strand synthesis, and when nascent strands are isolated and analyzed with strand-specific riboprobes, the initiation site must lie in the region where nascent strand polarity shifts. Some problems with this method are the uncertainty of how emetine blocks lagging strand synthesis and what other effects it may have on replication physiology, and low (6 - 15 kb) resolution (Falaschi et al., 1993).

A related mapping strategy is the "Okazaki fragment strand switching" technique, where strand-specific riboprobes are used to screen 100 - 300 bp Okazaki fragments isolated by denaturing gel electrophoresis, which by their nature are derived exclusively from the lagging strand (Burhans et al., 1990). A panel of short, unique riboprobes and their complements can localize the initiation site to within 450 bp, making this the most high resolution technique currently available. However, the low levels of specific Okazaki fragments present at any single point in the cell cycle means that the technique only works when applied to origins activated at the beginning of S in cells synchronized with metabolic inhibitors. Furthermore, hybridization of probes does not turn out to be perfectly strand-specific (Carroll et al., 1993), and this method would yield complicated or misleading

results if, for example, initiation can occur in multiple closely-spaced points across the length of a DUE.

Functional Assays to Identify *cis*-Acting Genetic Control Elements

Genetic assays, designed to identify "replicators," cloneable origin-determining control sequences, have potential advantages over assays which merely localize initiation sites. By subcloning, mutagenesis, and sequencing experiments, the minimal components necessary for origin function can be defined, and origins can be cloned into new contexts for experimental investigation.

The *S. cerevisiae* ARS assay represents the paradigm in this field. While only about half of yeast sequences conferring autonomous replication onto plasmids turned out to be *bona fide* yeast chromosomal origins (Hamlin et al., 1994), ARS assays serve as an extremely sensitive test for origin function. This work eventually led to the definition of the minimal yeast origin consensus sequence, the identification of the ORC complex, and the construction of yeast artificial chromosomes.

Small amounts of extrachromosomal closed circular (episomal) DNA can be found in essentially all types of eukaryotic cells (Renault et al., 1993), although its biological role remains uncertain (Gaubatz and Flores, 1990). In many cases, such sequences are thought to be temporary byproducts of recombination events, but some episomes replicate and are maintained in the cell, including double minute chromosomes as small as a few hundred kb in size (Dolf et al., 1991), or larger ring chromosomes which can in some cases be heritable (Kosztolanyi et al, 1991).

Autonomous replication assays in mammalian cells have been attempted by many groups. Suspected origin sequences are cloned into a prokaryotic vector and transfected into cultured cells, most commonly by calcium phosphate coprecipitation (Chen and Okayama, 1987). Even if the plasmids do contain functional origins and replicate efficiently, they lack a centromere and do not partition accurately at mitosis, meaning that they are inherently unstable and are lost from the population without a selective pressure. Thus, replication is assessed after transient transfection of 48 - 72 h, or after longer term culture under drug selection for a marker present in the vector. Autonomous replication is usually assayed by one of two methods: "semiconservative replication assay" in the presence of BrdU, looking for plasmids containing one or two completely BrdU-substituted "heavy" strands, or alternatively by digestion with methylation-sensitive enzymes (*Dpn*I and *Mbo*I), which recognize, under appropriate digestion conditions (Sanchez et al., 1992), whether the bacterial methylation pattern has been lost through replication in the eukaryotic

system. With the latter method, enzyme-resistant DNA can be assessed by careful analysis of Southern blots (Rao and Martin, 1988), by bacterial back-transformation assay (Vassilev and Johnson, 1988), or by PCR (Cooper et al., 1994).

Whereas in yeast, ARS assay is perfectly sensitive, if not entirely specific (Dubey et al., 1991), autonomous replication assays in mammalian cells seem to be considerably less sensitive (Hamlin et al., 1994). When plasmid libraries of random human shotgun fragments were transfected into human cells, it was not possible to isolate actively replicating origin-containing clones (Biamonti et al., 1985). In addition, two separate groups have been unable to detect autonomous replication activity in fragments containing the known initiation site from the hamster *DHFR* locus (Burhans et al., 1990; Caddle and Calos, 1992). Episomal replication in human cells is complicated by inaccurate partition and possibly by depressed origin activation in the plasmid context (Falaschi et al., 1993), and analysis is made more difficult by non-replicative (repair) synthesis, and by potential integration of transfected plasmid into the host genome (Gilbert et al., 1989).

One group (Krysan et al., 1989) attempted to circumvent the problem of inaccurate partition by designing a crippled EBV-based transfection vector, retaining the family of repeats element of oriP, which is believed to direct (in the presence of EBNA-1) equal partition at mitosis. However, subsequent work showed that any cloned human insert greater than about 12 kb replicated in their system (Heinzel et al., 1991); yeast inserts replicated about as well as human DNA (Tran et al., 1993), and prokaryotic inserts, while considerably less efficient, were also active when arranged into large multimers greater than 30 kb (Krysan et al., 1993). Initiation sites on the plasmids, assessed by neutral-neutral 2DGE, were randomly positioned (Krysan and Calos, 1991). The same group has recently published similar results showing non-specific autonomous replication of human, fly, and bacterial sequences in *Drosophila* cells (Smith and Calos, 1995). In summary, their results suggest that either metazoan origins do not require specific or uncommon sequence motifs, or that autonomous replication assays in metazoans show greatly reduced specificity, and can use almost any sequence as a potential origin.

Although the authors briefly address various problems with their system (choice of cell line, effects of viral EBNA-1, partial function of the crippled vector, and inability to differentiate episomal replication from integrated constructs when linearizing restriction enzymes are used), the strongest challenge to their results may be the simple fact that many other groups have successfully used autonomous replication assays in mammalian cells to detect specific, short DNA sequences that confer putative origin activity, once per S phase. Masukata et al. (1993) demonstrated that, in 293S cells transfected with a purely prokaryotic vector, replication efficiency is not dependent on simple size, but on specific

cloneable subfragments, and BrdU semiconservative replication assay suggested that autonomously replicating plasmids were thirty times more active than vector alone. An origin 3' to the murine adenosine deaminase locus replicates autonomously when cloned as a 4 kb insert in a defective EBV-based vector; initiation is specifically localized to the insert by nascent strand PCR (Virta-Pearlman et al., 1993). A different origin, on the 5' side of the same mouse locus, is active on B-1/50 cell episomes; Okazaki fragment strand switching and earliest labeled fragment assays show that the same specific initiation site is used in the episomal and the chromosomal context of the adenosine deaminase gene (Carroll et al., 1993). Okazaki fragment strand switching also shows that the same origin is used by the hamster *CAD* gene when replicating episomally as when replicating in its *in vivo* chromosomal context (Kelly et al., 1995). Fragments from the upstream region of the human c-*myc* gene, where an initiation site has been localized by multiple techniques, direct autonomous replication in short and long term BrdU and *Dpn*I/*Mbo*I assay (McWhinney and Leffak, 1990). Another group has subcloned a 210 bp region of the 5' c-*myc* region which replicates autonomously, is maintained as an episome *in vivo* (Sudo et al., 1990), and contains minimal consensus elements and footprints of potential initiators (Ariga et al., 1989). DNA sequences isolated from monkey cells by nascent strand extrusion also possess autonomous replication capacity by *Dpn*I and BrdU assay (Frappier and Zannis-Hadjopoulos, 1987); this *ors* activity can be subcloned and searched for consensus elements (Todd et al., 1995), or used as a probe for potential initiator proteins (Ruiz et al., 1995). Anti-cruciform immunoaffinity purified fragments 0.2 - 3.2 kb in size also confer autonomous replication capacity, under conditions where random human DNA fragments do not (Bell et al., 1991). Some human cDNA clones have also been demonstrated to replicate autonomously following transfection into human cells (Wu et al., 1993a), and one such fragment has been subsequently proven to act as a *bona fide* chromosomal origin in its native position on chromosome 6 (Wu et al., 1993b). More recently, autonomous replication assays have agreed with *in vivo* mapping of chromosomal origin function in the human heat shock protein 70 promoter region (Taira et al., 1994).

Autonomous replication can also be assayed by using plasmid DNA as template in an *in vitro* replication system based on human cell extracts (Pearson et al., 1991). While a background incorporation of radionucleotide occurs through repair activity in crude extracts, initiation occurs specifically within the autonomously replicating insert DNA, as assessed by EM mapping of *ors* clones (Pearson et al., 1994b) and Okazaki strand switch assay of plasmids carrying the 5' c-*myc* origin (Berberich et al., 1995).

REFERENCES

Almasan A, Mishra NC (1990) Characterization of a novel plasmid-like element in *Neurospora crassa* derived mostly from the mitochondrial DNA. *Nucl. Acids Res.* 18:5871-5877

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

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290: 457-464

Ariga H, Imamura Y, Iguchi-Ariga SMM (1989) DNA replication origin and transcriptional enhancer in c-*myc* gene share the c-*myc* protein binding sequences. *EMBO J.* 8:4273-4279

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

Ayer DE, Kretzner L, Eisenmann RN (1993) Mad: a heterodimeric partner for max that antagonizes myc transciptional activity. *Cell* 72:211-222

Barnett MA, Buckle VJ, Evans EP, Porter ACG, Rout D, Smith AG, Brown WRA (1993) Telomere directed fragmentation of mammalian chromosomes. *Nucl. Acids Res.* 21:27-36

Bell D, Sabloff M, Zannis-Hadjopoulos M, Price GB (1991) Anti-cruciform DNA affinity purification of active mammalian origins of replication. *Biochim. Biophys. Acta* 1089:299-308

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

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

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

Bellanne-Chantelot C, Lacroix B, Ougen P, Billault A, Beaufils S, Bertrand S, Georges I, Gilbert F, Susini L, Codani JJ, Gesnouin P, Pook S, Vaysseix G, Lu-Kuo J, Ried T, Ward D, Chumakov I, Le Paslier D, Barillot E, Cohen D (1992) Mapping the whole human genome by fingerprinting yeast artificial chromosomes. *Cell* 70:1059-1068

Benard M, Pallotta D, Pierron G (1992) Structure and identity of a late-replicating and transcriptionally active gene. *Exp. Cell Res.* 201:506-513

Benbow RM, Zhao J, Larson DD (1992) On the nature of origins of DNA replication in eukaryotes. *BioEssays* 14:661-670

Bennett ER, Naujokas M, Hassell JA (1989) Requirements for species-specific papovavirus DNA replication. *J. Virol.* 63:5371-5385

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

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

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

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

Bianchi A, Wells RD, Heintz NH, Caddle MS (1990) Sequences near the origin of replication of the DHFR locus of Chinese hamster ovary cells adopt left-handed Z-DNA and triplex structures. *J. Biol. Chem.* 265:21789-21796

Blackburn EH (1992) Telomerases. *Annu. Rev. Biochem.* 61:113-129

Bloom K (1993) The centromere frontier: kinetochore components, microtubule-based motility, and the CEN-value paradox. *Cell* 73:621-624

Bonnefoy E, Takahashi M, Rouviere-Yaniv J (1994) DNA-binding parameters of the HU protein of *Escherichia coli* to cruciform DNA. *J. Mol. Biol.* 242:116-129

Boulikas T (1992) Homeotic protein binding sites, origins of replication, and nuclear matrix anchorage sites share the ATTA and ATTTA motifs. *J. Cell. Biochem.* 50:111-123

Boulikas T (1993) Homeodomain protein binding sites, inverted repeats, and nuclear matrix attachment regions along the human β-globin gene complex. *J. Cell. Biochem.* 52:23-36

Boulikas T (1995) A model of transcription replication enhancers. *Oncol. Rep.* 2:171-181

Boulikas T, Kong CF (1993) Multitude of inverted repeats characterizes a class of anchorage sites of chromatin loops to the nuclear matrix. *J. Cell. Biochem.* 53:1-12

Bowater R, Aboul-ela F, Lilley DMJ (1991) Large-scale opening of supercoiled DNA in response to temperature and supercoiling in (A+T)-rich regions that promote low-salt cruciform extrusion. *Biochemistry* 30:11495-11506

Boyer JC, Thomas DC, Maher VM, McCormick JJ, Kunkel TA (1993) Fidelity of DNA replication by extracts of normal and malignantly transformed human cells. *Cancer Res.* 53:3270-3275

Brenot-Bosc F, Gupta S, Margolis RM, Fotedar R (1995) Changes in the subcellular localization of replication initiation proteins and cell cycle proteins during G1- to S-phase transition in mammalian cells. *Chromosoma* 103:517-527

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

Brewer BJ, Fangman WL (1991) Mapping replication origins in yeast chromosomes. *BioEssays* 13:317-322

Brewer BJ, Fangman WL (1993) Initiation at closely spaced replication origins in a yeast chromosome. *Science* 262:1728-1731

Brown CJ, Hendrich BD, Rupert JL, Lafreniere RG, Xing Y, Lawrence J, Willard HF (1992) The human *XIST* gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* 71:527-542

Brown WRA, Dobson MJ, MacKinnon P (1990) Telomere cloning and mammalian chromosome analysis. *J. Cell Sci.* 95:521-526

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

Burke DT, Carle GF, Olson MV (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* 236:806-812

Buxton JL, Chan JCT, Gilbert H, Clayton-Smith J, Burn J, Pembrey M, Malcolm S (1994) Angelman syndrome associated with a maternal 15q11-13 deletion of less than 200 kb. *Hum. Mol. Genet.* 3:1409-1413

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

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

Cardoso MC, Leonhardt H, Nadal-Ginard B (1993) Reversal of terminal differentiation and control of DNA replication: cyclin A and cdk2 specifically localize at subnuclear sites of DNA replication. *Cell* 74:979-992

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

Chaly N, Chen X (1993) Assembly of adenovirus-specific nuclear inclusions in lytically infected HeLa cells: an ultrastructural and cytochemical study. *Biochem. Cell Biol.* 71:475-487

Chartier FL, Keer JT, Sutcliffe MJ, Henriques DA, Mileham P, Brown SDM (1992) Construction of a mouse yeast artificial chromosome library in a recombination-deficient strain of yeast. *Nature Genet.* 1:132-136

Chattoraj DK, Mason RJ, Wickner S (1988) Mini-P1 plasmid replication: the autoregulation-sequestration complex. *Cell* 52:551-557

Chen C, Okayama H (1987) High-efficiency transformation of mammalian cells by plasmid DNA. *Mol. Cell. Biol.* 7:2745-2752

Chong JPJ, Mahbubani HM, Khoo CY, Blow JJ (1995) Purification of an MCM-containing complex as a component of the DNA replication licensing system. *Nature* 375:416-421

Clarke HJ (1992) Nuclear and chromatin composition of mammalian gametes and early embryos. *Biochem. Cell Biol.* 70:856-866

Clayton DA (1992) Transcription and replication of animal mitochondrial DNAs. *Int Rev Cytol* 141: 217-232

Cleaver JE (1994) It was a very good year for DNA repair. *Cell* 76:1-4

Coffman FD, Fresa KL, Hameed M, Cohen S (1993a) Characteristics of DNA replication in isolated nuclei initiated by an aprotinin-binding protein. *J. Cell. Biochem.* 51:157-164

Coffman FD, Georgoff I, Fresa KL, Sylvester J, Gonzalez I, Cohen S (1993) *In vitro* replication of plasmids containing human ribosomal gene sequences: origin localization and dependence on an aprotinin-binding cytosolic protein. *Exp. Cell Res.* 209:123-132

Cohen S, Hassin D, Karby S, Lavi S (1994) Hairpin structures are the primary amplification products: a novel mechanism for generation of inverted repeats during gene amplification. *Mol. Cell. Biol.* 14:7782-7791

Cooper MJ, Hatzivassiliou G, Miron S (1994) A rapid, PCR-based assay for episomal replication. *BioTechniques* 16:21-23

Counter CM, Gupta J, Harley CB, Leber B, Bacchetti S (1995) Telomerase activity in normal leukocytes and in hematologic malignancies. *Blood* 85:2315-2320

Coverley D, Laskey RA (1994) Regulation of eukaryotic DNA replication. *Annu. Rev. Biochem.* 63:745-776

Cross SM, Sanchez CA, Morgan CA, Schimke MK, Ramel S, Idzerda RL, Raskind WH, Reid BJ (1995) A p53-dependent mouse spindle checkpoint. *Science* 267:1353-1356

Dani GM, Zakian VA (1983) Mitotic and meiotic stability of linear plasmids in yeast. *Proc. Natl. Acad. Sci. USA* 80:3406-3410

de Lange T, Shiue L, Myers RM, Cox DR, Naylor SL, Killery AM, Varmus HE (1990) Structure and variability of human chromosome ends. *Mol. Cell. Biol.* 10:518-527

Delouya D, Nobrega FG (1991) Mapping of the ARS-like activity and transcription initiation sites in the non-canonical yeast mitochondrial *ori6* region. *Yeast* 7:51-60

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

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

DePamphilis ML (1993b) Origins of DNA replication in metazoan chromosomes. *J. Biol. Chem.* 268:1-4

Dershowitz A, Newlon CS (1993) The effect on chromosome stability of deleting replication origins. *Mol. Cell. Biol.* 13:391-398

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

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

Diffley JFX, Stillman B (1990) The intiation of chromosomal DNA replication in eukaryotes. *Trends Genet.* 6:427-432

Dijkwel PA, Hamlin JL (1988) Matrix attachment regions are positioned near replication initiation sites, genes, and an interamplicon junction in the amplified dihydrofolate reductase domain of Chinese hamster ovary cells. *Mol. Cell. Biol.* 8:5398-5409

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

Dimitrova D, Vassilev L, Anachkova B, Russev G (1993) Isolation and cloning of putative mouse DNA replication initiation sites: binding to nuclear protein factors. *Nucl. Acids Res.* 21:5554-5560

Dobbs DL, Shaiu WL, Benbow RM (1994) Modular sequence elements associated with origin regions in eukaryotic chromosomal DNA. *Nucl. Acids Res.* 22:2479-2489

Dolf G, Meyn RE, Curley D, Prather N, Story MD, Boman BM, Siciliano MJ, Hewitt RR (1991) Extrachromosomal amplification of the epidermal growth factor receptor gene in a human colon carcinoma cell line. *Genes Chrom. Cancer* 3:48-54

Dowdy SF, Hinds PW, Louie K, Reed SI, Arnold A, Weinberg RA (1993) Physical interaction of the retinoblastoma protein with human D cyclins. *Cell* 73:499-511

Dubey DD, Davis LR, Greenfeder SA, Ong LY, Zhu J, Broach JR, Newlon CS, Huberman JA (1991) Evidence suggesting that the *ARS* elements associated with silencers of yeast mating-type locus *HML* do not function as chromosomal DNA replication origins. *Mol. Cell. Biol.* 11:5346-5355

Durcan MJ, Goldman D (1993) Genomic imprinting: implications for behavioral genetics. *Behavior Genet.* 23:137-143

Earnshaw WC, Bernat RL (1991) Chromosomal passengers: toward an integrated view of mitosis. *Chromosoma* 100:139-146

Earnshaw WC, Tomkiel JE (1992) Centromere and kinetochore structure. *Curr. Opin. Cell Biol.* 4:86-93

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

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

Farr C, Fantes J, Goodfellow P, Cooke H (1991) Functional reintroduction of human telomeres into mammalian cells. *Proc. Natl. Acad. Sci. USA* 88:7006-7010

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

Forget BG (1993) YAC transgenes: bigger is probably better. *Proc. Natl. Acad. Sci. USA* 90:7909-7911

Foss M, McNally FJ, Laurenson P, Rine J (1993) Origin recognition complex (ORC) in transcriptional silencing and DNA replication in *S. cerevisiae. Science* 262:1838-1843

Fotedar R, Roberts JM (1992) Cell cycle regulated phosphorylation of RPA-32 occurs within the replication initiation complex. *EMBO J.* 11:2177-2187

Frappier L, O'Donnell M (1991) Epstein-Barr nuclear antigen 1 mediates a DNA loop within the latent replication origin of Epstein-Barr virus. *Proc. Natl. Acad. Sci. USA* 88:10875-10879

Frappier L, Price GB, Martin RG, Zannis-Hadjopoulos M (1987) Monoclonal antibodies to cruciform DNA structures. *J. Mol. Biol.* 193:751-758

Frappier L, Price GB, Martin RG, Zannis-Hadjopoulos M (1989) Characterization of the binding specificity of two anticruciform DNA monoclonal antibodies. *J. Biol. Chem.* 264:334-341

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

Fuller RS, Kaguni JM, Kornberg A (1981) Enzymatic replication of the origin of the *Escherichi coli* chromosome. *Proc. Natl. Acad. Sci. USA* 78:7370-7374

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

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

Galli I, Iguchi-Ariga SMM, Ariga H (1993) Mammalian genomic sequences can substitute for the SV40 AT stretch in sustaining replication of the SV40 origin of replication. *FEBS Lett.* 318:335-340

Gasser SM, Laemmli UK (1986) Cohabitation of scaffold binding proteins with upstream/enhancer elements of three developmentally regulated genes of D. melanogaster. *Cell* 46:521-530

Gassmann M, Donoho G, Berg P (1995) Maintenance of an extrachromosomal plasmid vector in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. USA* 92:1292-1296

Gaubatz JW, Flores SC (1990) Tissue-specific and age-related variations in repetitive sequences of mouse extrachromosomal circular DNAs. *Mutation Res.* 237:29-36

Gekeler V, Wilisch A, Probst G, Kugel A, Brischwein K, Engelcke M, Probst H (1993) Staurosporine suppresses replicon initiation in mammalian cells. *FEBS Lett.* 327:150-156

Georgiev GP, Vassetzky YS, Luchnik AN, Chernokhvostov VV, Razin SV (1991) Nuclear skeleton, DNA domains and control of replication and transcription. *Eur. J. Biochem.* 200:613-624

Ghivizzani SC, Madsen CS, Nelen MR, Ammini CV, Hauswirth WW (1994) In organello footprint analysis of human mitochondrial DNA: human mitochondrial transcription factor A interactions at the origin of replication. *Mol. Cell. Biol.* 14:7717-7730

Gilbert D, Cohen SN, Grummt F (1989) Autonomous replication in mouse cells: a correction. *Cell* 56:143-144

Gilbert DM, Miyazawa H, Nallaseth FS, Ortega JM, Blow JJ, DePamphilis ML (1993) Site-specific initiation of DNA replication in metazoan chromosomes and the role of nuclear organization. *Cold Spring Harbor Symp. Quant. Biol.* LVIII:475-485

Gilbert DM, Neilson A, Miyazawa H, DePamphilis ML, Burhans WC (1995) Mimosine arrests DNA synthesis at replication forks by inhibiting deoxyribonucleotide metabolism. *J. Biol. Chem.* 270:9597-9606

Griffin BE, Björck E, Bjursell G, Lindahl T (1981) Sequence complexity of circular Epstein-Barr virus DNA in transformed cells. *J. Virol.* 40:11-19

Guo ZS, Gutierrex C, Heine U, Sogo JM, DePamphilis ML (1989) Origin auxilliary sequences can facilitate initiation of simian virus 40 DNA replication in vitro as they do in vivo. *Mol. Cell. Biol.* 9:3593-3602

Haaf T (1995) The effects of 5-azacytidine and 5-azadeoxycytidine on chromosome structure and function: implications for methylation-associated cellular processes. *Pharmac. Ther.* 65:19-46

Hadlaczky G, Praznovszky T, Cserpan I, Kereso J, Peterfy M, Kelemen I, Atalay E, Szeles A, Szelei J, Tubak V, Burg K (1991) Centromere formation in mouse cells cotransformed with human DNA and a dominant marker gene. *Proc. Natl. Acad. Sci. USA* 88:8106-8110

Hahnenberger KM, Baum MP, Polizzi CM, Carbon J, Clarke L (1989) Constuction of artificial minichromosomes in the fission yeast *Schizosaccharomyces pombe*. *Proc. Natl. Acad. Sci. USA* 86:577-581

Hamlin JL, Mosca PJ, Dijkwel PA, Lin HB (1993) Initiation of replication at a mammalian chromomal origin. *Cold Spring Harbor Symp. Quant. Biol.* LVIII:467-474

Hamlin JL, Mosca PJ, Levenson VV (1994) Defining origins of replication in mammalian cells. *Biochim. Biophys. Acta* 1198:85-111

Hanish JP, Yanowitz JL, de Lange T (1994) Stringent sequence requirements for the formation of human telomeres. *Proc. Natl. Acad. Sci.* USA 91:8861-8865

Harley VR, Goodfellow PN (1994) The biochemical role of *sry* in sex determination. *Mol. Reprod. Devel.* 39:184-193

Harriman W, Volk H, Defranoux N, Wabl M (1993) Immunoglobulin class switch recombination. *Annu. Rev. Immunol.* 11:361-384

Heald R, McLoughlin M, McKeon F (1993) Human wee1 maintains mitotic timing by protecting the nucleus from cytoplasmically activated cdc2 kinase. *Cell* 74:463-373

Heichman KA, Roberts JM (1994) Rules to replicate by. *Cell* 79:557-562

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

Held PG, Heintz NH (1992) Eukaryotic replication origins. *Biochim. Biophys. Acta* 1130:235-246

Hieter P, Mann C, Snyder M, Davis RW (1985) Mitotic stability of yeast chromosomes: a colony color assay that measures nondisjunction and chromosome loss. *Cell* 40:381-392

Higgins DG, Sharp PM (1989) Fast and sensitive multiple sequence alignments on a microcomputer. *Comput. Applic. Biosci.* 5:151-153.

Hozak P, Hassan AB, Jackson DA, Cook PR (1993) Visualization of replication factories attached to a nucleoskeleton. *Cell* 73:361-373

Huang RY, Kowalski D (1993) A DNA unwinding element and an ARS consensus comprise a replication origin within a yeast chromosome. *EMBO J.* 12:4521-4531

Huberman JA (1995) A license to replicate. *Nature* 375:360-361

Huberman JA, Riggs AD (1968) On the mechanism of DNA replication in mammalian chromosomes. *J. Mol. Biol.* 32:327-341

Hwang DS, Kornberg A (1990) A novel protein binds a key origin sequence to block replication of an E. coli minichromosome. *Cell* 63:325-331

Hsu MT (1985) Electron microscopic evidence for the cruciform structrue in intracellular SV40 DNA. *Virology* 143:617-621

Hyman BC, Garcia-Garcia F (1993) Titration of replication activity by increasing ARS dosage in yeast plasmids. *Curr. Genet.* 23:141-147

Iguchi-Ariga SMM, Itani T, Kiji Y, Ariga H (1987) Possible function of the c-*myc* product: promotion of cellular DNA replication. *EMBO J.* 6:2365-2371

Iguchi-Ariga SMM, Ogawa N, Ariga H (1993) Identification of the initiation region of DNA replication in the murine heavy chain gene and possible function of the octamer motif as a putative DNA replication origin in mammalian cells. *Biochim. Biophys. Acta* 1172:73-81

Jankelevich S, Kolman JL, Bodnar JW, Miller G (1992) A nuclear matrix attachment region organizes the Epstein-Barr viral plasmid in Raji cells into a single DNA domain. *EMBO J.* 11:1165-1176

Kang CH, Berger I, Lockshin C, Ratliff R, Moyzis R, Rich A (1995) Stable loop in the crystal structure of the intercalated four-stranded cytosine-rich metazoan telomere. *Proc. Natl. Acad. Sci. USA* 92:3874-3878

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

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

King AJ, van der Vliet PC (1994) A precursor terminal protein-trinucleotide intermediate during initiation of adenovirus DNA replication: regeneration of molecular ends in vitro by a jumping back mechanism. *EMBO J.* 13:5786-5792

Kipling D, Kearsey SE (1990) Reversion of autonomously replicating sequence mutations in *Saccharomyces cerevisiae*: creatio of a eucaryotic replication origin within procaryotic vector DNA. *Mol. Cell. Biol.* 10:265-272

Kirchmaier AL, Sugden B (1995) Plasmid maintenance of derivitaves of oriP of Epstein-Barr virus. *J Virol.* 69:1280-1283

Kitsberg D, Selig S, Brandeis M, Simon I, Keshet I, Driscoll DJ, Nicholls RD, Cedar H (1993a) Allele-specific replication timing of imprinted gene regions. *Nature* 364:459-463

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

Klar AJS, Bonaduce MJ (1993) The mechanism of fission yeast mating-type interconversion: evidence for two types of epigenetically inherited chromosomal imprinted events. *Cold Spring Harbor Symp. Quant. Biol.* LVIII:457-465

Knoll JHM, Wagstaff J, Lalande M (1993) Cytogenetic and molecular studies in the Prader-Willi and Angelman syndromes: an overview. *Am J. Med. Genet.* 46:2-6

Knoll JHM, Cheng SD, Lalande M (1994) Allele specificity of DNA replication timing in the Angelman/Prader-Willi syndrome imprinted chromosomal region. *Nature Genet.* 6:41-46

Kornberg A, Baker TA (1992) *DNA Replication (2nd ed).* Freeman, New York

Kosztolanyi G, Mehes K, Hook EB (1991) Inherited ring chromosomes: an analysis of published cases. *Hum. Genet.* 87:320-324

Krajewski WA, Razin SV (1992) Organization of specific DNA sequence elements in the region of the replication origin and matrix attachment site in the chicken α-globin gene domain. *Mol. Gen. Genet.* 235:381-388

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

Krysan PJ, Haase SB, Calos MP (1989) Isolation of human sequences that replicate autonomously in human cells. *Mol. Cell. Biol.* 9:1026-1033

Krysan PJ, Smith JG, Calos MP (1993) Autonomous replication in human cells of multimers of specific human and bacterial DNA sequences. *Mol. Cell. Biol.* 13:2688-2696

Kubota Y, Mimura S, Nishimoto S, Takisawa H, Nojima H (1995) Identification of the yeast MCM3-related protein as a component of Xenopus DNA replication licensing factor. *Cell* 81:601-609

Kuspa A, Vollrath D, Cheng Y, Kaiser D (1989) Physical mapping of the *Myxococcus xanthus* genome by random cloning in yeast artificial chromosomes. *Proc. Natl. Acad. Sci. USA* 86:8917-8921

Lalande M (1994) In and around *SNRPN*. *Nature Genet.* 8:5-7

Lambowitz AM, Belfort M (1993) Introns as mobile genetic elements. *Annu. Rev. Biochem.* 62:587-622

Landry S, Zannis-Hadjopoulos M (1991) Classes of automously replicating sequences are found among early replicating monkey DNA. *Biochim. Biophys. Acta* 1088:234-244

LaSalle JM, Lalande M (1995) Domain organization of allele-specific replication within the *GABRB3* gene cluster requires a biparental chromosome 15q11-q13 contribution. *Nature Genet.* 9:386-394

Leu TH, Hamlin JL (1992) Activation of a mammalian origin of replication by chromosomal rearrangement. *Mol. Cell. Biol.* 12:2804-2812

Lewis DL, Farr CL, Farquhar AL, Kaguni LS (1994) Sequence, organization, and evolution of the A+T region of Drosophila melanogaster mitochondrial DNA. *Mol. Biol. Evol.* 11:523-538

Li JJ, Alberts BM (1992) Eukaryotic initiation rites. *Nature* 357:114-115

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

Li JJ, Kelly TJ (1984) Simian virus 40 DNA replication in vitro. *Proc. Natl. Acad. Sci. USA* 81:6973-6977

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

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

Lin CC, Sasi R, Lee C, Fan YS, Court D (1993) Isolation and identification of a novel tandemly repeated DNA sequence in the centromeric region of human chromosome 8. *Chromosoma* 102:333-339

Lindahl T (1993) Instability and decay of the primary structure of DNA. *Nature* 362:709-715

Lindahl T, Barnes DE (1992) Mammalian DNA ligases. *Annu. Rev. Biochem.* 61:251-281

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

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

Lusky M, Botchan M (1981) Inhibition of SV40 replication in simian cells by specific pBR322 DNA sequences. *Nature* 293:79-81

Lyon MF (1993) Epigenetic inheritance in mammals. *Trends Genet.* 9:123-128

Ma C, Martin S, Trask B, Hamlin JH (1993) Sister chromatid fusion initiates amplification of the dihydrofolate reductase gene in Chinese hamster cells. *Genes Dev.* 7:605-620

Mader S, Chambon P, White JH (1993) Defining a minimal estrogen receptor DNA binding domain. *Nucl. Acids Res.* 21:1125-1132

Madine MA, Khoo CY, Mills AD, Laskey RA (1995) MCM3 complex required for cell cycle regulation of DNA replication in vertebrate cells. *Nature* 375:421-424

Madsen CS, Chivizzani SC, Hauswirth WW (1993) Protein binding to a single termination-associated sequence in the mitochondrial D-loop region. *Mol. Cell. Biol.* 13:2162-2171

Mah DCW, Shihab-El-Deen A, Price GB, Zannis-Hadjopoulos M (1992) ORS12, a mammalian autonomously replicating DNA sequence, is present at the centromere of CV-1 cell chromosomes. *Exp. Cell Res.* 203:435-442

Mah DCW, Dijkwel PA, Todd A, Klein V, Price GB, Zannis-Hadjopoulos M (1993) *ors12*, a mammalian autonomously replicating sequence, associates with the nuclear matrix in a cell cycle-dependent manner. *J. Cell Sci.* 105:807-818

Mahbubani HM, Paull T, Elder JK, Blow JJ (1992) DNA replication initiates at multiple sites on plasmid DNA in Xenopus egg extracts. *Nucl. Acids Res.* 20:1457-1462

Manfredi JJ, Prives C (1993) Primary rat cells expressing a hybrid polyomavirus-simian virus 40 large T antigen have altered growth properties. *J. Virol.* 67:4750-4759

Mann C, Davis RW (1986) Structure and sequence of centromeric DNA of chromosome 4 in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 6:241-245

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

Marahrens Y, Stillman B (1994) Replicator dominance in a eukaryotic chromosome. *EMBO J.* 13:3395-3400

Marians KJ (1992) Prokaryotic DNA replication. *Annu. Rev. Biochem.* 61:673-719

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

Masukata H, Satoh H, Obuse C, Okazaki T (1993) Autonomous replication of human chromosomal DNA fragments in human cells. *Mol. Biol. Cell* 4:1121-1132

McAlear M, Ward GK, Mckenzie R, Price GB, Zannis-Hadjopoulos M (1989) Biphasic activation of DNA replication. *Mol. Gen. (Life Sci. Adv.)* 8:11-14

McWhinney C, Leffak M (1988) Episomal persistence of a plasmid containing human c-*myc* DNA. *Cancer Cells* 6:467-471

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

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

Miller DL, Dill FJ, Lichter JB, Kidd KK, Goodfellow PJ (1992) Isolation and high-resolution mapping of new DNA markers from the pericentromeric region of chromosome 10. *Genomics* 13:601-606

Moarefi IF, Small D, Gilbert I, Hopfner M, Randall SK, Schneider C, Russo AA, Ramsperger U, Arthur AK, Stahl H, Kelly T, Fanning E (1993) Mutation of the cyclin-dependent kinase phosphorylation site in simian virus 40 (SV40) large T antigen specifically blocks SV40 origin DNA unwinding. *J. Virol.* 67:4992-5002

Mosca PJ, Lin HB, Hamlin JL (1995) Mimosine, a novel inhibitor of DNA replication, binds to a 50 kDa protein in Chinese hamster cells. *Nucl. Acids Res.* 23:261-268

Mukherjee S, Erickson H, Bastia D (1988) Enhancer-origin interaction in plasmid R6K involves a DNA loop mediated by initiator protein. *Cell* 52:375-383

Murakami Y, Naitou M, Hagiwara H, Shibata T, Ozawa M, Sasanuma S, Sasanuma M, Tsuchiya Y, Soeda E, Yokoyama K, Yamazaki M, Tashire H, Eki T (1995) Analysis of nucleotide sequence of chromosome VI from *Saccharomyces cerevisiae*. *Nature Genet.* 10:261-268

Murray AW, Szostak JW (1983) Construction of artificial chromosomes in yeast. *Nature* 305:189-193

Nakajima M, Sheikh QI, Yamaoka K, Yui Y, Kajiwara S, Shishido K (1993) Bending of DNA segments with *Saccharomyces cerevisiae* autonomously replicating sequence activity, isolated from basidiomycete mitochondrial linear plasmids. *Mol. Gen. Genet.* 237:1-9

Natale DA, Schubert AE, Kowalski D (1992) DNA helical stability accounts for mutational defects in a yeast replication origin. *Proc. Natl. Acad. Sci. USA* 89:2654-2658

Negishi Y, Nishita Y, Saëgusa Y, Kakazaki I, Galli I, Kihara F, Tamai K, Miyajima N, Iguchi-Ariga SMM, Ariga H (1994) Identification and cDNA cloning of single-stranded DNA binding proteins that interact with the region upstream of the human c-*myc* gene. *Oncogene* 9:1133-1143

Newlon CS (1993) Two jobs for the origin recognition complex. *Science* 262:1830-1831

Newlon CS, Theis JF (1993) The structure and function of yeast ARS elements. *Curr. Opin. Genet. Devel.* 3:752-758

Nicholls RD (1993) Genomic imprinting and uniparental disomy in Angelman and Prader-Willi syndromes: a review. *Am. J. Med. Genet.* 46:16-25

Nicholls RD (1994) New insights reveal complex mechanisms involved in genomic imprinting. *Am. J. Hum. Genet.* 54:733-740

Niller HH, Glaser G, Knuchel R, Wolf H (1995) Nucleoprotein complexes and DNA 5'-ends at oriP of Epstein-Barr virus. *J. Biol. Chem.* 270:12864-12868

Nobile C, Martin RG (1986) Stable stem-loop and cruciform DNA structures: isolation of mutants with rearrangements of the palindromic sequence at the simian virus 40 replication origin. *Intervirology* 25:158-171

O'Keefe RT, Henderson SC, Spector DL (1992) Dynamic organization of DNA replication in mammalian cell nuclei: spatially and temporally defined replication of chromosome-specific α-satellite DNA sequences. *J. Cell Biol.* 116:1095-1110

Oliver S (1995) Size is important, but... *Nature Genet.* 10:253-254

Parry D, Bates S, Mann DJ, Peters G (1995) Lack of cyclin D-cdk complexes in Rb-negative cells correlates with high levels of $p16^{INK4/MTS1}$ tumor suppressor gene product. *EMBO J.* 14:503-511

Paull TT, Johnson RC (1995) DNA looping by Saccharomyces cerevisiae high mobility group proteins NHP6A/B - consequences for nucleoprotein complex assembly and chromatin condensation. *J. Biol. Chem.* 270:8744-8754

Pearson CE, Frappier L, Zannis-Hadjopoulos M (1991) Plasmids bearing mammalian DNA-replication origin-enriched (*ors*) fragments initiate semiconservative replication in a cell-free system. *Biochim. Biophys. Acta* 1090:156-166

Pearson CE, Ruiz MT, Price GB, Zannis-Hadjopoulos M (1994a) Cruciform DNA binding protein in HeLa cell extracts. *Biochemistry* 33:14185-14196

Pearson CE, Shihab-El-Deen A, Price GB, Zannis-Hadjopoulos M (1994b) Electron microscopic analysis of in vitro replication products of ors 8, a mammalian origin enriched sequence. *Somat. Cell Mol. Genet.* 20:147-152

Peng M, Lemke PA, Singh NK (1993) A nucleotide sequence involved in replicative transformation of a filamentous fungus. *Curr. Genet.* 24:114-121

Ramsperger U, Stahl H (1995) Unwinding of chromatin by the SV40 large T antigen DNA helicase. *EMBO J.* 14:3215-3225

Rao BS, Martin RG (1988) DpnI assay for DNA replication in animal cells: enzyme-resistant material can result from factors not related to replication. *Nucl. Acids Res.* 16:4171

Rao BS, Zannis-Hadjopoulos M, Price GB, Reitman M, Martin RG (1990) Sequence similarities among monkey *ori*-enriched (*ors*) fragments. *Gene* 87:233-242

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

Rattner JB (1991) The structure of the mammalian centromere. *BioEssays* 13:51-56

Razin A, Cedar H (1994) DNA methylation and genomic imprinting. *Cell* 77:473-476

Razin SV, Yassetskii YS, Georgiev GP (1990) Replication origins of higher eukaryotes-- a minireview. *Biomed. Sci.* 1:18-22

Rein T, Muller M, Zorbas H (1994) In vivo footprinting of the IRF-1 promoter: inducible occupation of a GAS element next to a persistent structural alteration of the DNA. *Nucl. Acids Res.* 22:3033-3037

Reis A, Kunze J, Ladanyi L, Enders H, Klein-Vogler U, Niemann G (1993) Exclusion of the $GABA_A$-receptor β3 subunit as the Angelman's syndrome gene. *Lancet* 341:122-123

Renault S, Degroote F, Picard G (1993) Despite its high representation in extrachromosomal circular DNAs from *Drosophila* embryos, the dodecasatellite does not allow autonomous replication in cultured cells. *Biol. Cell* 79:51-54

Rivier DH, Rine J (1992) An origin of DNA replication and a transcription silencer require a common element. *Science* 256:659-663

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

Runge KW, Zakian VA (1993) *Saccharomyces cerevisiae* linear chromosome stability (*lcs*) mutants increase the loss rate of artificial and natural linear chromosomes. *Chromosoma* 102:207-217

Ruiz MT, Pearson CE, Nielsen T, Price GB, Zannis-Hadjopoulos M (1995) Cofractionation of HeLa cell replication proteins with *ors*-binding activity. *J. Cell. Biochem.* 58:221-236

Saëgusa Y, Sato M, Galli I, Nakagawa T, Ono N, Iguchi-Ariga SMM, Ariga H (1993) Stimulation of SV40 DNA replication and transcription by *Alu* family sequence. *Biochim. Biophys. Acta* 1172:274-282

Sanchez JA, Marek D, Wangh LJ (1992) The efficiency and timing of plasmid DNA replication in *Xenopus* eggs: correlations to the extent of prior chromatin assembly. *J. Cell Sci.* 103:907-918

Savilahti H, Bamford DH (1993) Protein-primed DNA replication: role of inverted terminal repeats in the *Escherichia coli* bacteriophage PRD1 life cycle. *J. Virol.* 67:4696-4703

Schaeffer L, Roy R, Humbert S, Moncollin V, Vermeulen W, Hoeijmakers JHJ, Chambon P, Egly JM (1993) DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor. *Science* 260:58-63

Schaper S, Messer W (1995) Interaction of the initiator protein dnaA of *Escherichia coli* with its DNA target. *J. Biol. Chem.* 270:17622-17626

Schatz DG, Oettinger MA, Schlissel MS (1992) V(D)J recombination: molecular biology and regulation. *Annu. Rev. Immunol.* 10:359-383

Schlessinger D (1990) Yeast artificial chromosomes: tools for mapping and analysis of complex genomes. *Trends Genet.* 6:248-258

Schuller G, Holt SE, Hsu J, Wilson VG (1994) The bovine papillomavirus type 1 genome contains multiple loci of static DNA bending, but bends are absent from the functional origin of replication. *Virus Res.* 31:203-217

Shay JW, Werbin H (1992) New evidence for the insertion of mitochondrial DNA into the human genome: significance for cancer and aging. *Mutation Res.* 275:227-235

Shervington A (1993) T2G4 telomere repeats does not provide telomere function to a BPV-1 containing DNA fragment in mouse C127 cells. *Biochem. Mol. Biol. Int.* 29:661-672

Shihab-El-Deen A, Wu C, Hamerton JL, Zannis-Hadjopoulos M, Price GB (1993) Chromosomal localization of a sequence with in vivo activity for initiation of DNA replication. *Somat. Cell Mol. Genet.* 19:103-109

Sinnett D, Wagstaff J, Glatt K, Woolf E, Kirkness EJ, Lalande M (1993) High-resolution mapping of the γ-aminobutyric acid receptor subunit β3 and α5 gene cluster on chromosome 15q11-q13, and localization of breakpoints in two Angelman syndrome patients. *Am. J. Hum. Genet.* 52:1216-1229

Smith DR, Smyth AP, Moir DT (1990) Amplification of large artificial chromosomes. *Proc. Natl. Acad. Sci. USA* 87:8242-8246

Smith JG, Calos MP (1995) Autonomous replication in *Drosophila melanogaster* tissue culture cells. *Chromosoma* 103:597-605

Smithies O (1982) The control of globin and other eukaryotic genes. *J. Cell Physiol* (Suppl.) 1:137-143

Snyder M, Sapolsky RJ, Davis RW (1988) Transcription interferes with elements important for chromosome maintenance in *Saccharomyces cerevisiae. Mol. Cell. Biol.* 8:2184-2194

Staden R (1984) Computer methods to locate signals in nucleic acid sequences. *Nucl. Acids Res.* 12:505-519

Starborg M, Brundell E, Gell K, Larsson C, White I, Daneholt B, Hoog C (1995) A murine replication protein accumulates temporarily in the heterochromatic regions of nuclei prior to initiation of DNA replication. *J. Cell Sci.* 108:927-934

Staudt LM, Lenardo MJ (1991) Immunoglobulin gene transcription. *Annu. Rev. Immunol.* 9:373-398

Steiner NC, Clarke L (1994) A novel epigenetic effect can alter centromere function in fission yeast. *Cell* 79:865-874

Stillman B (1994) Initiation of chromosomal DNA replication in eukaryotes. *J. Biol. Chem.* 269:7047-7050

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

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

Sun TQ, Fenstermacher DA, Vos JMH (1994) Human artificial episomal chromosomes for cloning large DNA fragments in human cells. *Nature Genet.* 8:33-41

Surosky RT, Newlon CS, Tye BK (1986) The mitotic stability of deletion derivatives of chromosome III in yeast. *Proc. Natl. Acad. Sci. USA* 83:414-418

Sutcliffe JS, Nakao M, Christian S, Orstavik KH, Tommerup N, Ledbetter DH, Beaudet AL (1994) Deletions of a differentially methylated CpG island at the *SNRPN* gene define a putative imprinting control region. *Nature Genet.* 8:52-58

Suzuki M, Savoysky E, Izunta S, Tatebe M, Okajima T, Yoshida S (1993) RNA priming coupled with DNA synthesis on natural template by calf thymus DNA polymerase α -primase. *Biochemistry* 32:12782-12792

Szostak JW, Blackburn EH (1982) Cloning yeast telomeres on linear plasmid vectors. *Cell* 29:245-255

Taira T, Iguchi-Ariga SMM, Ariga H (1994) A novel DNA replication origin identified in the human heat shock protein 70 gene promoter. *Mol. Cell. Biol.* 14:6386-6397

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

Tanaka S, Halter D, Livingstone-Zatchej M, Reszel B, Thoma F (1994) Transcription through the yeast origin of replication ARS1 ends at the ABF1 binding site and affects extrachromosomal maintenance of minichromosomes. *Nucl. Acids Res.* 22:3904-3910

Tang WJ, Berger SL, Triezenberg SJ, Folk WR (1987) Nucleotides in the polyomavirus enhancer that control viral transcription and DNA replication. *Mol. Cell. Biol.* 7:1681-1690

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

Taylor JH (1960) Asynchronous duplication of chromosomes in cultured cells of chinese hamster. *J. Biophys. Biochem. Cytol.* 7:455-464

Thorsness PE, Fox TD (1990) Escape of DNA from mitochondria to the nucleus in *Saccharomyces cerevisiae*. *Nature* 346:376-379

Todd A, Landry S, Pearson CE, Khoury V, Zannis-Hadjopoulos M (1995) Deletion analysis of minimal sequence requirements for autonomous replication of *ors8*, a monkey early-replicating DNA sequence. *J. Cell. Biochem.* 57:280-289

Tran CT, Caddle MS, Calos MP (1993) The replication behavior of *Saccharomyces cerevisiae* DNA in human cells. *Chromosoma* 102:129-136

Traver CN, Klapholz S, Hyman RW, Davis RW (1989) Rapid screening of a human genomic library in yeast artificial chromosomes for single-copy sequences. *Proc. Natl. Acad. Sci. USA* 86:5898-5902

Tribioli C, Biamonti G, Giacca M, Colonna M, Riva S, Falaschi A (1987) Characterization of human DNA sequences synthesized at the onset of S-phase. *Nucl. Acids Res.* 15:10211-10231

Troelstra C, van Gool A, de Wit J, Vermeulen W, Bootsma D, Hoeijmakers JHJ (1992) *ERCC6*, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes. *Cell* 71:939-953

Tyler-Smith C (1993) Mammalian chromosome structure. *Curr. Opin. Genet. Devel.* 3:390-397

Tyler-Smith C, Oakey RJ, Larin Z, Fisher RB, Crocker M, Affara NA, Ferguson-Smith MA, Muenke M, Zuffardi O, Jobling MA (1993) Localization of DNA sequences required for human centromere function through an analysis of rearranged Y chromosomes. *Nature Genet.* 5:368-375

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

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

Vassilev L, Johnson EM (1988) Evaluation of autonomous plasmid replication in transfected mammalian cells. *Nucl. Acids Res.* 16:7742

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

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

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

Wadakiyama Y, Kiyama R (1995) Conservation and periodicity of DNA bend sites in the human beta-globin gene locus. *J. Biol. Chem.* 270:12439-12445

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

Wang PZ, Projan SJ, Henriquez V, Novick, RP (1993) Origin recognition specificity in pT181 plasmids is determined by a functionally asymmetric palindromic DNA element. *EMBO J.* 12:45-52

Wang SS, Zakian VA (1990) Telomere-telomere recombination provides an express pathway for telomere acquisition. *Nature* 345:456-458

Wanka F (1991) Control of eukaryotic DNA replication at the chromosomal level. *BioEssays* 13:613-618

Ward GK, McKenzie R, Zannis-Hadjopoulos M, Price GB (1990) The dynamic distribution and quantification of DNA cruciforms in eukaryotic nuclei. *Exp. Cell Res.* 188:235-246

Ward GK, Shihab-El-Deen A, Zannis-Hadjopoulos M, Price GB (1991) DNA cruciforms and the nuclear supporting structure. *Exp. Cell Res.* 195:92-98

Welch PJ, Wang JYJ (1993) A C-terminal protein-binding domain in the retinoblastoma protein regulates nuclear c-abl tyrosine kinase in the cell cycle. *Cell* 75:779-790

White E (1994) p53, guardian of Rb. *Nature* 371:21-22

Williams DL, Kowalski D (1993) Easily unwound DNA sequences and hairpin structures in the Epstein-Barr virus origin of plasmid replication. *J. Virol.* 67:2707-2715

Wu C, Friedlander P, Lamoureux C, Zannis-Hadjopoulos M, Price GB (1993a) cDNA clones contain autonomous replication activity. *Biochim. Biophys. Acta* 1174:241-257

Wu C, Zannis-Hadjopoulos M, Price GB (1993b) *In vivo* activity for initiation of DNA replication resides in a transcribed region of the human genome. *Biochim. Biophys. Acta* 1174:258-266

Wu Y, Hickey R, Lawlor K, Wills P, Yu F, Ozar H, Starr R, Quan JY, Lee M, Malkas L (1994) A 17S multiprotein form of murine cell DNA polymerase mediates polyomavirus DNA replication in vitro. *J. Cell. Biochem.* 54:32-46

Yan H, Newport J (1995) An analysis of the regulation of DNA synthesis by cdk2, Cip1, and licensing factor. *J. Cell Biol.* 129:1-15

Yates JL, Guan N (1991) Epstein-Barr virus-derived plasmids replicate only once per cell cycle and are not amplified after entry into cells. *J. Virol.* 65:483-488

Yeo JP, Aleruccio F, Toh BH (1994) A new chromosomal protein essential for mitotic spindle assembly. *Nature* 367:288-291

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

Yu GL, Blackburn EH (1990) Amplification of tandemly repeated origin control sequences confers a replication advantage on rDNA replicons in Tetrahymena thermophila. *Mol. Cell. Biol.* 10:2070-2080

Zannis-Hadjopoulos M, Frappier L, Khoury M, Price GB (1988) Effect of anti-cruciform DNA monoclonal antibodies on DNA replication. *EMBO J.* 7:1837-1844

Zhu JY, Abate M, Rice PW, Cole CN (1991) The ability of simian virus 40 large T antigen to immortalize primary mouse embryo fibroblasts cosegregates with its ability to bind to p53. *J. Virol.* 65:6872-6880

Zhu JG, Carlson DL, Dubey DD, Sharma K, Huberman JA (1994) Comparison of the two major ARS elements of the URA4 replication origin regions with other ARS elements in the fission yeast, *Schizosaccharomyces pombe. Chromosoma* 103:414-422

CHAPTER TWO

THESIS GOALS

Autonomous replication assays represent a powerful technique for the functional mapping of mammalian origins of DNA replication, opening the door to the isolation of new origins, and to further studies that will increase our knowledge of the nature of replication origins in higher eukaryotes. *In vivo* transfection and *in vitro* replication systems have been developed for the genetic assay of origin function. However, the question remains: are these techniques sufficiently sensitive to detect, reliably, replication activity in cloned origin sequences? To investigate this problem, fragments from a known mammalian origin, previously well-defined by a series of alternative physical mapping methods, will be assayed for autonomous replication, using the transfection and *in vitro* techniques developed in our laboratory.

If mammalian autonomous replication assays do have the sensitivity and specificity needed to identify sequences containing *cis*-active origin control elements, they will subsequently be applied to the isolation of large numbers of new replication origins. To date, only a handful of replication origins have been identified in vertebrate systems, including those associated with histone H5 (Dobbs et al., 1994) and α-globin (Krajewski and Razin, 1992) genes in the chicken, mouse origins mapped in the immunoglobulin heavy chain enhancer (Ariizumi et al., 1993) and at both ends of the adenosine deaminase gene (Carroll et al., 1993; Virta-Pearlman et al., 1993), origins mapped near or in the dihydrofolate reductase (Hamlin et al., 1993), rhodopsin (Gale et al., 1992), ribosomal protein S14 (Tasheva and Roufa, 1994) and *CAD* (Kelly et al., 1995) loci in hamster cell lines, and monkey *ors* clones (Zannis-Hadjopoulos et al., 1992). In the human system, the list of identified chromosomal origins remains short: origins mapped immediately 5' to exon I of c-*myc* (Vassilev and Johnson, 1990) and *hsp70* (Taira et al., 1994), the lamin B2-associated origin at 19p13.3 (Biamonti et al., 1992), the 343 origin located at the tip of the long arm of chromosome 6 (Shihab-El-Deen et al., 1993), and an origin mapped between the δ and β genes of the human β-globin domain (Kitsberg et al., 1993). Autonomous replication assays will be adapted to the screening of large numbers of sequences for origin activity, using as a starting point libraries enriched for origins by anti-cruciform immunoaffinity purification (Bell et al., 1991). If a large number of origins from a single system (the human) can be identified, they could be sequenced and analyzed for the presence of common sequence elements in an effort to determine the *cis*-active sequences required for origin function.

Where a large genomic region has been cloned and sequenced, the possible presence of a replication origin can be investigated by identifying clusters of elements associated with

origin activity, such as transcription factor binding sites, matrix attachment regions, DNA unwinding elements, and cruciform or bent secondary structures. Autonomous replication assays then provide a convenient, complementary method for screening a series of contiguous clones in an effort to identify the ones most likely to include a functional origin. Indeed, the convenience of autonomous replication studies, and the possibility to adapt *in vitro* replication assays for mapping initiation sites, will be applied to the examination of suspected replication origins, including those which may conceivably have a role in such phenomena as insertional mutagenesis and genomic imprinting.

Finally, the functional assay techniques and newly-identified origins will be applied in the construction of a human artificial chromosome. If it can be achieved, a human artificial chromosome would be a valuable tool for the investigation of chromosome biology, permitting origins to be arranged and analyzed in a large-scale defined context, allowing the cloning and study of the human centromere, and perhaps creating a model for epigenetic effects such as genetic imprinting and chromosomal inactivation. In addition, such a construct could serve as a human transfection vector with no practical size limit to the information it could carry, allowing assembly of complementation panels for gene mapping, and genetic transfer of complete transcriptional units without the need for any viral material. As a tool for gene therapy, a human artificial chromosome could see application in the treatment of inherited metabolic disorders (Kay and Woo, 1994), cancer (Culver and Blaese, 1994), or even some neurological disorders (Friedmann, 1994). Knowledge of replication origins can permit the first steps to be taken along this road.

REFERENCES

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

Bell D, Sabloff M, Zannis-Hadjopoulos M, Price GB (1991) Anti-cruciform DNA affinity purification of active mammalian origins of replication. *Biochim. Biophys. Acta* 1089:299-308

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

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

Culver KW, Blaese RM (1994) Gene therapy for cancer. *Trends Genet.* 10:174-178.

Dobbs DL, Shaiu WL, Benbow RM (1994) Modular sequence elements associated with origin regions in eukaryotic chromosomal DNA. *Nucl. Acids Res.* 22:2479-2489

Friedmann T (1994) Gene therapy for neurological disoders. *Trends Genet.* 10:210-214

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

Hamlin JL, Mosca PJ, Dijkwel PA, Lin HB (1993) Initiation of replication at a mammalian chromomal origin. *Cold Spring Harbor Symp. Quant. Biol.* LVIII:467-474

Kay MA, Woo SLC (1994) Gene therapy for metabolic disorders. *Trends Genet.* 10:253-257

Kelly R, DeRose M, Draper BW, Wahl GM (1995) Identification of an origin of bidirectional replication in the ubiquitously expressed mammalian *CAD* gene. *Mol. Cell. Biol.* 15:4136-4148

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

Krajewski WA, Razin SV (1992) Organization of specific DNA sequence elements in the region of the replication origin and matrix attachment site in the chicken α-globin gene domain. *Mol. Gen. Genet.* 235:381-388

Shihab-El-Deen A, Wu C, Hamerton JL, Zannis-Hadjopoulos M, Price GB (1993) Chromosomal localization of a sequence with in vivo activity for initiation of DNA replication. *Somat. Cell Mol. Genet.* 19:103-109

Taira T, Iguchi-Ariga SMM, Ariga H (1994) A novel DNA replication origin identified in the human heat shock protein 70 gene promoter. *Mol. Cell. Biol.* 14:6386-6397

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

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

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

Zannis-Hadjopoulos M, Pearson CE, Bell D, Mah D, McAlear M, Price GB (1992) Structural and functional characteristics of autonomously replicating mammalian origin-enriched sequences (ORS). In: *DNA Replication: The Regulatory Mechanisms*. Hughes P, Fanning E, Kchiyama M (eds) Berlin: Springer-Verlag, pp107-116

CHAPTER THREE

AUTONOMOUS REPLICATION IN VIVO AND IN VITRO OF CLONES SPANNING THE REGION OF THE *DHFR* ORIGIN OF BIDIRECTIONAL REPLICATION (*ORIβ*)

Maria Zannis-Hadjopoulos, Torsten O. Nielsen, Andrea Todd and Gerald B. Price

This chapter was published December 30, 1994, in the journal *Gene* (151:273-277). Experimental results include work using cell transfection autonomous replication assays, which were performed by Maria Zannis-Hadjopoulos and Andrea Todd, and *in vitro* replication assays, which I performed and wrote up for the first draft of the manuscript, otherwise put together by Maria Zannis-Hadjopoulos and Gerald B. Price. Journal reviewers requested several changes, additions and clarifications involving all portions of the manuscript, which I dealt with personally to produce the final revised form of the manuscript, presented here as a reprint of the journal article, with permission from the

Gene, 151 (1994) 273–277

GENE 08386

Autonomous replication in vivo and in vitro of clones spanning the region of the *DHFR* origin of bidirectional replication (*oriβ*)

(Hamster; episomal replication; in vitro replication; dihydrofolate reductase)

Maria Zannis-Hadjopoulos, Torsten O. Nielsen, Andrea Todd and Gerald B. Price

McGill Cancer Centre, McGill University, Montréal, Québec H3G 1Y6, Canada

Received by R. Padmanabhan: 11 April 1994; Revised/Accepted: 15 July/1 August 1994; Received at publishers: 29 August 1994

SUMMARY

Plasmids containing the origin of bidirectional replication (*oriβ*) of the Chinese hamster dihydrofolate reductase-encoding gene (*DHFR*) were tested for autonomous replication in vivo and in vitro. The results show that plasmids pX24 and p*neo*S13, that contain a 4.8- and a 11.5-kb fragment, respectively, spanning the *oriβ* region, are able to replicate autonomously in human cells and in a cell-free system that uses human cell extracts. Another plasmid, pX14, containing a 4.8-kb fragment that is immediately adjacent to the *oriβ* region, also replicated in these two assays.

INTRODUCTION

Several different methods have been employed for the isolation of mammalian DNA replication origins (*ori*) (reviewed in Vassilev and DePamphilis, 1992). However, a major difficulty in their identification as functional *ori*, has been the lack of a simple, reliable and sufficiently sensitive functional assay. One of the best candidates for such an assay is the *Dpn*I resistance assay (Peden et al., 1980), which has been very useful in the functional assessment of viral *ori* (Vassilev and DePamphilis, 1992). This assay has been successfully employed by us (Frappier and Zannis-Hadjopoulos, 1987; Landry and Zannis-Hadjopoulos, 1991; Bell et al., 1991; Nielsen et al., 1994) and others (Iguchi-Ariga et al., 1988; McWhinney and Leffak, 1990; Virta-Pearlman et al., 1993) to study the function of putative mammalian chromosomal *ori*. In addition, we recently developed an in vitro DNA replication system, which supports the specific initiation of one round of semiconservative replication in plasmids containing putative mammalian *ori* (Pearson et al., 1991; 1994). As we have shown, the *ori*-containing plasmids that are capable of autonomous replication in vivo, by transfection, can also replicate in the in vitro replication assay (Pearson et al., 1991; Nielsen et al., 1994). In this paper, we used the *Dpn*I resistance assay, in vivo and in vitro, to functionally assess a series of plasmids containing the bidirectional replication *oriβ* of the dihydrofolate reductase-encoding gene *DHFR* (Burhans et al., 1990).

EXPERIMENTAL AND DISCUSSION

(a) Transient replication assays in vivo by transfection

The autonomous replication of a series of plasmids containing sequences from the coding (pDG1a; 7.5-kb *Hin*dIII fragment of the *DHFR* coding region cloned in pUC19) and 3′ non-coding (pX14, pX24, p*neo*S13 region of the *DHFR* gene (gifts of Dr. M.L. DePamphilis, Roche Institute of Molecular Biology) was investigated (Fig. 1)

Correspondence to: Dr. M. Zannis-Hadjopoulos, McGill Cancer Centre, 3655 Drummond Street, Montréal, Québec H3G 1Y6, Canada. Tel. (1-514) 398-3536; Fax (1-514) 398-6769;
e-mail: hadjopoulos@medcor.mcgill.ca

Abbreviations: bp, base pair(s); DHFR, dihydrofolate reductase; *DHFR*, gene encoding DHFR; kb, kilobase(s) or 1000 bp; nt, nucleotide(s); *ori*, origin(s) of DNA replication; *oriβ*, bidirectional *ori* of hamster *DHFR* region; *ors*, *ori*-enriched sequences; 2D, two dimensional.

SSDI 0378-1119(94)00597-4

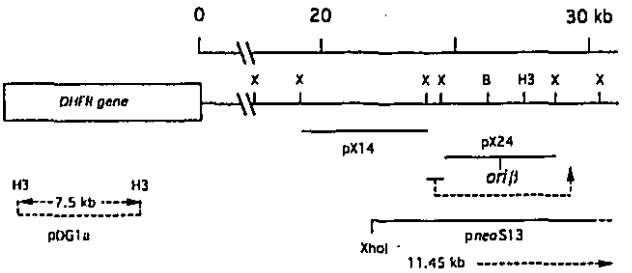


Fig. 1. Diagram of the *DHFR* gene and downstream noncoding regions indicating the fragments contained in plasmids pDG1a, pX14, pX24 and p*neo*S13. The position of the *oriβ*, as mapped by Burhans et al. (1990), is also indicated. B, *Bam*HI; H3, *Hin*dIII; X, *Xho*I.

Appropriate control plasmids of equivalent size containing bacteriophage λ DNA sequences (pDGλ8.6, 6.6+2.0-kb *Hin*dIII fragments of λ DNA; and pDGλ6.6, 6.6-kb *Hin*dIII fragment of λ DNA; gifts of Dr. M.L. DePamphilis) were also used. A monkey autonomously replicating *ors12* sequence (Zannis-Hadjopoulos et al., 1985; Frappier and Zannis-Hadjopoulos, 1987; Rao et al., 1990) was used as a positive control. DNAs (5 μg) from these plasmids were transfected in HeLa cells by the Ca·phosphate co-precipitation method (Graham and Van der Eb, 1973) and 48-h later Hirt supernatants (Hirt, 1967) were prepared. Plasmid DNA recovered in these supernatants was digested with *Dpn*I (Peden et al., 1980), which cleaves unreplicated (fully methylated) input DNA, but does not cleave DNA replicated (hemimethylated or unmethylated) in HeLa cells, and subjected to Southern blot analyses (Frappier and Zannis-Hadjopoulos, 1987; Landry and Zannis-Hadjopoulos, 1991; Wu et al., 1993a). Blots were probed with either nick-translated pBR322 (Fig. 2A) or a mixture of pDG1a and p*neo*S13 (Fig. 2B), which are plasmids containing vectors pUC19 and

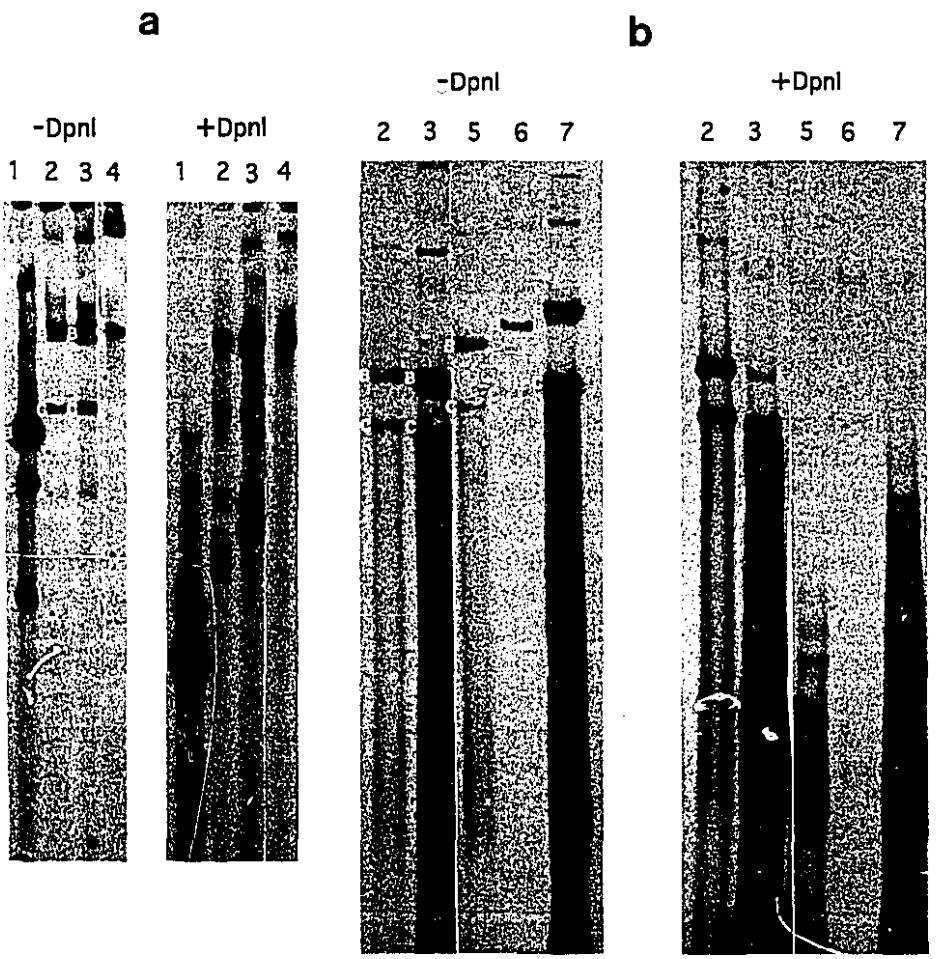


Fig. 2. Autonomous replication assay of *oriβ*-containing plasmids. HeLa cells were transfected with 5 μg of each plasmid DNA. After 48 h th recovered plasmid DNA samples were split into two halves, one half left untreated (−DpnI) and the other digested with 1.5 units of *Dpn*I for 2 (+DpnI), electrophoresed on 1% agarose, blot-transferred and probed with a mixture of nick-translated pDG1a and p*neo*S13 DNAs. Full digestio by *Dpn*I was verified by including 200 ng of λ DNA in each reaction, as described previously (Frappier and Zannis-Hadjopoulos, 1987). Lanes: *ors12*; 2, pX14; 3, pX24; 4, p*neo*S13; 5, pDGλ6.6; 6, pDG1a; 7, pDGλ8.6. Panels a and b represent two experiments, performed in the same manne excepting comb size and the plasmids tested. Supercoiled form I (A), relaxed circular form II (B) and linear form III (C) molecules are indicated i white. Bands migrating below form I represent *Dpn*I digestion products; bands migrating above form II include catenated circular products o replication and replicative intermediates.

pdMMTneo that carry the various DNA fragments used in the replication studies. As an internal control for full digestion by *Dpn*I, 500 ng of methylated λ DNA (NE Biolabs, Beverly, MA, USA) were included in all reactions, and the digestion products were verified by ethidium bromide staining and Southern blot hybridization (data not shown). The recovered plasmid DNAs were of the correct size expected for p*ors12* (5.4 kb), pX14 (7.5 kb), pX24 (7.5 kb), pDG1a (10.2 kb), p*neo*S13 (18 kb), pDGλ6.6 (9.3 kb) and pDGλ8.6 (11.3 kb), as in previous assays (Frappier and Zannis-Hadjopoulos, 1987; Landry and Zannis-Hadjopoulos, 1991; Wu et al., 1993b). The results (Fig. 2) show that *ors12*, pX14, pX24 and p*neo*S13 generated *Dpn*I-resistant bands, while pDG1a and the two control clones of bacteriophage λ DNA did not. The *ors12* element has been shown previously to have autonomously replicating activity (Frappier and Zannis-Hadjopoulos, 1987). Plasmids pX24 and p*neo*S13 both contain the *oriβ* region of *DHFR* (Burhans et al., 1990). Plasmid pX14 includes a region in which initiation was detected by 2D-gel electrophoresis, but at a lower frequency than the fragments which contain the *oriβ* (Dijkwel and Hamlin, 1992). The production of *Dpn*I-resistant supercoiled (form I) DNA is variable and generally lower than that of relaxed circular (form II) and linear (form III) forms, as has also been observed previously (Landry and Zannis-Hadjopoulos, 1991). This may be caused either by variability in topoisomerase activity (Pearson et al., 1991) or by the sensitivity of hemimethylated form-I DNA to nicking (Sanchez et al., 1992) that converts it to form II.

These results are consistent with previous observations that small (approx. 0.1–2 kb) *ors* (Zannis-Hadjopoulos et al., 1985; Rao et al., 1990; Landry and Zannis-Hadjopoulos, 1991) or other selected sequences (the 5′ c-*myc* origin: McWhinney and Leffak, 1990; cruciform-enriched human genomic fragments: Bell et al., 1991; Nielsen et al., 1994; a sequence 3′ to the mouse adenosine deaminase gene; Virta-Pearlman et al., 1993) of mammalian DNA can initiate autonomous replication in vivo (Frappier and Zannis-Hadjopoulos, 1987; McWhinney and Leffak, 1990; Landry and Zannis-Hadjopoulos, 1991; Bell et al., 1991, Virta-Pearlman et al., 1993; Nielsen et al., 1994) and in vitro (Pearson et al., 1991; 1994; Nielsen et al., 1994). In contrast, some studies have reported that only large fragments (> 10 kb) were observed to support autonomous DNA replication in human (Heinzel et al., 1991; Krysan et al., 1993) or rodent (Krysan and Calos, 1993) cells, including those that contain the *DHFR oriβ* (Caddle and Calos, 1992).

(b) In vitro replication assay

We have recently developed an in vitro replication system, in which monkey (*ors*) (Pearson et al., 1991) and human (Nielsen et al., 1994) autonomously replicating sequences can initiate and carry out one round of aphidicolin-sensitive semiconservative DNA replication (Pearson et al., 1991). We next assessed the ability of this system to detect autonomously replicating plasmids among those used above, containing sequences both from within the coding region or downstream from *DHFR*. In vitro replication was carried out as described in Pearson et al. (1991), using equimolar amounts (approx. 100 ng) of each plasmid DNA as template. Since pNeo.Myc-2.4 (9.3 kb) can be successfully replicated in our in vitro system, consistent with in vivo transfection data of McWhinney and Leffak (1988), we used it and *ors8* plasmid DNA (4.9 kb) (Pearson et al., 1991) as positive controls (Fig. 3). We observed that the same clones (pX14, pX24 and p*neo*S13) that replicated in vivo after transfection were also capable of replicating in vitro, generating *Dpn*I-resistant bands, while pDG1a and the two λ clones did not (Fig. 3). All the unreplicated input DNAs, before their use either in vivo or in vitro, were *Dpn*I-sensitive indicating that they were fully methylated (Peden et al., 1980; Nielsen et al., 1994). As previously, in all plasmids we observed material which migrates slower than form II, indicative of catenated dimers and replicative intermediates (Pearson et al., 1991). Control experiments in which template DNAs were not included in in vitro reactions yielded no products (Pearson et al., 1991). Thus, as before (Pearson et al., 1991; Nielsen et al., 1994), we found that the in vivo and in vitro assays are both capable of detecting autonomous replicating activity of plasmids containing putative mammalian *ori*.

Burhans et al. (1990), using various, different transfection protocols and a similar in vitro system with a variety of mammalian extracts, had difficulty observing significant *Dpn*I-resistant replication with either pX24 or pX14. Occasionally, however, under conditions that permitted borderline-detectable replication, both plasmids showed equivalent activity. Caddle and Calos (1992), transfecting 293S cells, observed autonomous replication of a 13.3-kb 3′ *DHFR* fragment (incorporating the inserts of both pX24 and pX14), but did not observe any activity from a shorter construct with the same 4.3-kb *Xba*I fragment present in pX24.

(c) Conclusions

(*1*) Autonomous replication activity can be detected in transient episomal replication assays of plasmids carrying the *DHFR oriβ*, regardless of their size (4.8-kb fragment, pX24; 11.5-kb fragment, p*neo*S13).

(*2*) The in vitro replication assay is an acceptable alternative method to the in vivo assay for detection of autonomous replicating activity of plasmids containing mammalian putative *ori*, corroborating our previous

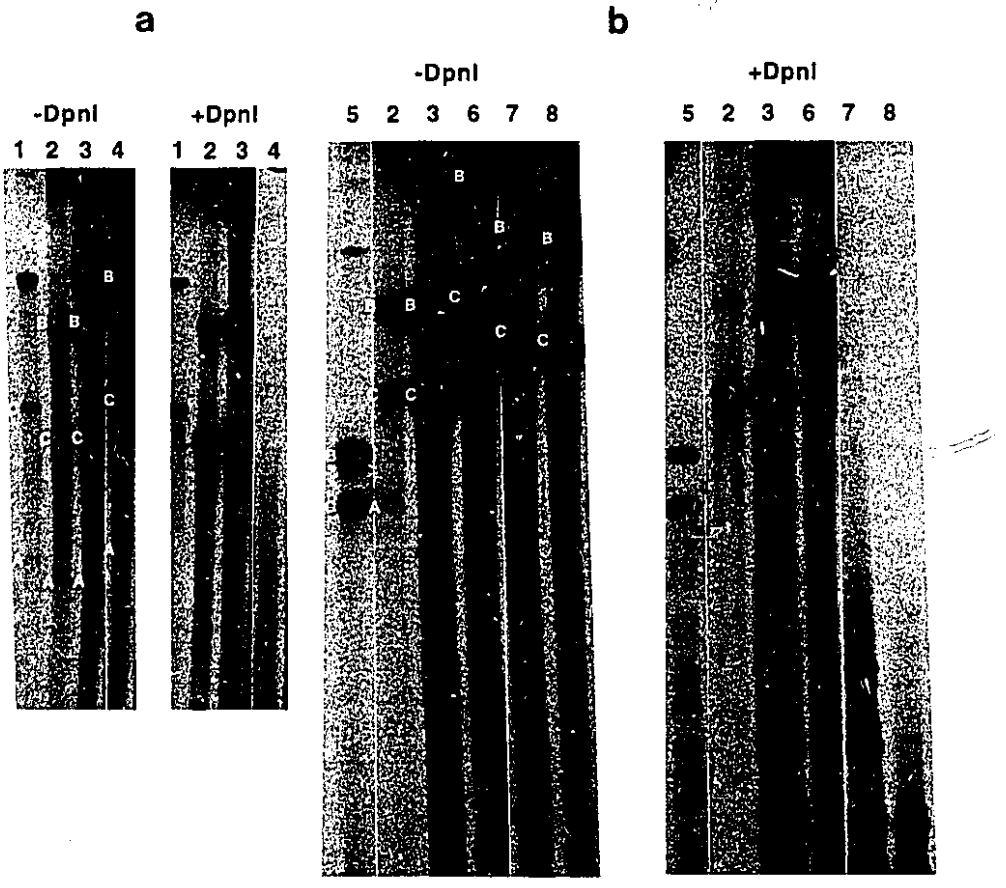


Fig. 3. In vitro replication assay. Equimolar amounts of each plasmid DNA were incubated for 1 h at 30°C in reaction mixtures containing HeLa cell extracts, as previously described (Pearson et al., 1991). The product DNAs were purified, concentrated, and divided into two halves; one half (approx. 50 ng) of each reaction was left untreated (– DpnI) and the other was digested with 1 unit of *Dpn*I for 1.5 h (+ DpnI). Lanes: 1, pNeo.Myc-2.4; 2, pX24; 3, pX14; 4, pDG1a; 5, *ors*8; 6, p*neo*S13; 7, pDGλ8.6; 8, pDGλ6.6. Panels a and b, as well as electrophoresis and the different replication products are as described in the legend to Fig. 2.

observations (Pearson et al., 1991; 1994; Nielsen et al., 1994).

(*3*) In both the in vivo and in vitro replication assays, control plasmids carrying λ DNA inserts are replication-negative, whereas the p*neo*S13 plasmid does replicate autonomously, as does the smaller plasmid, pX24 (discussed in Burhans et al., 1990); both of the latter two plasmids carry inserts that contain *oriβ*. A third plasmid, pX14, containing a 4.8-kb insert from the *DHFR* downstream region immediately 5′ to pX24, also displays autonomous replication; however, pDG1a, carrying 7.5 kb of the *DHFR* coding region, does not. The above results suggest that replication occurs only in plasmids containing mammalian inserts that carry an autonomously replicating sequence (a potential *ori*) and is not the result of random initiation from nicked, gapped or cut templates.

The observation that the inserts in pDG1a (7.5 kb) and pDGλ8.6 (8.6 kb) do not replicate in our assays, whereas those in pX14, pX24, pNeo.Myc-2.4 and *ors12* (4.8, 4.8, 2.4 and 0.8-kb inserts, respectively) do, implies that these results cannot be explained as simple fragment size dependent DNA replication (Heinzel et al., 1991). While replication from the *oriβ* contained within pX24 is consistent with results obtained by many mapping techniques, including nascent DNA PCR (Vassilev et al., 1990), and Okazaki fragment distribution (Burhans et al., 1990), the possibility of origin activity from pX14 is supported by several lower resolution mapping techniques (reviewed in Burhans et al., 1990), and by 2D-gel analysis (Vaughn et al., 1990).

ACKNOWLEDGEMENTS

We would like to thank Dr. M.L. DePamphilis (Roche Institute of Molecular Biology) for providing the plasmids pX14, pX24, pDG1a, p*neo*S13, pDGλ8.6 and pDGλ6.6. We thank Dr. P. Leder (Harvard Medical School) for providing the lambda phage clone containing the human *c-myc* gene and flanking DNA, and Dr. M. Leffak (Wright State University), who subsequently made

available a portion of this clone as pNeo.Myc-2.4. We also thank Drs. J.L. Hamlin (University of Virginia) and D. Gilbert (Roche Institute of Molecular Biology) for helpful advice. This work was funded by grants from the Medical Research Council (MRC) of Canada to M.Z.-H. (MT-7965) and the Cancer Research Society, Inc. to G.B.P. M.Z.-H. is an MRC Scientist, T.O.N. is a recipient of an MRC studentship, A.T. is a recipient of studentships from F.C.A.R. and Le Défi Corporatif Canderel.

REFERENCES

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

Bell, D., Sabloff, M., Zannis-Hadjopoulos, M. and Price, G.B.: Anticruciform DNA affinity purification of active mammalian origins of replication. Biochim. Biophys. Acta 1089 (1991) 299–308.

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

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 62 (1990) 955–965.

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

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 the plant amino acid mimosine. Mol. Cell. Biol. 12 (1992) 3715–3722.

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

Graham, F.L. and Van der Eb, A.J.: A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology 52 (1973) 456–467.

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 source of the DNA. Mol. Cell. Biol. 11 (1991) 2263–2271.

Hirt, B.: Selective extraction of polyoma DNA from infected mouse cultures. J. Mol. Biol. 26 (1967) 365–369.

Iguchi-Ariga, S.M.M., Ogawa, N. and Ariga, H.: Identification of the initiation region of DNA replication in the murine immunoglobulin heavy chain gene and possible function of the octamer motif as a putative DNA replication origin in mammalian cells. Biochim. Biophys. Acta 1172 (1993) 73–81.

Krysan, P.J. and Calos, M.P.: Epstein-Barr virus-based vectors that replicate in rodent cells. Gene 136 (1993) 137–143.

Krysan, P.J., Smith, J.G. and Calos, M.P.: Autonomous replication in human cells of multimers of specific human and bacterial DNA sequences. Mol. Cell. Biol. 13 (1993) 2688–2696.

Landry, S. and Zannis-Hadjopoulos, M.: Classes of autonomously replicating sequences are found among early-replicating monkey DNA. Biochim. Biophys. Acta 1088 (1991) 234–244.

McWhinney, C. and Leffak, M.: Episomal persistence of a plasmid containing human c-*myc* DNA. In: Stillman, B. and Kelly, T.J. (Eds.), Cancer Cells, Vol. 6, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1988, pp. 467–471.

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

Nielsen, T., Bell, D., Lamoureux, C., Zannis-Hadjopoulos, M. and Price, G.B.: A reproducible method for identification of human genomic DNA autonomously replicating sequences. Mol. Gen. Genet. 242 (1994) 280–288.

Pearson, C.E., Frappier, L. and Zannis-Hadjopoulos, M.: Plasmids bearing mammalian DNA-replication origin-enriched (*ors*) fragments initiate semiconservative replication in a cell-free system. Biochim. Biophys. Acta 1090 (1991) 156–166.

Pearson, C.E., Shihab-El-Deen, A., Price, G.B. and Zannis-Hadjopoulos, M.: Electron microscopic analysis of in vitro replication products of *ors8*, a mammalian origin enriched sequence. Somat. Cell Mol. Genet. 20 (1994) 147–152.

Peden, K.W.C., Pipas, J.M., Pearson-White, S. and Nathans, D.: Isolation of mutants of an animal virus in bacteria. Science 209 (1980) 1392–1396.

Rao, S.B., Zannis-Hadjopoulos, M., Price, G.B. and Martin, R.G.: Sequence similarities among monkey *ori*-enriched (*ors*) fragments. Gene 87 (1990) 233–242.

Sanchez, J.A., Marek, D. and Wangh, L.J.: The efficiency and timing of plasmid DNA replication in *Xenopus* eggs: correlations to the extent of prior chromatin assembly. J. Cell Sci. 103 (1992) 907–918.

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

Vassilev, L.T. and DePamphilis, M.L.: Guide to identification of origins of DNA replication in eukaryotic cell chromosomes. Crit. Rev. Biochem. Mol. Biol. 27 (1992) 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 61 (1990) 1075–1087.

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

Wu, C., Friedlander, P., Lamoureux, C., Zannis-Hadjopoulos, M. and Price, G.B.: cDNA clones contain autonomous replication activity. Biochim. Biophys. Acta 1174 (1993a) 241–257.

Wu, C., Zannis-Hadjopoulos, M. and Price, G.B.: In vivo activity for initiation of DNA replication resides in a transcribed region of the human genome. Biochim. Biophys. Acta 1174 (1993b) 258–266.

Zannis-Hadjopoulos, M., Kaufmann, G., Wang, W.-S., Lechner, R.L., Karawya, E., Hesse, J. and Martin, R.G.: Properties of some monkey DNA sequences obtained by a procedure that enriches for DNA replication origins. Mol. Cell. Biol. 5 (1985) 1621–1629.

CHAPTER FOUR

A REPRODUCIBLE METHOD FOR IDENTIFICATION OF HUMAN GENOMIC AUTONOMOUSLY REPLICATING SEQUENCES

Torsten Nielsen, David Bell, Claude Lamoureux, Maria Zannis-Hadjopoulos and Gerald Price

Chapter 3 demonstrated that autonomous replication assays have sufficient sensitivity to detect origin activity in DNA fragments derived from a mammalian chromosomal initiation site. In Chapter 4, I adapt these techniques to allow mass screening of human DNA libraries that are enriched for sequences which form cruciforms *in vivo*, in an effort to isolate large numbers of novel replication origins.

This chapter was published by *Molecular and General Genetics* in February, 1994 (242:280-288). I was responsible for all of the experimental work, with the exception of the construction of the initial anti-cruciform immunoaffinity purified λ phage libraries (performed by David Bell), and I was aided by Claude Lamoureux in the construction of pBluescript secondary libraries. Gerald Price wrote the Introduction and Discussion sections of the paper, but I wrote the remainder of the first draft and made all revisions for the published version. Maria Zannis-Hadjopoulos and Gerald Price provided supervision in the initial planning of experiments, and edited all drafts of the manuscript.

The chapter is presented as a reprint of the published article (with permission from the copyright owner, Springer-Verlag) but I have included an addendum which elaborates on the points made by Table 3, Figure 5 and the text at the end of the results section, relating the transfection and *in vitro* assays in mass screening and individual clone experiments.

Mol Gen Genet (1994) 242:280–288

MGG

A reproducible method for identification of human genomic DNA autonomously replicating sequences

Torsten Nielsen, David Bell, Claude Lamoureux, Maria Zannis-Hadjopoulos, Gerald Price

McGill Cancer Centre, McGill University, Montreal, Quebec, Canada H3G 1Y6

Received: 24 May 1993 / Accepted: 27 July 1993

Abstract. We demonstrate a method for the isolation of autonomously replicating sequences from pools of clones obtained from genomic DNA libraries constructed using affinity purification of cruciform DNA. The selection of autonomously replicating sequences was based on their differential ability to replicate as episomes after transfection of pools of plasmid clones into human HeLa cells. Two separate libraries containing affinity-purified cruciform DNA were used, one prepared from DNA of log phase primary human genital fibroblasts and the other prepared from DNA of log phase SW48 colon adenocarcinoma cells. Representative samples of the entire phage libraries were converted to phagemid clones by filamentous helper phage-mediated mass excision to produce pBluescript libraries in *Escherichia coli*. Clones were grown up individually and the bacteria pooled into groups of 48 for recovery of plasmid DNA. Plasmid pools of 48 independent clones (120 μg total) were then transfected by calcium phosphate coprecipitation onto log phase HeLa cells, which were allowed to grow for 3 days before recovery of plasmid by Hirt lysis. The recovery of plasmid from each transfection was estimated to range from 10 to 60 ng. *Dpn*I digestion was then used to digest plasmids which had not been replicated and therefore retained a bacterial methylation pattern which was sensitive to digestion. We estimated from agarose electrophesis gels that 40–200 pg of recovered plasmid DNA per transfected pool of DNA was resistant to *Dpn*I and therefore was capable of transforming competent *E. coli* cells. The *Dpn*I-resistant fraction yielded from one to seven independent clones from each pool, with genomic DNA inserts ranging in size from 0.35 to 3.4 kb. The fidelity of the procedure was demonstrated by performing duplicate transfections from the same pool of plasmid DNA, and identifying bands which were apparently common between duplicate transfections by size and sequence analysis. A second method of mass screening, using an in vitro DNA replication assay instead of transfections, resulted in similar yields and led to the isolation of an overlapping subset of selected clones.

Key words: Origins – DNA replication – Human – In vitro replication assay – Library screening

Communicated by G.P. Georgiev

Correspondence to: G. Price

Introduction

Inverted repeat DNA (palindromic) sequences are widely distributed in chromosomal DNA of eukaryotes (Wilson and Thomas 1973, 1974; Schmid et al. 1975). They have the potential to form intrastrand base pairs, which, in a negatively supercoiled molecule, can lead to the formation of cruciform structures. It has been suggested that such structures may form under physiological conditions to serve as recognition signals for specific regulatory proteins of DNA replication. Inverted repeat sequences have been found to be a regular feature of mammalian autonomously replicating sequences (Rao et al. 1990; Landry and Zannis-Hadjopoulos 1991). In an effort to facilitate the detection of cruciforms and evaluate their role in DNA replication, we recently produced monoclonal antibodies with unique specificity for cruciform DNA molecules (Frappier et al. 1987). These antibodies recognize conformation determinants specific to DNA cruciforms and do not bind linear double-stranded DNA, linear single-stranded DNA, or single-stranded DNA containing a hairpin; the conformation determinant has been shown to be associated with the secondary structure characteristic of the base, and the junction of the cruciform structure (Frappier et al. 1989). When this anti-cruciform antibody was used in permeabilized cells, we observed a 2- to 11-fold enhancement of DNA replication and enhancement in copy number of low copy genetic elements (Zannis-Hadjopoulos et al. 1988). Recently, we have successfully used anti-cruciform monoclonal antibody and negatively supercoiled genomic DNA of monkey CV-1 cells to make an affinity-purified library of sequences for analysis in assays of autonomous

replication in human cells (Bell et al. 1991). We found that the library was enriched for sequences capable of autonomous replication after transfection into HeLa cells.

We have made additional libraries of human genomic sequences from negatively supercoiled DNA of primary genital fibroblasts and colon adenocarcinoma cells. The individual isolation of random clones, subcloning, and individual assessment of autonomous replication activity cannot accommodate or use, to appropriate advantage, such a large number of putatively autonomously replicating sequences. In this study, we report a method for assessment of large numbers of clones simultaneously. The loss in sensitivity of detection due to the more stringent conditions that are necessitated by these types of mass transfections is counterbalanced by the ability to screen large numbers of independent recombinants from the libraries.

Materials and methods

Construction of lambda bacteriophage libraries. Lambda ZAPII origin-enriched human genomic libraries were produced by the anti-cruciform DNA affinity purification method of Bell et al. (1991). Briefly, high molecular weight DNA was extracted from log phase primary human genital fibroblasts and from SW48 (ATCC CCL231) human colon adenocarcinoma cells, incubated with anti-cruciform monoclonal antibody 2D3 (Frappier et al. 1989) specific for DNA cruciform structures, and exhaustively digested with *Eco*RI (BRL). The DNA was then specifically recovered with anti-mouse immunoglobulin immunobeads (BioRad). For each source, DNA from 12 such reactions was pooled, ligated to *Eco*RI-digested Lambda ZAPII vector DNA (Stratagene, La Jolla, Calif.) and packaged into phage particles as detailed by Bell et al. (1991). The phage were then used to infect XL-1 Blue bacteria at a low multiplicity, and the bacteria plated in the presence of X-gal (Pharmacia) and IPTG (Pharmacia) to determine the percentage of recombinants in the library. After determining the titer of the packaging reaction mixture, the library was amplified. Hereafter, the libraries made with *Eco*RI-digested, 2D3 anti-cruciform affinity-purified DNA from primary genital fibroblasts and SW48 colon carcinoma cells will be called PHF and SW48II libraries, respectively.

Subcloning and pBluescript libraries. Mass excision of phage to obtain phagemid clones was facilitated by the presence in the Lambda ZAPII vector of the f1 filamentous helper phage origin of replication and termination/packaging signals. Coinfection of host bacteria by both library lambda phage and filamentous helper phage results in the production of single-stranded pBluescript clones packaged as defective filamentous phage. For the PHF library, 2×10^5, and for the SW48II library, 4×10^5 lambda phage were mixed with a 15:1 excess of filamentous helper phage R408 and 200 µl log phage host XL1-Blue *Escherichia coli;* they were incubated together for 15 min at 37° C to allow infection. Bacteria were then shaken for 3 h at 37° C in 5 ml 2 × YT media to allow replication and extrusion of filamentous phage; any remaining host cells and lambda phage were killed by heating to 70° C for 20 min, and removed by centrifugation. The supernatant (stored with 10 µl chloroform) contains a mixture of defective pBluescript clone-containing particles and intact helper phage. pBluescript phagemid titers and optimal reinfection ratios were determined by induction of ampicillin resistance in fresh XL1-Blue cells. pBluescript double-strand plasmid libraries were then produced by infecting excess XL1-Blue bacteria with suitably diluted mass excision supernatant and plating on LB agar containing 80 µg/ml ampicillin, 80 µg/ml X-gal, and 0.5 mM IPTG. The plates were incubated overnight at 42° C to further limit the chance of coinfection by the temperature-sensitive intact helper phage. Individual recombinant colonies were picked and transferred to 96-well plates containing LB/ampicillin media and grown overnight at 42° C; after addition of glycerol to final concentration of 15%, the plates were stored at −80° C.

Production of plasmid DNA pools. The bacterial colonies were reconfirmed to contain recombinant plasmid by growth on LB/ampicillin/X-gal/IPTG agar plates. Aliquots of 20 µl from bacteria containing individual recombinant plasmid clones were then singly removed from 96-well plates and used to inoculate 15 ml cultures of LB/ampicillin; the 48 individual cultures were then pooled and the recombinant plasmid isolated. Alkaline lysis and large-scale plasmid preparation were performed according to standard methods (Maniatis et al. 1982); the plasmid DNA was purified using two consecutive CsCl gradient ultracentrifugations.

Functional purification of human genomic clones by transfection assay. HeLa cells (human cervical carcinoma, ATCC CCL2) recently recovered from frozen stocks were placed in T-175 flasks (Nunclon) at a density of $1 \times 10^4/cm^2$, and incubated overnight (37° C, 5% CO_2) in 50 ml α-minimal essential medium (α-MEM; Gibco) supplemented with 10% fetal calf serum (Flow Laboratories), 50 U/ml penicillin (Gibco), 50 µg/ml streptomycin (Gibco), 292 mg/l L-glutamine (Sigma), and 44 mg/l L-asparagine (Sigma). Immediately prior to transfection, the medium was removed and replaced with 25 ml fresh medium; calcium phosphate coprecipitates (Graham and van der Eb 1973; Bell et al. 1991) of 120 µg aliquots of the purified plasmid DNA pools were added to the cells, and the cells were then incubated at 37° C for 20 h. Subsequently, the cells were rinsed twice with serum-free α-MEM before addition of 50 ml of fresh growth medium. The transfected HeLa cells were allowed to grow 72 h and the low molecular weight DNA was isolated by the method of Hirt (1967) as previously described (Bell et al. 1991; Landry et al. 1991). Plasmids that had undergone replication in transfected HeLa cells were selected on the basis of differential methylation and resistance to *Dpn*I digestion. A 75 µl aliquot of the Hirt supernatant containing low molecular weight DNA was digested for 2 h with 8 U *Dpn*I; the digested plasmid was

immediately used to transform competent *E. coli* DH5α cells. Aliquots of 25 µl of *Dpn*I-digested plasmid were used to transform 100 µl competent cells. As controls for the bacterial transformation reactions, we used undigested and *Dpn*I-digested pBluescript vector DNA. Aliquots were then plated on LB agar containing 80 µg/ml ampicillin to determine transformation efficiency. The rest of the transformed bacteria was used to inoculate an overnight culture of 50 ml LB broth containing 80 µg/ml ampicillin. After culture, LB/ampicillin agar plates were streaked with the broth culture and individual colonies were picked for screening of clones and characterization of insert size after digestion with *Eco*RI.

Analysis of functionally purified clones. Samples representing 5% of the Hirt supernatant, before and after *Dpn*I treatment, were analyzed by Southern blot (Southern 1975). DNA was separated by electrophoresis in an 0.8% agarose gel in 1 × TBE buffer (Maniatis et al. 1982), and blotted onto 0.45 µm Nytran membranes (Schleicher and Schuell). Blots were prehybridized according to the manufacturer's specifications and then hybridized at 65° C with a pBluescript probe, prepared with a nick translation kit (Amersham) using α-[^{32}P]dCTP (3000 Ci/mmol, Amersham). The blots were washed extensively and then an autoradiograph was prepared with Kodak XAR-5 X-ray film exposed at −70° C.

Demonstration of the reproducibility of the procedure. The fidelity of the procedure, the capacity to reproducibly isolate the same clones from the same input DNA, was tested by performing duplicate transfections of HeLa cells, and treating these identically thereafter. If matching *Eco*RI banding patterns were present in clones isolated from duplicate transfections, homology of the clones was proven by thymidine-tracking (T-tracking). Both ends of each clone were tracked and suspected homologs were compared. Sequencing and T-tracking were performed on plasmid DNA using a T7 sequencing kit (Pharmacia), either T3 or T7 primers and α-[^{35}S]dATP according to the manufacturer's specifications.

Functional purification of human genomic clones by in vitro replication. Aliquots of 250 ng of the same plasmid DNA pools used for transfection assays were replicated in vitro using the system of Pearson et al. (1991), in a reaction mix containing HeLa nuclear extract, HeLa low salt cytoplasmic extract, an ATP regenerating system, polyethylene glycol, and non-radiolabelled dNTPs. The product was digested 2 h with 5 U *Dpn*I, bacterial transformation and subsequent isolation of unique clones then proceeded as described above. T-tracking was employed to compare isolated clones with suspected homologs previously isolated using the transfection assay.

Assay of in vitro replication capacity of individual clones. Selected clones were grown in *E. coli* and the plasmid DNA isolated by alkaline lysis and two CsCl ultracentrifugations. Samples (50 ng) were assayed using the same in vitro replication system described above, this time including 10 µCi each of α-[^{32}P]dCTP and α-[^{32}P]dTTP. To assess replication, duplicate reactions were performed in the presence of 30 µM aphidicolin (Boehringer-Mannheim), or alternatively, the reaction products were digested with *Dpn*I (1 U, 1.5 h). The purified product was electrophoresed 16 h in a 1% 1 × TAE (Maniatis et al. 1982) gel at 2.5 V/cm. The gel was dried down and used to expose X-ray film directly.

Assay of replication capacity of individual clones by transfection assay. Samples (5 µg) of CsCl-purified DNA from single clones were separately transfected into HeLa cells in T-25 flasks (Nunclon) using a scaled-down version of the mass transfection assay described above. One-quarter of the product was set aside, and the remainder digested with 1 U *Dpn*I for 1 h at 37° C. An equivalent amount of digested DNA was electrophoresed alongside the uncut aliquots, Southern blotted and probed with nick-translated pBluescript vector. The rest, half of the original lysate, was used to transform competent DH5α cells, which were immediately plated on LB/ampicillin to provide a colony count assay of replication efficiency.

Results

Anti-cruciform DNA affinity purification produced cruciform-enriched primary human fibroblast (PHF) and adenocarcinoma (SW48II) lambda ZAPII libraries containing 1.87×10^4 and 1.40×10^4 independent recombinants, respectively. Based on bromodeoxyuridine repli-

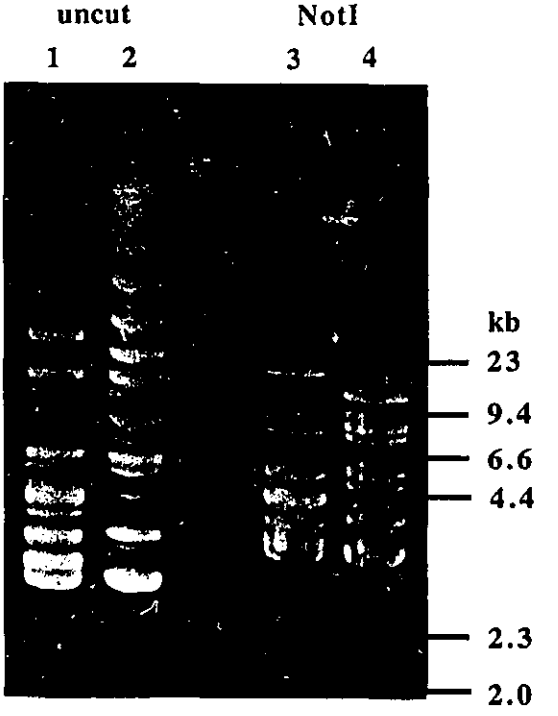


Fig. 1. Distribution of clones in plasmid DNA pools. Aliquots c 2 µg of two different pools of 48 individual clones were electrc phoresed in 0.8% agarose gels and stained with ethidium bromide Lanes 1 and 2, undigested; lanes 3 and 4, linearized by *Not*I diges tion

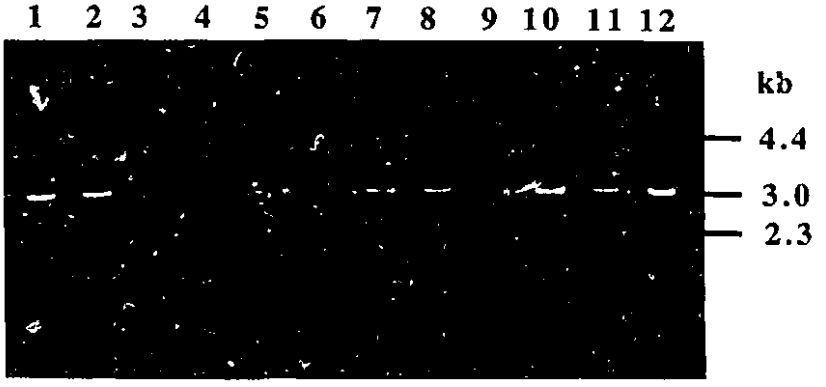


Fig. 2. Plasmid yield from individual pBluescript library clones. *Eco*R1-digested plasmid from 12 individual 3 ml cultures was electrophoresed in 0.8% agarose and stained with ethidium bromide (lanes 1–12). Scanning densitometry using the Bio Image densitometer compared the intensities of the 3.0 kbp vector DNA band to a standard (not shown), except for the non-cutting lane 9, for which the form I and II band intensities were summed (see Results)

cation assays of a CV-1 (monkey kidney) library produced by the same procedure (Bell et al. 1991), nearly one-half of these clones are likely to possess autonomous replication activity.

Before attempting a functional purification of these affinity-purified human genomic DNA clones, it was necessary to subclone the DNA into a bacterial vector. We converted phage to phagemid (plasmid) by using filamentous helper phage-mediated mass excision from the lambda ZAPII vector. Burst ratios were of the order of 20000 defective filamentous pBluescript phagemid particles for each lambda library clone. Reinfection of XL1-Blue bacteria completed the conversion; individual bacterial colonies representing a total of 3072 PHF and 3552 SW48II plasmid clones were picked and stored as pBluescript source libraries for further analysis.

Pools of individual cultures of bacterial clones yielded 0.4–1.1 mg plasmid DNA. Figure 1 shows the distribution of clones within two of the pools of plasmid DNA that were used for transfections. After digesting the pool of plasmid DNA with *Not*I to linearize plasmid DNA clones, we used quantitative densitometry to resolve 13 and 15 distinct bands, respectively; a comparison of the relative density of some of the bands suggested that multiple clones had comigrated. However, it was clear that fewer than 48 clones had been successfully propagated. In an effort to determine the proportion of clones which did grow sufficiently to contribute to the final pool, individual clones were tested for recovery of plasmid DNA from 3 ml of bacterial cultures according to standard procedures (Maniatis et al. 1982). The relative proportions of DNA obtained from individual cultures are shown in Fig. 2. After quantitative densitometry, we determined that 9 of 12 clones yielded similar amounts of plasmid, averaging 440 ± 80 ng (95% CI, confidence interval) per 1.5 ml culture, while three yielded significantly less, averaging 50 ± 30 ng (95% CI) per 1.5 ml culture, and would therefore be inadequately represented in the transfection input pools. This result was confirmed, and the mean size of inserts in the subcloned anti-cruciform libraries was calculated to be 1.4 kb ± 0.4 (95% CI), in close agreement with the size range observed in previous anti-cruciform library constructions (Bell et al. 1991).

Thus, the 120 µg DNA pools used for calcium phosphate coprecipitation transfections of HeLa cells represent an estimated 36 clones at approximately 3 µg plasmid DNA per clone. After removal of the transfection medium, HeLa cells remained in log phase throughout the subsequent 72 h, with an estimated doubling time of 15 h. After lysis and isolation of low molecular weight DNA (Hirt 1967), *Dpn*I was used to digest non-replicated plasmid; after digestion, the plasmids were used to transform competent DH5α bacteria.

A Southern blot of the Hirt low molecular weight DNA fraction, before and after *Dpn*I digestion, is shown in Fig. 3. The majority of the recovered DNA is *Dpn*I-sensitive, but with longer exposures, several undigested bands can be seen. Some bands which were present be-

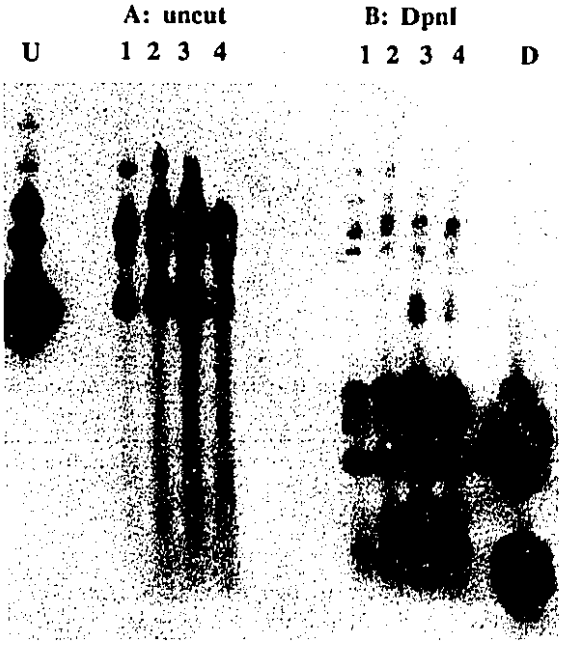


Fig. 3A, B. Southern blot showing plasmid DNA recovered from transfected HeLa cells (probe: pBluescript). Separation by agarose-gel electrophoresis of plasmid DNA recovered from HeLa cells after transfection with pools of plasmids. A Undigested plasmid DNA. B Plasmid DNA after digestion with *Dpn*I. Duplicate transfections of a pool of genomic clones obtained from DNA of the SW48II human colon adenocarcinoma (lanes 1, 2) and of a pool of genomic clones obtained from DNA of normal human genital fibroblast (lanes 3, 4). Lane U, 50 ng undigested pBluescript DNA. Lane D, 50 ng pBluescript DNA digested under the same conditions as the samples in B

Table 1. Summary of autonomous replication assays on pools of anti-cruciform antibody affinity-purified human genomic DNA clones

Cruciform-enriched lambda library: source of cell type	PHF Normal human genital fibroblasts	SW48II Colorectal adeno-carcinoma cell line
Approximate total number of clones transfected[a]	108	180
Number of transfections	6	8
Dpn I-resistant DNA recovered, as determined by transformation efficiency	40–200 pg per 120 µg transfection	60–110 pg per 120 µg transfection
Independent isolated clones[b]	18 (1–7/transfection)	23 (1–7/transfection)

[a] We determined that approximately 75% of clones in transfection pools grew sufficiently well to be fairly represented in the pool and in an assay of their autonomous replication potential (see also Results and Fig. 2)
[b] Where duplicate transfections were performed to show fidelity of the procedure, duplicated homologous clones are not included in the final total number of independent clones recovered. However, they were included in the analysis of clones recovered per transfection

Table 2. Fidelity of recovery of autonomously replicating human genomic clones in duplicate transfections

Source of DNA	Number of clones isolated[a]	Homologous pairs[b]	Unique clones[c]
SW48 fidelity pool 1	4, 3	2	3
SW48 fidelity pool 2	7, 7	2	10
PHF fidelity pool 1	3, 1	1	2
PHF fidelity pool 2	5, 6, 7	4	10

[a] The number of separate clones, as identified by insert size, which were recovered from duplicate transfections of a common pool of recombinant plasmids
[b] The number of identical clones, as determined by T-tracking sequence analysis, which were isolated from duplicate transfections
[c] The number of clones recovered from only one of the duplicate transfections

fore *Dpn*I treatment are no longer visible after *Dpn*I digestion. These results also indicate the reproducibility of the method of isolation of autonomously replicating sequences, since lanes from duplicate transfections (1 and 2 are separate assays from the same SW48II pool; 3 and 4 are separate assays from the same PHF pool) gave very similar banding patterns before and after digestion.

The amounts of plasmid DNA recovered from transfected HeLa cells after digestion with *Dpn*I were estimated at 40–200 pg of DNA. The presence of plasmid DNA forms I, II, and III, and the possibility that overlapping clones occur made it necessary to isolate single

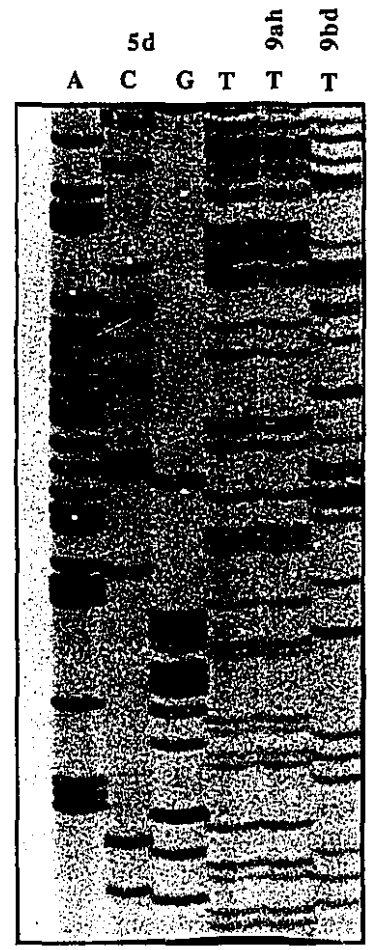


Fig. 4. Confirmation of recovery of homologous clones from dupli cate transfections by T-track sequence analysis. Clone 5d, recovere after transfection of a pool of DNA clones from the PHF librar is shown to be identical to clone 9ah and different from clone 9b both recovered from one of two duplicate transfection assays

transformants in order to determine the number an identity of those clones which had replicated after trans fection into HeLa cells. The number of clones that w needed to analyze, in order to obtain examples of a replicated clones and measure the specific insert size, wa determined using the Poisson distribution and the equa tion ln(0.05)/ln((x-1)/x), where x = estimated maximur number of *Dpn*I-resistant clones observed by Souther blot. Thus, there was greater than 95% certainty of sam pling all independent clones surviving the abov procedure. The mean number of unique clones from eac pool of independent clones, which were replicated i HeLa cells, was 3.6 ± 1.3 (95% CI, range 1–7). Combine results of all transfections are shown in Table 1.

After recovery of plasmids with the ability to replicat autonomously in HeLa cells, we determined the fidelit of this method, i.e. the ability to isolate reproducibl identical clones from the same pool of independe recombinant plasmids. To test the fidelity of ou procedure, duplicate transfections were run under ident cal conditions. Very similar banding patterns were ol served after gel electrophoresis of either undigested c

Table 3. Contingency comparisons of autonomously replicating DNA assayed by transfection and in vitro replication

	Contingency tables				Hypothesis
A	Replicated using transfection assay				H_0: The chance of a clone replicating in the in vitro assay is independent and unrelated to its chance of replicating in the transfection assays[a] $P = 2.5 \times 10^{-4}$[b]
		+	−	Total	
Replicated using in vitro replication assay	+	11	7	18	
	−	23	103	126	
	Total	34	110	144	
B	Replication in one in vitro replication assay				H_0: The clones identified by transfection are not the same as those identified by the in vitro replication assay $P = 0.25$
		+	−	Total	
Replication in a single transfection	+	7	10	17	
	−	11	14	25	
	Total	18	24	42	

[a] Null hypothesis
[b] P, probability determined by Fisher's exact test

*Dpn*I-digested plasmid which was recovered from duplicate cultures of transfected HeLa cells (Fig. 3). We used sequence analysis (T-tracking) to confirm with confidence the actual number of recombinant clones which could be reproducibly recovered from HeLa cells (Fig. 4). These results are summarized in Table 2. Some clones were reproducibly recovered from duplicate transfections; however, not all clones recovered from one of two duplicate transfections were recovered in the other. The probability that 9 clones of affinity-purified genomic DNA could be reproducibly obtained from duplicate cultures by random chance, when 34 clones with autonomous replicating potential were detected and none of 12 random genomic clones have autonomous replicating activity (Bell et al. 1991), is $P = 0.048$ (Fisher's exact test). Thus, mass transfections, like those described here, may be judiciously used to screen large numbers of independent clones to recover autonomously replicating human genomic DNA clones from libraries of anti-cruciform affinity-purified DNA.

Next, we undertook to test the ability of an in vitro DNA replication assay system (Pearson et al. 1991) to detect autonomously replicating sequences among pools of plasmid DNA clones. We performed four separate assays of 250 ng DNA/assay from two pools of plasmids (approximately 36 clones/pool), each from PHF and SW48II. From two to seven independent *Dpn*I-resistant subclone transformants were recovered from each (mean 4.5 ± 2, 95% CI), with inserts of size range 0.2–2.6 kb. The number of clones recovered as having autonomous replication activity was very similar to that obtained from the transfection-based assays. We estimated from the transformation efficiency for each pool of plasmid DNA, tested in vitro after digestion with *Dpn*I, that we recovered from 20 to 170 pg *Dpn*I-resistant DNA. This represents up to 500 times greater yield of in vitro replicated DNA than could be obtained from the in vivo transfection assays for autonomous DNA replication. We observed that of the 34 clones replicating after in vivo transfection of these pools, 11 were also recovered as replicated in the in vitro assay. Of the estimated 110 clones which failed to be replicated by in vivo transfection, 7 were recovered by the in vitro replication assay. The probability that this classification of clones by autonomous replication potential as assayed by transfection and by replication in vitro was random is $P = 2.5 \times 10^{-4}$ (Fisher's exact test, see Table 3A). Thus, there is a major component of similarity in these two assays for autonomous replicating activity. However, when we examine only those clones of pools which were detectable as having been replicated by a single in vivo transfection assay, and then assayed those clones in vitro, we could not conclude that there was an exact equivalency of the two assays (Table 3B: $P = 0.25$ for the null hypothesis, i.e. considering replicating clones only, the subset of clones which replicated in vitro is not necessarily the same as that which replicated on in vivo transfection). Overall, these results suggest that the same subsets of cruciform-purified clones are not identically recoverable in both systems, and therefore those clones which have replication potential may be activated with different efficiencies by the two different replication assay systems.

Figure 5 demonstrates that the in vitro replication of representative individual clones derived from pools previously screened is dependent upon α/δ DNA polymerases. Clones S14 and F6 were previously identified only by transfection, whereas F2 was identified only by in vitro screening, and F4 was identified by both systems. Treatment with aphidicolin inhibited incorporation of radionucleotides significantly, i.e. incorporation was DNA polymerase α/δ dependent. Clone F4 showed the greatest radionucleotide incorporation, implying that those clones isolated from both the in vitro and the transfection systems may possess the strongest individual replication capacity. However, the other three clones, whether isolated by transfection or in vitro methodology, also had replication potential. Further tests showed that individual clones F1, F11, F12, F15, F20, and S15 were

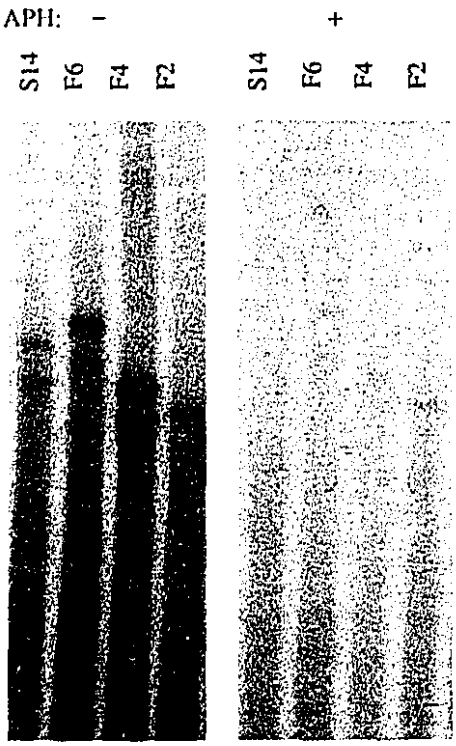


Fig. 5. In vitro replication assay of individual autonomously replicating clones. Plasmid clones S14 and F6, isolated by transfection assay but not in vitro replication assay, clone F4, isolated by both assay systems, and clone F2, isolated by in vitro replication but not by transfection assay, were replicated in the presence of radiolabelled dCTP and dTTP using the method of Pearson et al. (1991)

also able to replicate in vitro, yielding a product resistant to *Dpn*I under conditions which completely digested a control plasmid (pBluescript containing a 0.7 kb human cDNA insert). In summary, ten of ten anti-cruciform mass screening isolates tested had demonstrable autonomous replication activity when individually subjected to the in vitro assay.

Autonomous replication activity of isolates was also confirmed by individual transfection assays. Seven of seven clones (F4, F5, F15, F20, S3, S9, S14) tested demonstrated strong activity, showing *Dpn*I-resistant bands on Southern blot analysis, and yielding an average of 30 times as many *Dpn*I-resistant transformants as control plasmid. Furthermore, we observed that those clones isolated by both the in vitro replication and in vivo mass transfection methodologies gave the greatest number of bacterial transformants, apparently indicative of greater efficiency in "activation" of eukaryotic DNA replication.

Thus, while the anti-cruciform affinity-purified libraries are highly enriched for sequences demonstrating autonomous replication capacity in an individual assay (estimated at about 50% in Bell et al. 1991), these mass screening procedures described here employ a sufficiently stringent selection to yield only about 10% of clones with apparent autonomous replicating activity; of these, the most active subset are isolated with consistency by both in vitro and transfection-based assays.

Table 4. Summary of clones with autonomous replication activity

Clone name	Size (kbp)	Method of identification[a]	
		In vivo	In vitro
F1	0.11	+	−
F2	0.25	−	+
F3	0.25	−	+
F4	0.30	+	+
F5	0.35	+	+
F6	0.80	+	−
F7	2.1	++[c]	−
F8	2.1	+	−
F9	0.50	+	−
F10	2.2	+	−
F11	0.40	−	+
F12	0.12	++	−
F13	0.65	+	−
F14	0.70	+	−
F15	0.69	++	+
F16	1.2	+	−
F17	2.0	−	+
F18	2.2	++	−
F19	2.7	+	−
F20	0.30	++	+
F21	1.5	−	+
F22	2.6	+	−
F23	2.6	−	+
F24	2.0	−	+
F25[b]		+	+
S1	0.30	+	−
S2	0.90	+	−
S3	1.1	++	+
S4	1.15	+	−
S5	2.1	+	−
S6	2.4	+	+
S7	3.4	+	−
S8	0.10	+	−
S9	0.25	++	−
S10	1.1	+	+
S11	1.4	+	+
S12	2.0	+	−
S13[b]		+	−
S14	0.50	+	−
S15	0.30	+	+
S16	1.2	+	+
S17	1.3	++	−

[a] Two methods of identification of the autonomously replicatin sequences were used, in vivo transfection into somatic cells and th in vitro replication assay system

[b] Although the pBluescript clone possessed a recombinant clon banding pattern, disruption of the *Eco*RI resistriction sites used i cloning fragments of genomic DNA excluded the possibility of simple assessment of size

[c] ++indicates those clones that were recovered from duplica transfections in vivo

Table 4 provides a summary of the autonomous replicating DNA clones recovered by mass screenir assays for those anti-cruciform clone pools which we analyzed by both in vivo and in vitro methodologies. Th size of the isolated inserts ranges from 0.1 to 3.4 kb, wi a mean of 1.20 kb ± 0.28 (95% CI), and is thus not signi cantly different from the size range estimated for t clones of the anti-cruciform affinity-purified librari (1.4 kb ± 0.4).

Discussion

It has previously been shown that cruciform-like structures are associated with mammalian DNA replication activity (Zannis-Hadjopoulos et al. 1984; 1988). In subsequent experiments, we demonstrated a biphasic distribution of cruciform-like structures in mammalian cell nuclei as they progressed from the G1/S boundary through S phase (Ward et al. 1990). Recently, the cruciform-like structure was further implicated in DNA replication by the demonstration that anti-cruciform affinity-purified DNA is enriched for autonomously replicating DNA sequences (Bell et al. 1991). Since large libraries of DNA fragments enriched for autonomous replicating sequences were now available, a method for mass transfection of large numbers of individual, independent clones of DNA was needed.

A method of analysis which allows the isolation of autonomously replicating sequences from mass transfections of numerous individual genomic DNA clones is described. In order to perform such mass transfections, however, it was necessary to reduce the amount of any single plasmid DNA in a transfection (120 µg from 36–48 plasmids/1.8×10^6 HeLa cells) below the levels we have previously used (5–10 µg/3×10^5 HeLa cells). Such a difference in concentration requires that any individual plasmid would have to be proportionately more efficient in initiation of DNA replication in order that sufficient plasmid could be replicated and recovered in subsequent steps of the assay. In the course of demonstrating the ability of this method to recover the same plasmid out of a pool of up to 48 different recombinant plasmids, it was realized that this would identify yet another subset of plasmids which would be incrementally more efficient than those recovered as autonomously replicating sequences from only one of two duplicate transfections of a given pool of plasmids.

When a mixture of different plasmids is transfected into mammalian cells, intermolecular recombination may occur; the plasmids are, of course, composed of about 3 kb of homologous sequence, excluding the human genomic DNA fragment. However, we have analyzed by restriction enzyme digests and by sequence analysis over 34 independent clones, recovered from HeLa cells after transfection of pools of plasmids, without any indication of recombination. The frequency of recombinant events for mammalian cells may be expected to be of the order of 2×10^{-3} per rescued plasmid (Desautels et al. 1991). The application of this mass transfection method, as we have described, should therefore not be greatly confounded by recombination between co-transfected plasmids. However, a series of restriction digests must be performed to ensure the preservation of the cloning site and the basic structure of the pBluescript vector for any clones that are to be further characterized.

The in vitro mass screening replication assay resulted in a recovery of autonomously replicating sequences similar to that for the transfection assay, each selecting a subset of perhaps 10% of the input clones. In addition, the much smaller amount of DNA required for the in vitro replication assay (250 ng per assay rather than 120 µg) offers the possibility of scaling up the number of clones present in each pool for a more rapid analysis of the anti-cruciform libraries. However, the in vitro systems does represent further distancing of the replication system from what actually goes on during in vivo chromosomal replication. It has been previously demonstrated (Pearson et al. 1991) that this in vitro system is capable of performing only one round of replication per plasmid DNA molecule. In addition accessory factors may not be proportionately represented or may function differently in vitro by comparison to the transfection assay, thereby resulting in a lack of a perfect correlation between the two assays.

These mass transfection and in vitro assay methods will now afford the opportunity to assemble large panels of autonomously replicating sequences for characterization. Sequence analysis of over 30 monkey autonomously replicating sequences have failed to reveal a single consensus sequence; rather the features most commonly present are structural and include regions of high AT content and inverted repeat sequences which have the potential to form cruciforms (Rao et al. 1990; Landry and Zannis-Hadjopoulos 1991). The inability to detect a single consensus sequence may be due to the existence of different subsets of origins of DNA replication, which in turn may account for various aspects of their differential regulation, e.g. temporal regulation. In order to test such a hypothesis, a larger number of characterized sequences will be necessary to resolve any apparent heterogeneity of primary sequence into its distinctive subgroups. Furthermore, there may also be a difference in the origins of DNA replication of normal compared to malignantly transformed cells; the number of origins of replication is estimated to at least double with viral transformation (Martin and Oppenheim 1977). Mass assays of the libraries prepared from normal fibroblast and adenocarcinoma cells can quickly provide large series of autonomously replicating sequences from each source for comparison. It may also be possible to modify the mass screenings into a series of incremental steps that would provide defined groups of plasmids carrying human autonomously replicating sequences with different levels of efficiency for initiation of DNA replication. Such stratification of autonomously replicating sequences by their efficiency may result in a compartmentalization that would reveal the variables in structure and sequence which are the essential regulatory elements of origins of DNA replication.

Acknowledgements. This work was supported by the National Science and Engineering Research Council of Canada and the Medical Research Council of Canada. T.N. is the recipient of a Medical Research Council of Canada studentship.

References

Bell D, Sabloff M, Zannis-Hadjopoulos M, Price GB (1991) Anti-cruciform DNA affinity purification of active mammalian origins of replication. Biochim Biophys Acta 1089:299–308

Desautels L, Brouillette S, Chartrand P (1991) Reciprocal homologous junctions generated in mouse cells. Mol Gen Genet 227:356–360

Frappier L, Price GB, Martin RG, Zannis-Hadjopoulos M (1987) Monoclonal antibodies to cruciform DNA structures. J Mol Biol 193:751–758

Frappier L, Price GB, Martin RG, Zannis-Hadjopoulos M (1989) Characterization of the binding specificity of two anticruciform DNA monoclonal antibodies. J Biol Chem 264:334–341

Graham FL, Van Der Eb AJ (1973) A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology 52:456–467

Hirt B (1967) Selective extraction of polyoma DNA from infected mouse cell cultures. J Mol Biol 26:365–369

Landry S, Zannis-Hadjopoulos M (1991) Classes of autonomously replicating sequences are found among early-replicating monkey DNA. Biochim Biophys Acta 1088:234–244

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Martin RG, Oppenheim A (1977) Initiation points for DNA replication in nontransformed and Simian Virus 40-transformed Chinese hamster lung cells. Cell 10:859–867

Pearson CE, Frappier L, Zannis-Hadjopoulos M (1991) Plasmids bearing mammalian DNA-replication origin-enriched (*ors*) fragments initiate semiconservative replication in a cell-free system. Biochim Biophys Acta 1090:156–166

Rao BS, Zannis-Hadjopoulos M, Price GB, Reitman M, Martin RG (1990) Sequence similarities among monkey ori-enriched (*ors*) fragments. Gene 87:233–242

Schmid CW, Manning JE, Davidson N (1975) Inverted repeat sequences in the Drosophila genome. Cell 5:159–172

Southern E (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503–517

Ward GK, McKenzie R, Zannis-Hadjopoulos M, Price GB (1990) The dynamic distribution and quantification of DNA cruciforms in eukaryotic nuclei. Exp Cell Res 188:235–246

Wilson DA, Thomas CA Jr (1973) Hydroxyapatite chromatography of short double-helical DNA. Biochim Biophys Acta 331:330–340

Wilson DA, Thomas CA Jr (1974) Palindromes in chromosomes. J Mol Biol 84:115–144

Zannis-Hadjopoulos M, Kaufmann G, Martin RG (1984) Mammalian DNA enriched for replication origins is enriched for snap-back sequences. J Mol Biol 179:577–586

Zannis-Hadjopoulos M, Frappier L, Khoury M, Price GB (1988) Effect of anti-cruciform DNA monoclonal antibodies on DNA replication. EMBO J 7:1837–1844

ADDENDUM TO CHAPTER FOUR

The results describing individual clone transfection assays are added as a supplement to the original manuscript. Subsequent work has ensured that a total of 13 clones, some recoverable by transfection mass screen only, some by *in vitro* mass screening only, and some recovered by both methodologies, were tested for individual autonomous replication activity under the standard single-clone transfection and *in vitro* assay conditions described at the end of the Materials and Methods section and presented, in part, in Figure 5. The results of these experiments are summarized in Table 5:

Table 5. Comparison of transfection and in vitro autonomous replication assays for identifying putative origins from anti-cruciform affinity purified libraries.

Clone name	Recovery by mass screening protocol		Activity in individual autonomous replication assay[a]	
	Transfection	*In vitro*	Transfection	*In vitro*
F1	+	-	+	+
F2	-	+	+	+
F4	+	+	+	+
F5	+	+	+	+
F6	+	-	+	+
F11	-	+	+	+
F12	+	-	+	+
F15	+	+	+	+
F20	+	+	+	+
F21	-	+	+	+
S3	+	+	+	+
S14	+	-	+	+
S15	+	+	+	+

[a] Negative controls for these experiments were pBluescript vector plasmid, and pBluescript carrying a random 0.7 kb insert (plasmid 30.4; see also Chapter 5, Figure 1, and Chapter 6, Figure 4).

This data elaborates on the conclusions drawn from Table 3. Both the transfection and *in vitro* mass screens reveal the subset of anti-cruciform clones which have the potential to replicate autonomously, but they reveal these putative origins with differing efficiencies. Each mass screening methodology is a stringent test where moderately efficient origins may not survive the selection procedure, and the subsets recovered by mass transfection and mass *in vitro* replication show only a partial overlap. Nevertheless, as shown here in Table 5, when isolates are tested under standard single-clone transfection and *in vitro* assay conditions, all are positive for autonomous replication activity by both systems, regardless of the method(s) successfully used for their initial isolation. Thus, the differences observed between the mass screenings probably reflect differential sensitivities of activation of replication from among those clones with the potential for autonomous replication. As mentioned in the Discussion section, the cell extracts used for the *in vitro* assay may contain full initiator activity, but different proportions (when compared to transfected cells) of accessory factor activities responsible for modifying the efficiency with which particular origins are used.

In summary, anti-cruciform clones can be separated into three classes through application of mass screening techniques. Over 70% of clones are double negatives: not isolated by either transfection or *in vitro* mass screens (103 of 144 clones, as shown in Table 3A). Since David Bell's data suggests that about 50% of anti-cruciform clones have autonomous replication activity (Bell et al., 1991), this class includes the other 50% of clones which are non-replicating, together with a proportion of autonomously replicating sequences whose function is too weak to be detected with the increased stringency of the mass screening procedures. A second class of clones are recovered by one of the two mass screens, but not both. As shown in Table 5, all of these plasmids replicate autonomously in both transfection and *in vitro* individual clone assays, confirming that they are indeed autonomously replicating putative origins of DNA replication. The final class consists of double positives, autonomously replicating plasmids which are recovered by both types of mass screening assays. Less than 10% (11 of 144 tested with both assays, Table 3A) of the anti-cruciform clones fall into this category, and are likely to represent the most efficient origins of replication present in the libraries.

CHAPTER FIVE

AUTONOMOUS REPLICATION ACTIVITY OF A HUMAN MITOCHONDRIAL DNA SEQUENCE INSERTED INTO GENOMIC DNA

Torsten O. Nielsen, Mieczyslaw A. Piatyszek, Jerry W. Shay, Christopher E. Pearson, Maria Zannis-Hadjopoulos and Gerald B. Price

Preceding chapters have demonstrated that both transfection-based and *in vitro* autonomous replication assays work to identify mammalian sequences which can function as *cis*-active origins. In Chapter 5, I use these methods, together with a modified *in vitro* protocol designed to map the predominant replication initiation site on an autonomously replicating plasmid, in a study of a mitochondrial DNA sequence, an investigation which touches on the relation of origin function with genetic instability and transcriptional control.

This chapter was published in the November 1994 issue of the *International Journal of Oncology* (5:1003-1008) and is presented (with permission) as a journal reprint. The work was done in collaboration with the laboratory of Jerry W. Shay (University of Texas Southwestern Medical Center, Dallas), who provided plasmid pHL-1 and its sequencing data, and helped edit the manuscript for submission. The experiment presented as Figure 2 was performed by Christopher E. Pearson, and the input provided by Maria Zannis-Hadjopoulos was important in the design and execution of the kinetic mapping experiment of Figures 4 and 5. Gerald B. Price managed the collaborative effort and contributed extensively to the planning of the experiments and editing of the manuscript. Otherwise, I performed all experiments, assembled the results and wrote all portions of the initial and revised forms of the paper.

Autonomous replication activity of a human mitochondrial DNA sequence inserted into genomic DNA

TORSTEN O. NIELSEN[1], MIECZYSLAW A. PIATYSZEK[2], JERRY W. SHAY[2], CHRISTOPHER E. PEARSON[1], MARIA ZANNIS-HADJOPOULOS[1] and GERALD B. PRICE[1]

[1]McGill Cancer Centre, McGill University, 3655 Drummond Street, Montréal, Québec H3G 1Y6, Canada; [2]Department of Cell Biology and Neuroscience, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75235-9039, USA

Contributed by J.W. Shay, September 2, 1994

Abstract. Mitochondrial DNA (mtDNA) fragments can be found inserted into nuclear DNA and may contribute to cancer and aging. A HeLaTG-cell nuclear transcript was shown to include human cytochrome oxidase subunit 3 (coxIII) mtDNA fused to c-*myc* sequences. Independently, a coxIII-containing cDNA was discovered to replicate autonomously in HeLa cells. We show that the HeLaTG mtDNA insertion has autonomous replication activity when transfected into HeLa cells and in a mammalian *in vitro* replication system. Replication *in vitro* starts within the fusion cDNA sequence. Insertion of an element which could function as a replication origin may affect cellular DNA replication and proliferation.

Introduction

Evidence from many sources suggests a link between transcription and the function of eukaryotic origins of DNA replication (1). Possible mechanisms for transcriptional activation of origin function include the production of RNA primers, induction of structural alterations in the origin region and the potential accessory role of transcription factors in loading of the replication complex. To study the association between these two fundamental processes, we tested (2) human embryonic lung fibroblast cDNA sequences with homology to monkey *ors* putative origins of DNA replication (3) for their ability to confer autonomous replication onto otherwise non-replicating prokaryotic plasmid vector upon transfection into human cells. One active isolate, NOA5 (2), contained neither repetitive DNA sequences, nor matches to the yeast ARS or scaffold-attachment region consensus, but did contain AT-rich regions and numerous small inverted repeat sequences which may have a role in chromosomal origin function (4-7). The cDNA insert of this clone matches human mitochondrial DNA (mtDNA) positions 9248-9998 (8), corresponding to the cytochrome c oxidase subunit 3 (coxIII) transcript (position 9210-9990 in mtDNA) and tRNA (9991-10058) before final processing and cleavage of adjoining tRNA (9).

A coxIII-containing cDNA specific to HeLaTG cells was isolated and cloned (a pBluescript clone, designated as pHL-1) on the basis of homology with both mtDNA and c-*myc*, as part of a study into the role of mtDNA sequences as mobile genetic elements (10). Fragments of mtDNA produced by incomplete autophagic degradation have the potential for insertion into nuclear DNA at relatively high frequency (11) and may thereby contribute to cancer and the aging process in general (12,13). Evidence from several groups is consistent with the possibility that mitochondrial genes may be present and even transcribed in the nucleus of several types of tumor cells (14,15). The cDNA contained in plasmid pHL-1 contains coxIII (mtDNA 9217-9990) and c-*myc* exons 2 and 3 (nucleotides 5139-5277 and 6654-7682, respectively) (16,17) and is derived from a fusion mRNA transcribed from one allele of c-*myc* in HeLaTG cells into which the coxIII mtDNA has become inserted (10).

The mitochondrial portions of pHL-1 and NOA5 contain neither the strand-specific unidirectional origins of mitochondrial DNA replication (oriL near mtDNA position 5750 and oriH in the D-loop region around position 200) (18), nor the mitochondrial transcription promoters (also found in the D-loop region) (19); in fact, the coxIII/c-*myc* fusion mRNA is transcribed as the mitochondrial L-strand sequence, which carries numerous stop codons. Furthermore the c-*myc* portion of pHL-1 does not contain the human genomic origin of DNA replication localized in HeLa cells 1.5 kb 5' to c-*myc* exon 1 (20) which has been shown to possess autonomous replication activity (21). However, there is a precedent for *S. cerevisiae* autonomous replication (ARS) activity in yeast mtDNA fragments (22,23), as well as for yeast ARS activity in mtDNA fragments obtained from other fungi (24,25), *Xenopus* (26), *Drosophila* (27) and from chloroplast DNA (28,29). These fragments can be completely independent of mitochondrial replication origins (30) and replicate autonomously outside the yeast mitochondria

Correspondence to: Dr Gerald B. Price, McGill Cancer Centre, 3655 Drummond Street, Montréal, Québec H3G 1Y6, Canada

Key words: DNA replication, *in vitro* replication, replication origin, mitochondrial DNA, insertional mutagenesis, cytochrome oxidase III

organelle (22,31); they generally possess AT-rich regions, homology with the yeast ARS consensus and inverted repeat elements.

Given the demonstrated autonomous replication capacity in human cell transfection assays, of clone NOA5 carrying coxIII outside its mitochondrial context and the disruptive presence of this mtDNA sequence in an oncogene with a closely associated origin of DNA replication, we analyzed the plasmid pHL-1 for its ability to replicate autonomously in the human system.

Materials and methods

Sequences and plasmids. All sequences may be accessed in GenBank. pHL-1: Locus HUMPHL1 and Accession No. X54629. The complete human mitochondrial genome: Locus HUMMTCG and Accession Nos. J01415, M12548, M58503, M63932, M63933. cDNA clone 343: Locus HUMAUTONJ and Accession No. L08443. NOA5: Locus HUMAUTONH and Accession No. L08441. *ors*8: Locus AGMORS8A and Accession No. M26221. The c-*myc* and flanking regions: Locus HUMMYC and Accession No. J00120. 950C was created by subcloning the 957 bp *Cla*I fragment of pHL-1, in reverse orientation, into the unique *Cla*I site of pBR322. Plasmid 30.4 is a pBluescript clone containing a random breast tumor cDNA insert of 0.7 kb. All plasmids used in these studies were propagated in bacterial hosts, HB101 or XL-1 Blue and media. Plasmid DNA was isolated by alkaline lysis of cells and purified by two rounds of CsCl/ethidium bromide gradient centrifugations (32).

Replication assay by plasmid transfection. HeLa cells were seeded at a density of 2.5×10^5 cells per 20 cm^2 flask and allowed to grow overnight in DMEM (Gibco) supplemented with 10% FCS (Flow Laboratories), 50 U/ml penicillin (Gibco), 50 µg/ml streptomycin (Gibco) and 292 mg/l L-glutamine (Sigma). Medium was replaced and the cells allowed to grow 3 h prior to calcium phosphate coprecipitation transfection (33), with 5 µg of test plasmid DNA: *ors*8 (3), pHL-1 (10), 30.4, pBR322, or 950C. After overnight incubation, the cells were washed and left to grow 48 h in 5 ml fresh medium. Cells were lysed by the method of Hirt (34) and low molecular weight DNA isolated and purified as described previously (35). Resulting DNA was digested for 1 h with 1 U *Dpn*I to digest unreplicated (bacterially methylated) plasmid (36), or divided into three parts, one of which was left uncut, the second was digested with *Dpn*I (1 U, 1 h) and the third digested with *Mbo*I (1 U, 1.5 h); products were electrophoresed in a 1% agarose gel, blotted and probed with nick-translated pBR322 DNA (37), or a random-primed 692 bp *Dra*I fragment of pBR322. Analysis by quantitative densitometry of exposed XAR-5 film (Kodak) was performed using the Bio Image (Millipore) Whole Band Analysis program.

In vitro replication assay. 100 ng supercoiled pBR322 and pHL-1 DNA prepared as described earlier were used as templates in a cell-free replication system (38) composed of HeLa low salt cytoplasmic extract, HeLa nuclear extract, an ATP regenerating system, PEG and a dNTP mix including [α^{32}P]-dCTP and [α^{32}P]-dTTP. Following 1 h incubation a 30°C, DNA isolation and purification, half of the product wa digested 6 h with 1 U *Dpn*I and electrophoresed alongsid the undigested samples, through a 1% agarose, 1x TAE g (16 h, 2.5 V/cm). The gel was dried and exposed to XAR- film overnight.

Kinetics of pHL-1 labeling by the in vitro replication syster 100 ng pHL-1 was used as template for triplicate *in vitr* replication assays as described above, but this time with th stop mix added to separate reactions at three evenly space time points up to 20 minutes. The purified product wa digested 1.5 h with 10 U *Rsa*I, cutting the vector twice ar the cDNA insert four times to yield fragments of 1902 bp ((which all but 77 bp is vector DNA), 1575 bp (entirel cDNA), 1090 bp (entirely vector DNA), 289 bp (of whic 243 bp is cDNA) and 30 and 16 bp (both entirely cDNA One-half the *Rsa*I product was further digested for 1.5 h wi 10 U *Sty*I, which cuts only the large cDNA fragmen liberating 748, 336, 248 and 243 bp subfragments ar samples were electrophoresed overnight at 1.2 V/cm in 2 agarose, 1x TAE. The gel was dried and used to expose bo XAR-5 film and a Fujix BAS 2000 phospho-imager scree for quantitation, exposures using phospho-imager screer providing a convenient and reliable system for 'linea quantitation of amount of radioactivity incorporated p band. Quantitative densitometry was performed aft scanning of the phospho-imager screen using the Bio Ima (Millipore) Whole Band Analysis program.

Results

The potential for autonomous replication can b demonstrated by the ability of a cloned fragment to allo non-replicating prokaryotic plasmid vector to replicate aft transfection into human host cells. Using the previous identified autonomously replicating clone *ors*8 (37) as positive control, we tested pHL-1 for its ability to replica into a *Dpn*I-resistant form following its transfection in HeLa cells by calcium phosphate coprecipitation ar subsequent short-term cell growth (48 h). Results a presented in Fig. 1A. pHL-1 produces the same strong for III (linear) band seen for *ors*8 and while producing somewh less form I (supercoiled) *Dpn*I-resistant product, the over level of replication of pHL-1 is similar to that of the high active positive control (i.e., by quantitative densitometry, t total *Dpn*I-resistant DNA obtained for pHL-1 is 70% of th seen for *ors*8). Form I DNA levels are expected to va somewhat in *Dpn*I-resistance assays like the one shown he In general, form I levels are usually lower than those forms II or III (35), in part because *Dpn*I may sometim result in nicking of hemimethylated form I plasmid (39) because of variability in topoisomerase activity.

For further confirmation of the autonomous replication coxIII-containing sequences, a 957 bp *Cla*I fragment of pHL- containing all of the coxIII portion, was subcloned into t pBR322 vector (to create plasmid 950C) and tested f autonomous replication activity in a transfection ass together with pHL-1 and negative control plasmids 30.4 a pBR322 (Fig. 1B), by digestion with *Dpn*I and *Mbo*I (40).

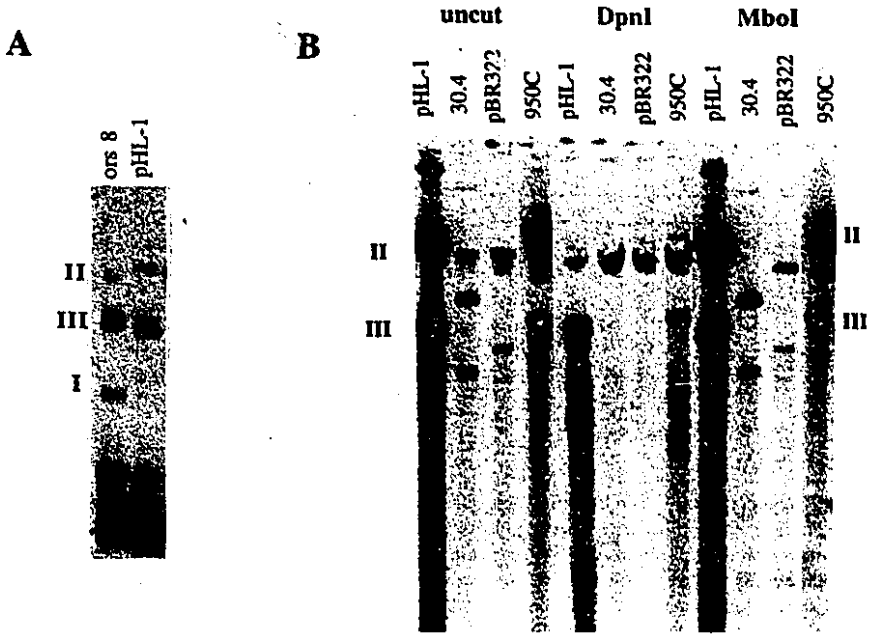


Figure 1. *In vivo* episomal DNA replication of pHL-1. For each indicated plasmid, 5 µg samples were transfected into HeLa host cells by calcium phosphate coprecipitation. After two days, cells were lysed and low molecular weight DNA isolated by the method of Hirt (34), was digested with *Dpn*I (A), or split evenly into aliquots that were either not treated (uncut), digested with *Dpn*I, or digested with *Mbo*I (B). Samples were then electrophoresed in a 1% agarose gel, blotted and probed with vector sequences. The approximate positions of forms I, II and III plasmid DNA are indicated.

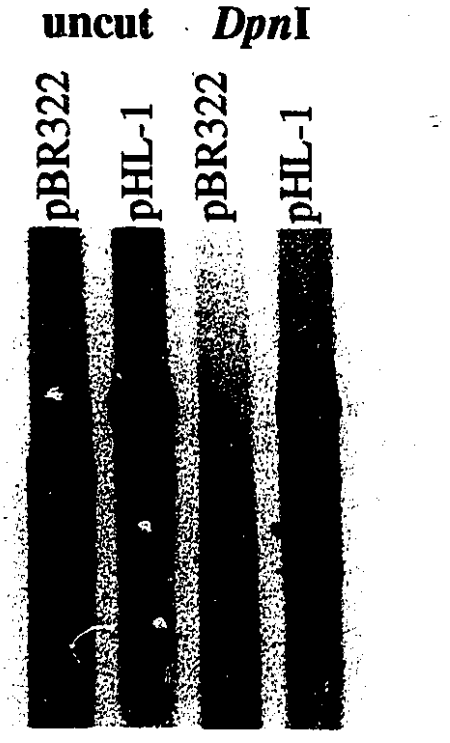


Figure 2. *In vitro* DNA replication of pHL-1. 100 ng aliquots of supercoiled pBR322 and pHL-1 DNA were replicated using the cell-free system of Pearson *et al* (38), in the presence of α-^{32}P-dCTP and α-^{32}P-TTP. The products were divided into two equal parts and analyzed by electrophoresis on 1% agarose gel either undigested or following digestion by *Dpn*I. The uppermost band is form II and the next lower band is form III plasmid DNA.

this experiment, little if any form I DNA was recovered (from all transfections, some contaminating linear genomic DNA produced an *Mbo*I-sensitive band present in the *Dpn*I and uncut lanes). More total episomal DNA was recovered at 48 h post-transfection from plasmid 950C and especially pHL-1 compared to plasmids 30.4 and pBR322 (Fig. 1B, uncut); furthermore, 950C and pHL-1 bands are *Dpn*I-resistant, unlike those of 30.4 and pBR322 (Fig. 1B, *Dpn*I). All plasmid bands are *Mbo*I-resistant, indicating that the pHL-1 and 950C transfections yield hemimethylated products from a single round of replication, as expected in a short term assay.

A second method for the demonstration of autonomous replication capacity utilizes an *in vitro* replication system (38) based on HeLa cell extracts and an ATP regenerating system. Fig. 2 presents the results of such an *in vitro* assay, using pHL-1 as a template and pBR322 as a negative control. While both plasmids incorporate precursor radionucleotide after one hour incubation, only the incorporation into pHL-1 is *Dpn*I-resistant, indicating that the plasmid has been replicated completely and thus all the *Dpn*I recognition sites have been converted to hemi-methylated or unmethylated sequences that are resistant to digestion. Comparison, by quantitative densitometry, of the relative levels of the *Dpn*I-resistant products that were generated by *in vitro* replication products showed that by this assay pHL-1 possesses 110% of the replication capacity of *ors*8. The coxIII-containing clone NOA5, previously shown to replicate autonomously by the transfection assay (2) was also found to be strongly positive by the *in vitro* replication assay (data not shown). Fig. 3 shows a comparison of the mtDNA sequences present in cox III-containing clones with autonomous replication activity.

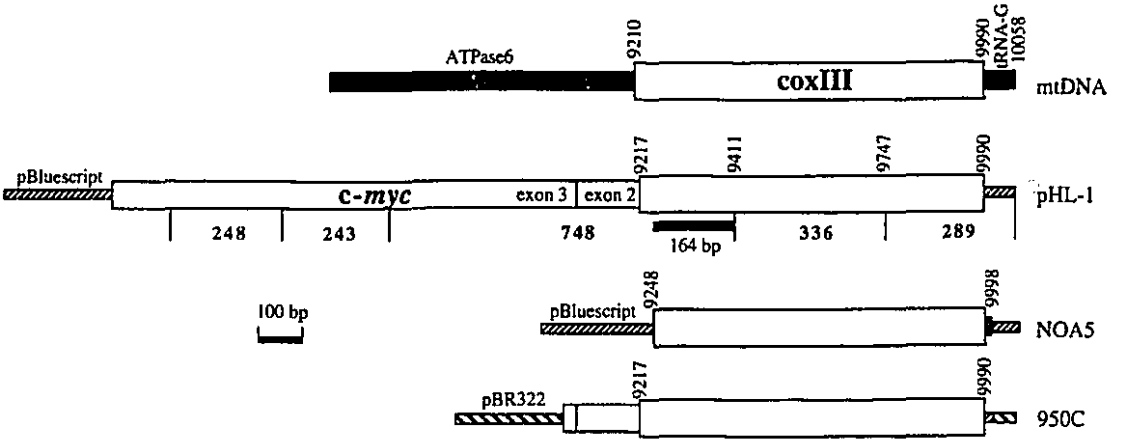


Figure 3. Schematic representation of coxIII-containing clones. Top line: the full length coxIII sequence in its native position in mitochondrial DNA, flanke by the ATPase6 and glycine-tRNA loci. Second line: insert region of autonomously replicating plasmid pHL-1. Numbers above indicate mtDNA positic equivalents for ends and relevant internal restriction sites. Positions of pHL-1-specific restriction fragments (748, 336, 289, 248 and 243 bp) generated b double digestion with *Rsa*I and *Sty*I are indicated (bold); the 164 bp region, common to the earliest labeled fragment and to all autonomously replicatin coxIII-containing clones, is also indicated (black box). Third and fourth lines: inserts of NOA5 and 950C, coxIII-containing clones which also posse autonomous replication activity, with type of vector and mtDNA positions of insert ends shown.

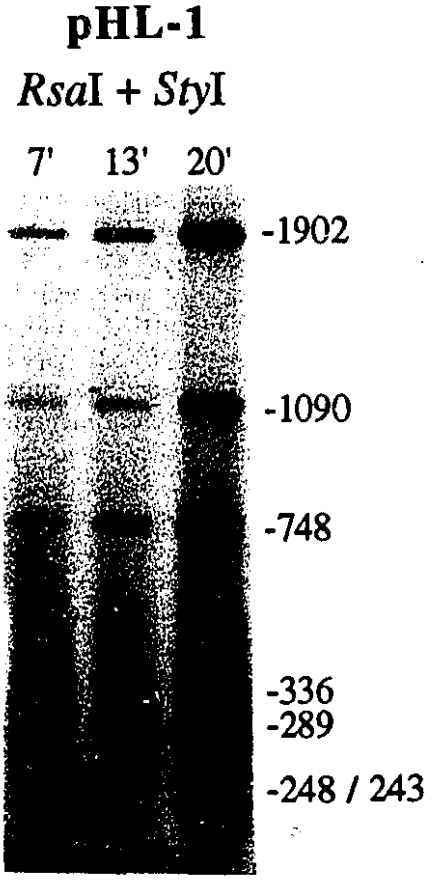


Figure 4. *In vitro* DNA replication kinetics of pHL-1. 100 ng aliquots of pHL-1 DNA were used as template in parallel *in vitro* replication reactions, which were stopped after 7, 13, or 20 minutes. After isolation and purification of DNA, the product was digested consecutively with *Rsa*I and then *Sty*I before electrophoretic separation in a 2% agarose gel. The gel was then dried and exposed to X-ray film. pHL-1 is cut into 1902 and 1090 bp fragments containing mainly pBluescript vector sequences and a series of smaller fragments carrying the coxIII/c-*myc* insert.

Because of the ease of its manipulation, high cop number of the template and inherent initial synchronizatior the *in vitro* replication system is convenient for analyzin origin position in an autonomously replicating plasmid. Thi is accomplished by premature termination of the reaction a early time points, followed by restriction enzyme mapping c the substrate plasmid to determine into which subfragment radiolabeled nucleotides are most efficiently incorporated i the initial stages of the reaction (38). This approach is simila to the 'earliest labeled DNA fragment' origin mappin strategy (41). Fig. 4 depicts the autoradiogram produce from a *Rsa*I/*Sty*I digest analysis of pHL-1 replicated *in vitr* in a time-course reaction (up to 20 minutes). The two large bands contain mainly vector sequences (the 1902 b fragment includes 77 bp of c-*myc* exon 3), while th remainder contains most of the cDNA sequence; two ver short fragments (30 and 16 bp) of cDNA are lost to furthe analysis. A schematic map of the pHL-1 fragments and the relationships is presented at the bottom of Fig. 5.

Quantitative densitometry, after background is subtracte and signal is normalized for fragment size, reveals that th greatest incorporation takes place within the cDNA inser specifically, the apparent initiation site is localized to the 74 bp fragment containing both coxIII and c-*myc* sequence Note that the 243 and 248 bp bands are not resolved on th gel and had to be analyzed as a unit. The signal-to-noise rati present on the gel in the region below 400 bp makes th potential measurement error relatively high for the 336, 28 and 248/243 bp fragments; however, background i insignificant relative to signal for the three largest bands meaning that even by the most conservative interpretatior the preferred intiation site for *in vitro* autonomous replicatio lies within the insert. The degree of labeling preference when expressed as a ratio of the incorporation per base of th maximally and minimally labeled fragments, is very simila for pHL-1 (2.1) as compared with *ors*3, *ors*8 and *ors*9 (2.2 2.4) (38). It should be noted that the validity of plasmi origin localization by *in vitro* fragment labeling has no

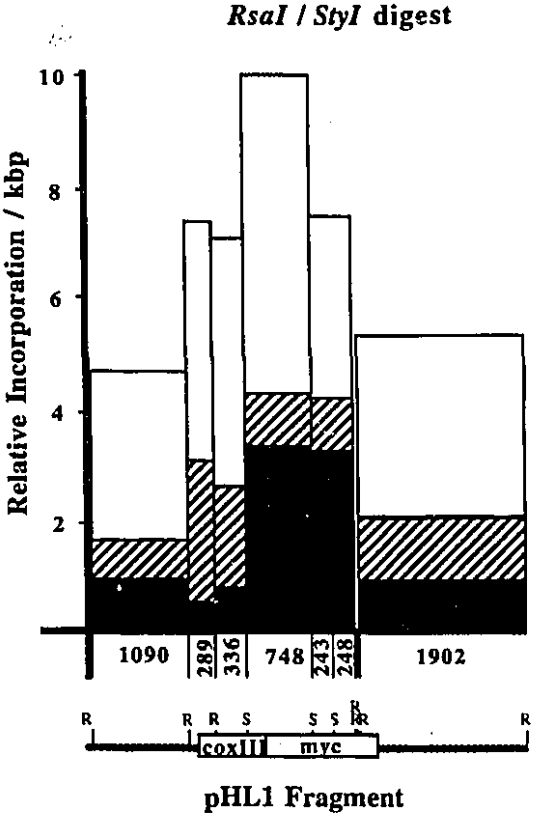


Figure 5. Graphic representation of *in vitro* replication of pHL-1 fragments. The gel shown in Fig. 4 was used to expose a phospho-imager plate and the bands analyzed by quantitative densitometry. Results are expressed as normalized for fragment size and relative to the strongest incorporating band (assigned a value of 10). Black bars denote relative incorporation/kbp at 7 minutes, while striped and white boxes represent the cumulative radionucleotide incorporation at 13 and 20 minutes, respectively. Immediately below the graph, the fragments corresponding to the linearized map of pHL-1 are indicated, where vector sequences are represented as a striped line and the coxIII/c-*myc* containing insert are shown as boxes. R=*Rsa*I site; S=*Sty*I site.

been independently confirmed, in the case of *ors*8, by electron microscopic mapping (42).

When the same analysis was performed with an *Rsa*I single digest, the strongest labeling was observed in the 1575 bp cDNA fragment (encompassing the 336, 748, 243 and 248 bp subfragments in the *RsaI/StyI* double digest) and when the pBluescript vector plasmid alone is analyzed for incorporation of radionucleotide *in vitro*, no preference is seen in the labeling of its two *Rsa*I fragments (data not shown).

Discussion

The 1941 bp insert of pHL-1 possesses autonomous replication activity which has been demonstrated by both the transfection and *in vitro* assay methodologies. NOA5, a cDNA clone previously shown to replicate autonomously (2) and pHL-1 share overlapping mitochondrial DNA sequences derived from the coxIII gene (mtDNA 9248-9990). If the preferred site of initiation on pHL-1 lies in the 748 bp *Sty*I central fragment, containing coxIII (mtDNA 9217-9411) and c-*myc* exons 2 and 3 sequences, a common initiation region between pHL-1 and NOA5 might lie in a 164 bp region (mtDNA 9248-9411) of coxIII. The various constructs can be directly compared in Fig. 3.

Autonomous replication activity may not always correspond to *in vivo* origin activity when the sequence is in its proper chromosomal context (43). However, a PCR-based *in vivo* origin mapping method, sensitive to chromosomal context, now exists (44). When the chromosomal region encompassing clone 343, a human cDNA isolated in the same fashion as NOA5 and demonstrating autonomous replication activity by transfection assay (2), was tested by nascent replicated strand PCR amplification, the presence of a functional *in vivo* origin was detected and localized to a 1.6 kb region inclusive of the 343 sequence (45); subsequently the 343 locus was mapped to the q22-qter region of human chromosome 6 (46). Thus, there is reason to believe that the coxIII sequence present in NOA5 and pHL-1 may also serve as a replication origin after its insertion into nuclear DNA and could possibly disrupt the normal functioning of the neighboring c-*myc* 5' origin in HeLaTG cells.

The effect of insertion into the nuclear genome of certain mitochondrial DNA sequences, which contain elements which are sufficient for initiation of DNA replication in nuclear DNA, would add significantly to arguments regarding putative roles for insertion of mtDNA as a mechanism of disruption of normal regulation of cellular proliferation. A disruption in not only transcription and expression of the c-*myc* gene, but also in the control and initiation of the gene's normal chromosomal replication could have major consequences for regulation of replication and cell proliferation and represent one type of dysfunction associated with malignant transformation and aging.

Acknowledgements

We thank P. Leder of Harvard Medical School for providing the lambda phage clone containing the human c-*myc* gene and flanking DNA. T.O.N. is a recipient of a MRC studentship. C.E.P. is a recipient of a Graduate Faculty Award (Faculty of Medicine). This research was supported by the National Cancer Institute of Canada with funds from the Canadian Cancer Society and by NIH grant CA4006 (J.W.S.).

References

1. DePamphilis ML: Eukaryotic DNA replication: anatomy of a origin. Annu Rev Biochem 62: 29-63, 1993.
2. Wu C, Friedlander P, Lamoureux C, Zannis-Hadjopoulos M an Price GB: cDNA clones contain autonomous replicatio activity. Biochim Biophys Acta 1174: 241-257, 1993.
3. Zannis-Hadjopoulos M, Kaufmann G, Wong S, Leehner I Karawya E, Hesse J and Martin R: Properties of some monke DNA sequences obtained by a procedure that enriches for DN replication origins. Mol Cell Biol 5: 1621-1629, 1985.
4. Rao BS, Zannis-Hadjopoulos M, Price GB, Reitman M ar Martin RG: Sequence similarities among monkey ori-enriche (ors) fragments. Gene 87: 233-242, 1990.
5. Zannis-Hadjopoulos M, Kaufmann G and Martin RC Mammalian DNA enriched for replication origins is enriche for snap-back sequences. J Mol Biol 179: 577-586, 1984.
6. Zannis-Hadjopoulos M, Frappier L, Khoury M and Price GI Effect of anti-cruciform DNA monoclonal antibodies on DN replication. EMBO J 7: 1837-1844, 1988.

7. Zannis-Hadjopoulos M, Pearson CE, Bell D, Mah D, McAlear M and Price GB: Structural and functional characteristics of autonomously replicating mammalian origin-enriched sequences (ors). In: DNA Replication: The Regulatory Mechanisms. Hughes P, Fanning E and Kohiyama M (eds) Berlin, Springer-Verlag, pp107-116, 1992.
8. Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R and Young IG: Sequence and organization of the human mitochondrial genome. Nature 290: 457-464, 1981.
9. Clayton DA: Transcription of the mammalian mitochondrial genome. Annu Rev Biochem 53: 573-594, 1984.
10. Shay JW, Baba T, Zhan Q, Kamimura N and Cuthbert JA: HeLaTG cells have mitochondrial DNA inserted into the c-*myc* oncogene. Oncogene 6: 1869-1874, 1991.
11. Thorsness PE and Fox TD: Escape of DNA from mitochondria to the nucleus in *Saccharomyces cerevisiae*. Nature 346: 376-379, 1990.
12. Shay JW and Werbin H: Cytoplasmic regulation of cellular differentiation: a role for mitochondria in carcinogenesis. In: Mechanisms of Differentiation, Vol 1. Fisher PB (ed) Florida, CRC Press, pp135-159, 1990.
13. Shay JW and Werbin H: New evidence for the insertion of mitochondrial DNA into the human genome: significance for cancer and aging. Mutation Res 275: 227-235, 1992.
14. Corral M, Baffet G, Kitzis A, Paris B, Tichonicky L, Kruth J, Guguen-Guillouzo C and Defer N: DNA sequences homologous to mitochondrial genes in nuclei from normal rat tissues and from rat hepatoma cells. Biochem Biophys Res Commun 162: 258-264, 1989.
15. Heerdt BG, Halsey HK, Lipkin M and Augenlicht LH: Expression of mitochondrial cytochrome c oxidase in human colonic cell differentiation, transformation and risk for colonic cancer. Cancer Res 50: 1596-1600, 1990.
16. Alitalo K, Schwab M, Lin CC, Varmus HE and Bishop JM: Homogeneously staining chromosomal regions contain amplified copies of an abundantly expressed cellular oncogene (c-*myc*) in malignant neuroendocrine cells from a human colon carcinoma. Proc Natl Acad Sci USA 80: 1707-1711, 1983.
17. Marcu KB, Harris LJ, Stanton LW, Erikson J, Watt R and Croce CM: Transciptionally active c-*myc* oncogene is contained within NIARD, a DNA sequence associated with chromosome translocations in B-cell neoplasia. Proc Natl Acad Sci USA 80: 519-523, 1983.
18. Clayton DA: Replication of animal mitochondrial DNA. Cell 28: 693-705, 1982.
19. Clayton DA: Transcription and replication of animal mitochondrial DNAs. Int Rev Cytol 141: 217-232, 1992.
20. Vassilev LT and Johnson EM: An initiation zone of chromosomal DNA replication located upstream of the c-*myc* gene in proliferating HeLa cells. Mol Cell Biol 10: 4899-4904, 1990.
21. McWhinney C and Leffak M: Autonomous replication of a DNA fragment containing the chromosomal replication origin of the human c-*myc* gene. Nucleic Acids Res 18: 1233-1242, 1990.
22. Blanc H and Dujon B: Replicator regions of the yeast mitochondrial DNA active *in vivo* and in yeast transformants. In: Mitochondrial Genes. Slonimski P, Borst P and Attardi G (eds) New York, Cold Spring Harbor Laboratory Press, pp279-294, 1981.
23. Delouya D and Norbrega FG: Mapping of the ARS-like activity and transcription initiation sites in the non-canonical yeast mitochondrial ori 6 region. Yeast 7: 51-60, 1991.
24. Tikhomirova LP, Kryokov VM, Strizhov NI and Bayev AA: mtDNA sequences of *Candida utilis* capable of supporting autonomous replication of plasmids in *Saccharomyces cerevisiae*. Mol Gen Genet 189: 479-484, 1983.
25. Tudzynski P and Esser K: Nuclear association in yeast of a hybrid vector containing mitochondrial DNA. Curr Genet 7: 165-166, 1983.
26. Zakian VA and Kupfer DM: Replication and segregation of a unstable plasmid in yeast. Plasmid 8: 15-28, 1982.
27. Marunouchi T, Matsumoto Y, Hosoya H and Okabayashi K: I addition to the ARS core, the ARS box is necessary fo autonomously replicating sequences. Mol Gen Genet 206: 6(65, 1987.
28. Ohtani T, Uchimiya H, Kato A, Harada H, Sugita M an Sugimura M: Location and nucleotide sequence of a tobacc chloroplast DNA segment capable of replication in yeast. Mc Gen Genet 195: 1-4, 1984.
29. Vallet JM, Rahire M and Rochaix JD: Localization an sequence analysis of chloroplast DNA sequences o *Chlamydomonas reinhardii* that promote autonomou replication in yeast. EMBO J 3: 415-421, 1984.
30. Mabuchi T, Nishigawa S and Wakabayashi K: Nucleotid sequence of an essential region for autonomous replication c cloned yeast mitochondrial DNA. J Biochem 95: 589-592, 198
31. Delouya D, Bonjardim CA and Nobrega FG: ARS activit along the yeast mitochondrial apocytochrome b regior correlation with the location of petite genomes and consensu sequences. Curr Genet 12: 583-589, 1987.
32. Maniatis T, Fritsch EF and Sambrook JF: Molecular Cloning: Laboratory Manual. New York, Cold Spring Harbor Laboratory 1983.
33. Graham FL and van der Eb AJ: A new technique for the assa of infectivity of human adenovirus 5 DNA. Virology 52: 45(467, 1973.
34. Hirt B: Selective extraction of polyoma DNA from infecte mouse cell cultures. J Mol Biol 26: 365-369, 1967.
35. Landry S and Zannis-Hadjopoulos M: Classes of autonomousl replicating sequences are found among early-replicatin monkey DNA. Biochim Biophys Acta 1088: 234-244, 1991.
36. Peden KWC, Pipas JM, Pearson-White S and Nathans D Isolation of mutants of an animal virus in bacteria. Science 20! 1392-1396, 1980.
37. Frappier L and Zannis-Hadjopoulos M: Autonomous replicatio of plasmids bearing monkey DNA origin-enriched sequence: Proc Natl Acad Sci USA 84: 6668-6672, 1987.
38. Pearson CE, Frappier L and Zannis-Hadjopoulos M: Plasmid bearing mammalian DNA-replication origin-enriched (or: fragments initiate semiconservative replication in a cell-fre system. Biochim Biophys Acta 1090: 156-166, 1991.
39. Sanchez JA, Marek D and Wangh LJ: The efficiency and timin of plasmid DNA replication in *Xenopus* eggs: correlations to th extent of prior chromatin assembly. J Cell Sci 103: 907-91! 1992.
40. Virta-Pearlman VJ, Genaratne PH and Chinault AC: Analysis c a replication initiation sequence from the adenosine deaminas region of the mouse genome. Mol Cell Biol 13: 5931-594: 1993.
41. Vassilev LT and DePamphilis ML: Guide to identification c origins of DNA replication in eukaryotic cell chromosomes. Cr Rev Biochem Mol Biol 27: 445-472, 1992.
42. Pearson CE, Shihab-El-Deen A, Price GB, Zannis-Hadjopoulos M Electron microscopic analysis of *in vitro* replication products c ors 8, a mammalian origin enriched sequence. Somat Cell Mc Genet 20: 147-152, 1994.
43. Umek RM, Linskens MHK, Kowalski D and Huberman JA: Ne beginnings in studies of eukaryotic DNA replication origin Biochim Biophys Acta 1007: 1-14, 1989.
44. Vassilev LT and Johnson EM: Mapping initiation sites of DN. replication *in vivo* using polymerase chain reactio amplification of nascent strand segments. Nucleic Acids Res 1' 7693-7705, 1989.
45. Wu C, Zannis-Hadjopoulos M and Price GB: *In vivo* activity fc initiation of DNA replication resides in a transcribed region c the human genome. Biochim Biophys Acta 1174: 258-26(1993.
46. Shihab-El-Deen A, Wu C, Hamerton JL, Zannis-Hadjopoulos M and Price GB: Chromosomal localization of a sequence with *vivo* activity for initiation of DNA replication. Somat Cell Mc Genet 19: 103-109, 1993.

CHAPTER SIX

IDENTIFICATION OF A PUTATIVE DNA REPLICATION ORIGIN IN THE γ-AMINOBUTYRIC ACID RECEPTOR SUBUNIT β3 AND α5 GENE CLUSTER ON CHROMOSOME 15q11q13, A REGION ASSOCIATED WITH ALLELE-SPECIFIC REPLICATION TIMING

D. Sinnett, E. Woolf, W. Xie, K. Glatt, E.F. Kirkness, T.O. Nielsen, M. Zannis-Hadjopoulos, G.B. Price, and M. Lalande

Preceding chapters have shown the utility of the *in vitro* autonomous replication assay as a test of origin function and as a method for mapping origin activity. In this chapter, these techniques are applied to the search for a replication origin in an imprinted genomic region, associated with human genetic diseases, that displays allele-specific replication timing asynchrony.

This chapter was completed as a collaborative effort with the laboratory of Marc Lalande (Harvard University, Boston), and has been submitted to the journal *Gene* for publication. My responsibilities included the design, execution, analysis and write-up of the functional screening of the panel of cloned fragments (section (e) and Figure 4) and the mapping of the predominant initiation site in p82 (section (f) and Figure 5). In addition, I developed and utilized a computer program for the statistical analysis of the significance of observed sequence motifs, the results of which are incorporated into section (b) and Table I. This program, written in BASIC, is included as an addendum to the manuscript, together with tabulated results from the statistical analysis, in a more thorough form than could be included in the submitted article. Finally, I edited and revised all portions of the manuscript through several drafts.

Identification of a putative DNA replication origin in the γ-aminobutyric acid receptor subunit β3 and α5 gene cluster on chromosome 15q11q13, a region associated with allele-specific replication timing.

Key words: parental imprinting, *in vitro* DNA replication assay, lambda phage cloning, bent DNA, DNA helical stability, topoisomerase II binding site.

D. Sinnett[1,2*], E. Woolf[1,3], W. Xie[1], K. Glatt[1], E.F. Kirkness[4], T. O. Nielsen[5], M. Zannis-Hadjopoulos[5], G.B. Price[5], and M. Lalande[1,2,3]. [1]*Genetics Division, Children's Hospital,* [2]*Department of Pediatrics, Harvard Medical School,* [3]*Howard Hughes Medical Institute, Boston, MA 02115 USA (617) 355-7387,* [4]*The Institute for Genomic Research, Gaithersberg, MD 20878 USA (301) 443-5880* and [5]*McGill Cancer Centre, McGill University, Montreal, H3G 1Y6 (514) 398-3536 Canada.*

Abbreviations: ARS, autonomously replicating sequence; *DHFR*, dihydrofolate reductase gene; DUE, DNA unwinding element; FISH, fluorescence *in situ* hybridization; GABRA5, γ-aminobutyric acid α5 receptor subunit gene; GABRB3, γ-aminobutyric acid β3 subunit gene; SAR, scaffold attachment region.

* Present address: Hematology-Oncology
Ste-Justine Hospital
3175 Côte Ste-Catherine
Montréal (Québec)
H3T 1C5 Canada

SUMMARY

The region containing the $GABA_A$ receptor β3 and α5 subunit genes is subject to parental imprinting and is organized in different allele-specific replication timing domains. A 60 kb domain displaying a maternal early/paternal late pattern of allele-specific replication timing asynchrony is nested within a larger region displaying the opposite pattern. The proximal portion of this maternal early replicating domain is incorporated into phage clone λ84. In order to identify DNA structures which may be associated with the boundary between the replication domains, phage λ84 has been subcloned into smaller fragments and several of these have been analyzed by nucleotide sequencing. A plot of helical stability for 13 kb of contiguous sequence reveals several A-T rich regions which display potential DNA unwinding. The plasmid subclones from phage λ84 have been assayed for bent DNA and one of these, p82, contains bent DNA and overlaps with the region of highest potential helical instability. Of the seven plasmids tested, only p82 shows strong autonomous replication activity in an *in vitro* replication assay, with replication initiating within the genomic insert. These results suggest that a putative origin of DNA replication contained within p82 may play a role in establishing the allele-specific replication timing domains in the $GABA_A$ receptor subunit gene cluster.

INTRODUCTION

Parental imprinting, which marks the parental origin of chromosomes, results in allele-specific changes in chromatin organization, transcription as well as replication. One of the best examples of parental imprinting in humans occurs in chromosome 15q11-q13 where the absence of a paternal contribution to this region, either by deletion or through uniparental disomy, results in Prader-Willi syndrome, while lack of maternal contribution results in Angelman syndrome. Allele-specific replication asynchrony has been detected at several loci surrounding the $GABA_A$ receptor β3 (GABRB3) and α5 (GABRA5) subunit gene cluster in chromosome 15q11-13 (Kitsberg et al., 1993; Knoll et al., 1994). These two genes, which are separated by 100 kb, are arranged in a head-to-head orientation (Sinnett et al., 1993). A domain of parental allele-specific replication has been localized to a 50-60 kb region between GABRA5 and GABRB3 by using λ phage probes to detect replication events by fluorescence *in situ* hybridization (FISH) (LaSalle and Lalande, 1995). For this small region, replication occurs at the beginning of S phase on the maternal chromosome 15, but is delayed until the end of S phase on the paternal homologue. In contrast, the genomic regions on either side of this maternal early replication domain exhibit the opposite pattern, with replication of the paternal loci preceding that of the maternal, but both events delayed until mid to late S phase (LaSalle and Lalande, 1995). In this paper, we report the characterization of the genomic DNA spanning the proximal boundary of the maternal early replication domain. A 13 kb region is identified that contains most DNA sequence motifs and structural features usually found at mammalian origins of replication (Benbow et al., 1992; Gale et al., 1992; DePamphilis, 1993; Caddle et al., 1990) including bent DNA, DNA unwinding elements, purine/pyrimidine tracts, scaffold associated regions and transcriptional factor binding sites. In addition, we found a small region associated with a very high autonomous replication activity *in vitro*, suggesting that the establishment and/or maintenance of the allele-specific replication domains in the $GABA_A$ receptor subunit gene cluster may be dependent on DNA replication origin activity.

RESULTS AND DISCUSSION

(a) Isolation and characterization of genomic clones

We have shown that GABRB3 and GABRA5 genes are arranged in a head-to-head configuration separated by approximately 100 kb (Sinnett et al., 1993). In the same paper, we reported a phage-contig assembly of 43 kb encompassing the 5' untranslated region

(UTR) of GABRB3 (delimited by λ60) and another of 21 kb corresponding to the 5' UTR of GABRA5 (defined by λ149). These two contigs did not overlap and were separated by a gap of approximately 60 kb (Sinnett et al., 1993). In order to obtain genomic clones that cover this chromosomal region, a large insert phage library enriched in sequences derived from chromosome 15 was initially screened with the oligodeoxynucleotide (5'-GCCATTTATGAAACCAG-3') as a probe. From this initial screening experiment, three phage genomic clones, λ84, λ86 and λ87, were isolated and characterized as described previously (Sinnett et al., 1993). The localization of these genomic clones was confirmed by FISH and long range restriction mapping (data not shown).

Single-copy probes (DNA fragments, PCR products and/or oligonucleotides) derived from these phage clones were used in subsequent screening experiments in order to extend the phage contig in both directions. This screening yielded three phage clones (λ327, λ324 and λ329) and one P1 clone (π205) and allowed the closure of the gap in the genomic DNA between GABRB3 and GABRA5. These genomic clones were analyzed by restriction mapping with a combination of KpnI, HindIII, BglII and several rare-cutting endonucleases (only the rare-cutting restriction sites are shown) and the resulting 140 kb contig is illustrated at the top of figure 1.

(b) DNA sequencing analysis

A region containing an allele-specific asynchronous DNA replication transition zone has been localized between the genes GABRB3 and GABRA5 in chromosome 15q11q13 (Knoll et al., 1994; LaSalle and Lalande, 1995). A maternal early replication domain spanning phage λ84 through λ149 has recently been identified (LaSalle and Lalande, 1995). The loci detected by these phages in the FISH replication assay display the predominant pattern where replication of the maternal locus occurs at the beginning of S phase, while that of the paternal locus is delayed until the end of S phase. The loci on either side of the maternal early domain, including the one detected by the λ87 and λ86 phage, display the opposite pattern of allele-specific replication timing, with the paternal locus replicating before the maternal. The transition between the imprinted replication domains occurs in the overlap between λ87/λ86 and λ84. For this reason, the genomic clone λ84 was further characterized in order to determine whether DNA sequences and/or structural features could be associated with the transition between the two imprinted domains.

The λ84 insert was digested with HindIII, and the restriction fragments were subcloned into pBluescript II SK+ and restriction mapped (fig. 1, bottom). Using the subclones p88, p163, p159, p160, p162 and p82 as sequencing templates, 13,022 nt of

contiguous DNA sequence from λ84 was generated. DNA sequence homology searching revealed the presence of several features usually found in eukaryotic origins of DNA replication (Table 1). In addition to 72 SAR elements, the 13 kb fragment contains 30 near (10/11) or perfect (11/11) matches to the yeast ARS elements (18 ARS-c and 12 ARS-p). Taking into account the overall A+T content (61%) of the 13kb region, the number of matches to both the SAR and ARS elements are approximately twice that expected by chance. Only a subset of the ARS elements could be expected to function as active sites of replication initiation in yeast, however, and to date no *in vivo* function has been assigned to yeast ARS matches in the mammalian genome.

Perfect matches to consensus binding sites for seven transcription factors were identified in the 13kb fragment (2 for Sp1, 3 for OTFI/NFIII, 2 for AP1, 6 for B1 and 7 for p53 consensus; see Table 1). Five close matches (14 out of 16 bp) to the PUR element were also observed. PUR elements are present near initiation zones for DNA replication in a variety of eukaryotes from yeast through humans (Bergemann and Johnson, 1992). The PUR protein shows a greater affinity for single-stranded DNA, suggesting a function as a helix-destabilizing protein (Bergemann and Johnson, 1992) in order to maintain an open DNA duplex. Sixty-two overlapping DNA polymerase start sites, encompassing an extensive polypyrimidine tract of 175 bp, are present between positions 700 and 900, and another cluster was found in plasmid p162 near the junction with p82. At the same junction, a cluster of three topoisomerase II binding site matches (13/15) were identified within only 100 nucleotides.

(c) Thermal energy analysis

DNA unwinding studies predict that the ARS activity should be associated with a DNA unwinding element (DUE) that facilitates helix unwinding (Huang and Kowalski, 1993). Computer assisted analysis of DNA helical stability has been applied to detect easily unwound sequences (Natale et al., 1993). DNA helical stability is the free energy difference (ΔG) between the double- and single-stranded states, and is based on the nucleotide sequence. The computer program Thermodyn has been designed to calculate the free energy needed for strand separation of a given DNA sequence based on experimentally-determined thermodynamic parameters (Natale et al., 1993). We used this computer program to generate helical stability profiles of the 13 kb DNA sequence described above. In figure 2, the free energy required to melt a given 100 bp (window size) sequence is graphed versus the nucleotide position at the center of that sequence. The relationship

between domains of local instability and the A+T content is also shown in figure 2. As would be predicted, the helical stability varies inversely with A-T content.

(d) Identification of bent DNA structure

Regions of bent DNA are highly conserved structural features of replication origins in both prokaryotes and eukaryotes (Caddle et al., 1990; Eckdahl and Anderson, 1990), and it has been suggested that they act as recognition sites for initiator proteins (Eckdahl and Anderson, 1990). To investigate the presence of such bent DNA elements within the putative replication origin, we tested a series of subclones spanning the 13 kb of DNA derived from the phage λ84 using a two-dimensional gel electrophoresis system. In this assay, the first dimensional separation is performed at room temperature in a tube agarose gel in order to fractionate the DNA fragments according to their length, while the second dimension of electrophoresis is run at 4°C in a slab polyacrylamide gel allowing the separation of DNA fragments according to both their shape and length. Under these conditions, bent DNA fragments show a low electrophoretic mobility in the second dimension gel as compared with the first, while non-bent DNA show a similar migration behavior in both dimensions (Anderson, 1986). Consequently, bent DNA fragments are observed above an arc formed by non-bent DNA molecular weight marker (123 bp ladder). Using this approach, fragments p159 and p82 showed an anomalous migration when compared to an internal control. Representative results of bent (p82) and non-bent (p162) DNA are shown in figure 3A. For the bent DNA analysis of p82, the HindIII plasmid insert was cleaved with EcoRI to generate fragments of 530 and 480 bp. Only the larger 530 bp subfragment contains bent DNA (figure 3A). The latter result was confirmed by using an alternative assay for bent DNA (von Kries et al., 1990) which involves comparing the mobility of DNA fragments in polyacrylamide gels in the presence or absence of ethidium bromide by one dimensional electrophoresis. Using this technique (von Kries et al., 1990), the 530bp HindIII/EcoRI fragment of p82 displays a significantly decreased relative mobility in polyacrylamide with ethidium bromide (compare lane 5 in figures 3B and 3C) and, thus, contains bent DNA. On the other hand, the relative mobility of the smaller 480bp fragment was not altered upon electrophoresis with or without ethidium bromide (figure 3B and 3C, lane 5). The results using the one dimensional assay also confirm that the p162 fragment does not contain bent DNA (figures 3B and 3C, lane 4) while the inserts of plasmids p180, p159 and p160 all display slight alterations in mobility (lanes 1, 2 and 3, figures 3B and 3C) suggesting that the fragments are bent to some degree. The two dimensional assay used in Figure 3A detected bent DNA in p159 but not

in p160 (data not shown) suggesting that the one dimensional assay may be slightly more sensitive for the larger (> 1kb) fragments).

(e) *In vitro* replication assay

Figure 4 shows the results of *in vitro* replication assay (Pearson et al., 1991; Zannis-Hadjopoulos et al., 1994) for autonomous replication activity of clones spanning the putative origin contained within the 13 kb region. Purified DNA product from a 1h *in vitro* replication reaction was analyzed by agarose gel electrophoresis, before and after digestion by DpnI, which cleaves only fully methylated input plasmid DNA. The hemimethylated product of one round of DNA replication that occurs in the in vitro system used here (Pearson et al., 1991) is resistant to DpnI cleavage. pX24, the 7.5 kb long plasmid containing oriβ, the DHFR origin of bidirectional replication (Burhans et al., 1990; Zannis-Hadjopoulos et al., 1994), used here as a positive control, shows strong incorporation of precursor nucleotide into forms II (relaxed circular) and III (linear) DNA product; while much of the DNA is sensitive to DpnI treatment, a resistant linear band is recovered, demonstrating a newly replicated product. 30.4, a 3.7 kb clone carrying a piece of human cDNA (Nielsen et al., 1994), also shows some incorporation of radioactive precursor which, however, is not due to its replication *in vitro*, since it is entirely DpnI-sensitive. Of the ten plasmid subclones (p164, p88, p163, p362, p363, p159, p160, p162, p82 and p161) obtained from the genomic clone λ84 that were tested, only p82 shows strong *in vitro* replication activity, as demonstrated by the presence (fig. 4) of strong DpnI-resistant bands (both forms II and III are visible). Indeed, analysis by quantitative densitometry showed that the fraction of DpnI-resistant material using p82 as a template was 250% of the amount obtained with pX24, suggesting that p82 may contain a highly potent origin of DNA replication. In addition to p82, very small amounts of DpnI-resistant form III bands were visible for p159 (<17% of pX24) and p362 (<14% of pX24); however, in the case of p362, no resistant material is obtained with this same region, when included in the larger context of p163. The p159 and p362 fragments, therefore, display a very low level of *in vitro* replication potential. These results also suggest that the differences in autonomous replication activity are not related to insert size but rather reflect some intrinsic property of the DNA fragment being tested.

(f) p82 replication *in vitro* initiates within the cloned genomic insert

Origin function in p82 implies not only the presence of a genetically determined origin control element -- termed a "genetic origin" (DePamphilis, 1993) or a "replicator" (Stillman, 1994) -- but also of an initiation site for DNA replication. One method for mapping the position of an initiation site is the determination of the earliest labeled DNA fragment, either *in vivo* (Vassilev and DePamphilis, 1992) or *in vitro* (Pearson et al., 1991). *In vitro* replication reactions are terminated at early time points and digested by appropriate enzymes, chosen to yield distinctly resolved fragments, which are then analyzed by autoradiography and densitometry for detecting the fragment with the highest specific activity of precursor incorporation. We have previously demonstrated that monkey and human origin-rich sequences placed in the context of plasmids pBR322 or pBluescript have been found to be preferentially labeled relative to the plasmid DNA, while the plasmids alone showed no preferential labeling *in vitro* (Pearson et al., 1991; Nielsen et al., 1994). Furthermore, analysis by EM of the initiation point of *ors8*, a mammalian origin-rich sequence, confirmed that it contained the site of initiation of DNA replication (Pearson et al., 1994).

Digestion of *in vitro* replicated p82 with the enzymes *Ban*I and *Bst*BI yields five fragments, as shown in figure 5. Among these, only the 464 and 792 bp fragments contain cloned genomic insert sequences (each includes an additional 36 and 201 bp of pBluescript vector, respectively); these are the two fragments which show the greatest radionucleotide incorporation density after *in vitro* replication periods of 4, 8, and 12 minutes. Taking into account the measurement error of densitometric quantification, the specific activities of the 464 and 792 bp fragments, expressed as relative incorporation per base pair, are equal to each other but significantly greater than those of the 389, 1231, and 1097 bp vector fragments. Thus, the genomic insert contains the preferred initiation site for replication in p82, which may be localized to a position relatively closer to the *Bst*BI cut site.

p82 is a putative origin of DNA replication

The above data indicate that a 13 kb region from the transition between two imprinted DNA replication domains (LaSalle and Lalande, 1995) shares many of the features of prokaryotic and eukaryotic origins of replication (Table 1). These sequence motifs are present at a higher frequency than can be explained by chance matching to any DNA sequence of equivalent A-T content (Table 1). Of particular interest is the p82 fragment, which displays helical instability and high A-T content (65%) (figures 2 and 6),

contains bent DNA (figure 3) and has potent replication activity *in vitro* (figure 4), including the earliest labeled fragment (figure 5). Another component of many prokaryotic and viral replication origins is the DNA cruciform, a potential secondary structure formed from inverted repeat elements. There is evidence suggesting that cruciforms may also have a role in the structure and function of mammalian origins (Zannis-Hadjopoulos et al., 1988). The GCG version of STEMLOOP was used to search for inverted repeat sequences with a minimum stem length of 6, a minimum number of bonds/stem of 12 and a maximum and minimum loop size of 20 and 3, respectively (Wu et al., 1993). The region between GABRB3 and GABRA5 contains several such small inverted repeat elements, including four from the p82 subclone with the potential to extrude as cruciforms containing 20 or more hydrogen bonds per stem (data not shown). Our interpretation of these results is that p82 may function as a putative origin of DNA replication in mammalian cells.

The activity of the putative origin in p82 must be profoundly influenced by the organization of the surrounding chromatin which, in this region, is a function of parental origin. Several features of the chromatin domain encompassing a putative origin could influence its activity. Chromatin domains are believed to be organized as loops which are attached to an A-T rich backbone by scaffold attachment regions (SARs) or matrix attachment regions (MARs) (Laemmli et al., 1992). The chromatin domains or loops, which are 5-100 kb in size (Laemmli et al., 1992), may serve as the fundamental units of replication and transcription in mammalian cells (Hassan and Cook, 1994). The major protein component of SARs is topoisomerase II (topo II) (Laemmli et al., 1992). SAR elements are specifically bound and contain multiple sites of cleavage by topoisomerase II (Sperry et al., 1989), thus creating targets for cellular DNA helicase activity. A good candidate region for SAR activity is the border between p162 and p82 (figure 6). In this region, a cluster of consensus sites for topoisomerase II has been identified that partly overlaps with a region of bent DNA. Clustering of topoisomerase II consensus binding sites has previously been observed in the SARs of several genes including the mouse immunoglobulin κ gene (Cockerill and Garrard, 1986), the Chinese Hamster DHFR gene (Käs and Chasin, 1987) and the human globin gene complex (Jarman and Higgs, 1988). It will be interesting to determine whether DNA fragments containing this cluster region display any *in vitro* or *in vivo* SAR activity and, hence, can play some role in regulating origin activity. Bent DNA is also a preferential substrate for topoisomerase II (Howard et al., 1991), suggesting that DNA bending regions may play a role in mediating attachment of DNA to the nuclear matrix. Stable attachment sites formed by bent DNA could provide the anchorage required for the introduction and maintenance of superhelical turns within

chromosomal loops or domains. An anchoring function would also serve to position replication origins, enhancers and promoters near the nuclear matrix. Both bent-DNA structures identified in this paper are associated with an autonomous replication activity, supporting the hypothesis that bent DNA structure may be a cis-acting element in eukaryotic replication origins. Intrinsically bent DNA elements are detected about once per 11 kb in random DNA fragments (Milot et al., 1992). This high frequency may suggest that juxtaposition of this structure and origins of replication is incidental rather than functional.

The potential for origin activity may, therefore, result from the overlap of features such as bent DNA, high A-T content and helical instability, rather than be associated with specific initiation sequences. This would be analogous in some respects to the so called "Jesuit model" (DePamphilis, 1993) where only a subset of all potential origins are used as initiation sites in the chromosome. In this model, the chromatin structure will evidently play a crucial role in selecting which origins are used in a given tissue and at a particular time in development. Further studies will be required to determine whether this origin is active *in vivo* and how chromosome parental origin is involved in modulating origin activity. For example, it is possible that the pattern of imprinting dictates the direction of the DNA replication fork either in the transcriptional orientation of GABRB3 (on the paternal chromosome) or in the transcriptional orientation of GABRA5 (on the maternal chromosome). This would imply that some origins are not bidirectional and their polarities are affected by transcription. In this regard, the human *c-myc* (Leffak and James, 1989) and avian globin (James and Leffak, 1986) genes are replicated in their transcriptional direction when they are transcriptionally active, whereas their quiescent germ-line counterparts are not. It is also possible that the maternal early replication domain, which includes the λ84 locus, could contain other genes. Such genes could display allele-specific expression and profoundly affect allele-specific replication timing. The identification of specific sites of initiation of DNA synthesis is crucial to future investigations of the mechanism resulting in the allele-specific replication asynchrony in this region of the genome.

CONCLUSIONS

1) A 13 kb region from an imprinted region of the human genome which displays DNA replication asynchrony contains several DNA sequence features commonly found at mammalian origins of DNA replication. These include stretches of A-T rich DNA which overlap with regions of significant DNA helical instability. There is also a clustering of yeast ARS elements, topoisomerase II consensus binding sites and transcription factor binding sites. In addition most plasmid subclones from the 13kb region contain bent DNA and inverted repeat elements with the potential to extrude as DNA crucifoms.

2) One plasmid subclone, p82, which is derived from an A-T rich region with highest potential helical instability, is the only plasmid clone tested which displays significant autonomous replication activity *in vitro*.

3) The suggestion that a putative origin of DNA replication is contained within p82 is further supported by the observation that replication *in vitro* initiates within the genomic insert. A replication origin within p82 could play an important role in establishing the allele-specific replication timing domains located within this imprinted region.

ACKNOWLEDGEMENTS

We thank Dr. D. Kowalski for providing the Thermodyn program, Dr. J. LaSalle for critical reading of the manuscript and Dixon Yun for expert assistance with the photography and computer graphics. This research was supported by grants from the Cancer Research Society, Inc. (G.B.P.), the Medical Research Council of Canada (M.Z.-H.), and NIH grant RO1NS30628 (ML). T.N. is a recipient of a MRC studentship and D.S. was supported by an MRC research fellowship. M.Z.-H. is an MRC Scientist and M.L. is an assistant investigator of the Howard Hughes Medical Institute.

REFERENCES

Anderson JN (1986) Detection, sequence patterns and function of unusual DNA structures. *Nucl. Acids Res.* 21:8513-8533

Angel P, Imagawa M, Chiu R, Stein B, Imbra RJ, Rahmsdorf HJ, Jonat C, Herrlich P, Karin M (1987) Phorbol ester-inducible genes contain a common *cis* element recognized by a PTA-modulated *trans*-acting factor. *Cell* 49:729-739

Benbow RM, Zhao J, Larson DD (1992) On the nature of origins of DNA replication in eukaryotes. *BioEssays* 14:661-670

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

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

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

Cockerill PN Garrard WT (1986) Chromosomal loop anchorage of the Kappa immunoglobulin gene occurs next to the enhancer in a region containing topoisomerase II sites. *Cell* 44:273-282

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

DePamphilis ML (1993) Origins of DNA replication in metazoan chromosomes. *J. Biol. Chem.* 268:1-4

Devereux J, Haeberli P, Smithies, O (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.* 12:387-395

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

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

Hassan AB, Cook PR (1994) Does transcription by RNA polymerase play a direct role in the initiation of replication ? *J. Cell Sci.* 107:1381-1387

Howard MT, Maxwell PL, Hsieh T, Griffith JD (1991) *Drosophila* topoisomerase II-DNA interactions are affected by DNA structure. *J. Mol. Biol.* 217:53-62

Huang RY, Kowalski D (1993) A DNA unwinding element and an ARS consensus comprise a replication origin within a yeast chromosome. *EMBO J.* 12:4521-4531

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

Jarman AP, Higgs DR (1988) Nuclear scaffold attachment sites in the human globin gene complexes. *EMBO J.* 7:3337-3344

Käs E, Chasin LA (1987) Anchorage of the Chinese Hamster dihydrofolate reductase gene to nuclear scaffold occurs in an intragenic region. *J. Mol. Biol.* 198:677-692

Kitsberg D, Selig S, Brandeis M, Simon I, Keshet I, Driscoll DJ, Nicholls RD, Cedar H (1993) Allele-specific replication timing of imprinted gene regions. *Nature* 364:459-463

Knoll JHM, Cheng SD, Lalande M (1994) Allele specificity of DNA replication timing in the Angelman/Prader-Willi syndrome imprinted chromosomal region. *Nature Genet.* 6:41-46

Laemmli UK, Käs E, Poljak L, Adachi Y (1992) Scaffold-associated regions: *cis* -acting determinants of chromatin structural loops and functional domains. *Curr. Opin. Genet. Devel.* 2 (1992) 275-285

LaSalle JM, Lalande M (1995) Domain organization of allele-specific replication within the *GABRB3* gene cluster requires a biparental chromosome 15q11-q13 contribution. *Nature Genet.* 9:386-394

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

Milot E, Belmaaza A, Wallenburg J, Gusew N, Bradley W.E.C., Chartrand P (1992) Chromosomal illegitimate recombination in mammalian cells is associated with intrinsically bent DNA elements. *EMBO J.* 11:5063-5070

Natale DA, Umek RM, Kowalski D (1993) Ease of DNA unwinding is a conserved property of yeast replication origins. *Nucl. Acids Res.* 21:555-560

Nielsen TO, Piatyszek MA, Shay JW, Pearson CE, Zannis-Hadjopoulos M, Price GB (1994) Autonomous replication activity of a human mitochondrial DNA sequence inserted into genomic DNA. *Int. J. Oncol.* 5:1003-1008

O'Neil EO, Kelly TJ (1988) Purification and characterization of nuclear factor III (origin recognition protein C) a sequence-specific DNA binding protein required for efficient initiation of adenovirus DNA replication. *J. Biol. Chem.* 263:931-937

Pearson CE, Frappier L, Zannis-Hadjopoulos M (1991) Plasmids bearing mammalian DNA-replication origin-enriched (*ors*) fragments initiate semiconservative replication in a cell-free system. *Biochim. Biophys. Acta* 1090:156-166

Pearson CE, Shihab-El-Deen A, Price GB, Zannis-Hadjopoulos M (1994b) Electron microscopic analysis of in vitro replication products of ors 8, a mammalian origin enriched sequence. *Somat. Cell Mol. Genet.* 20:147-152

Rao BS, Zannis-Hadjopoulos M, Price GB, Reitman M, Martin RG (1990) Sequence similarities among monkey *ori*-enriched (*ors*) fragments. *Gene* 87:233-242

Sander M, Hsieh T (1985) *Drosophila* topoisomerase II double-strand cleavage: analysis of DNA sequence homology at the cleavage site. *Nucl. Acids Res.* 13:1057-1072

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

Sinnett D, Wagstaff J, Glatt K, Woolf E, Kirkness EJ, Lalande M (1993) High-resolution mapping of the γ-aminobutyric acid receptor subunit β3 and α5 gene cluster on chromosome 15q11-q13, and localization of breakpoints in two Angelman syndrome patients. *Am. J. Hum. Genet.* 52:1216-1229

Sperry AO, Blasquez VC, Garrard WT (1989) Dysfunction of chromosomal loop attachment sites: illegitimate recombination linked to matrix association regions and topoisomerase II. *Proc. Natl. Acad. Sci. USA* 86:5497-5501

Stillman B (1994) Initiation of chromosomal DNA replication in eukaryotes. *J. Biol. Chem.* 269:7047-7050

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

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

von Kries JP, Phi-Van L, Diekmann S, Strätling, WH (1990) A non-curved chicken lysozyme 5' matrix attachment site is 3' followed by a strongly curved DNA sequence. *Nucl. Acids Res.* 18:3881-3885

Wu C, Friedlander P, Lamoureux C, Zannis-Hadjopoulos M, Price GB (1993) cDNA clones contain autonomous replication activity. *Biochim. Biophys. Acta* 1174:241-257

Zannis-Hadjopoulos M, Frappier L, Khoury M, Price GB (1988) Effect of anti-cruciform DNA monoclonal antibodies on DNA replication. *EMBO J.* 7:1837-1844

Zannis-Hadjopoulos M, Nielsen TO, Todd A, Price GB (1994) Autonomous replication *in vivo* and *in vitro* of clones spanning the region of the DHFR origin of bidirectional replication (ori β). *Gene* 151:273-277

TABLE 1

Region	position (nt)	ARS-c	ARS-p	SAR	PUR	TOP II	Pol start	(P,S,B,O,53)[a]
p88	1-1306	0	2	5	1	0	62[b]	(0,1,1,1,0)
p362	1307-4553	3	2	3	0	0	2	(0,0,3,1,1)
p363	4554-8509	7	2	23	2	2	7	(1,1,2,1,2)
p159	8510-9671	2	2	1	1	1	1	(0,0,0,0,0)
p160	9672-10820	3	3	14	0	0	1	(1,0,0,0,1)
p162	10821-11490	2	0	11	1	0	8	(0,0,0,0,0)
p82	11491-12502	1	1	8	0	4	0	(0,0,0,0,3)
	12503-13022	0	0	7	0	0	0	(0,0,0,0,0)
Total	1-13022	18	12	72	5	7	81	(2,2,6,3,7)

Table 1 footnote:

Sequence motifs associated with eukaryotic DNA replication origins. Computer analysis of the sequence was performed with either the Genetics Computer Group (GCG) software package (Devereux et al., 1984) or the HIBIO MacDNASIS Pro system (Hitachi). To determine the statistical significance of observed matches of short consensus elements to the genomic sequence, a program was written in BASIC that calculates the probable number of matches expected to occur if the same element was compared to a random double-stranded DNA sequence of identical size and A-T content (available from authors upon request), using the strategy and equations described previously (Rao et al., 1990). Only those elements present above the number expected by chance are included here. 1) Yeast ARS element: minimum 10 of 11 match to ARS-c, (*S. cerevisiae*) WTTTATRTTTW (Gale et al., 1992) and ARS-p (*S. pombe*), WRTTTATTTAW (Gale et al., 1992): 2) Scaffold attachment region (SAR): perfect 10 of 10 match to AATAAAYAAA, TTWTWTTWTT, WADAWAYAWW, TWWTDTTWWW (Benbow et al., 1992; Gale et al., 1992): 3) Purine-rich protein recognition element (PUR), minimum 14 of 16 match to GGNNGAGGGAGARRRR (Virta-Pearlman et al., 1993): 4) Topoisomerase II binding site (TOP II): minimum 13 of 15 match to GTNWAYATTNATNNR with a 6 of 6 match with the core WAYATT (Sander and Hsieh, 1985): 5) DNA polymerase start site (Pol start): minimum 16 of 17 match with YYYYYYYYYCTTTYYYY (Benbow et al., 1992): 6) Transcription factor binding sites [a]; AP1 (P), TGAGTCA (Angel et al., 1987); SP1 (S), CCGCCC (DePamphilis, 1988); B1 (B), AARRGGAA (Gale et al., 1992); OTFI/NFIII (O), ATTTGCAT (O'Neil and Kelly, 1988); p53 (p3), RRRCWWGYYY (Virta-Pearlman et al., 1993); [Codes for the consensus sequences are: W=A or T, R=A or G, Y=T or C, D=T, A or G, and N=A, C, G or T].

[b] most are contained in an extensive polypyrimidine tract of 172 bp

FIGURE 1. Chromosomal organization of the GABRB3/GABRA5 intergenic region. The 100 kb intergenic region that separates GABRB3 and GABRA5 was cloned and mapped. Phage genomic clones were obtained from the screening of a flow sorted chromosome 15 phage library as described previously (Sinnett et al., 1993). P1 phage clone, π205, was obtained from Genome Systems (St-Louis), by screening with PCR products derived from the λ84 insert. The Genome systems reference address for the P1 clone is plate 377, well A. The λ84 phage insert was digested with the restriction enzyme HindIII, and the fragments subcloned into pBluescript II SK+ (Stratagene). The 13 kb region that was analyzed by nucleotide sequencing is shown at the bottom. DNA sequence was generated by the dideoxy chain termination method (Sanger et al., 1977) using the CircumVent Thermal Cycle Dideoxy DNA Sequencing Kit (New England Biolabs). The sequence data were obtained either directly from phage clones or from plasmid subclones using universal sequencing primers and specific oligonucleotide primers. The p82 and p163 plasmids were sequenced by standard automated techniques using an Applied Biosystems instrument. The nucleotide sequence has been deposited in GenBank. (B, BssHII; Bgl, BglII; E, EagI; F, SfiI; H3, HindIII; k, KpnI; M, MluI; N, NotI; R, NruI; S, SacII; X, XhoI).

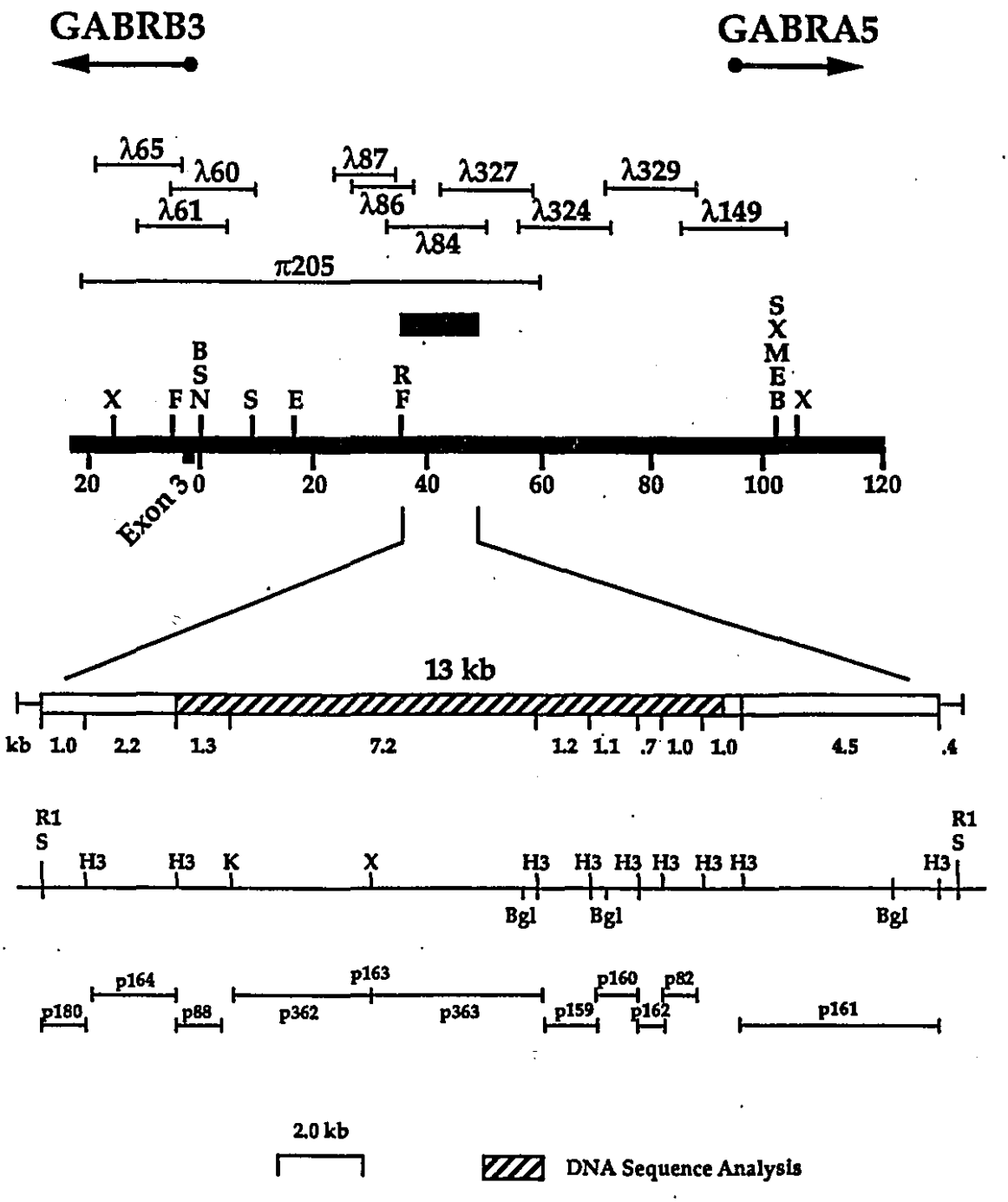

GABRB3
GABRA5
λ65
λ60
λ61
λ87
λ86
λ84
λ327
λ324
λ329
λ149
π205
X
B
S
F
N
S
E
R
F
S
X
M
E
B
X
20
0
20
40
60
80
100
120
Exon 3
13 kb
kb
1.0
2.2
1.3
7.2
1.2
1.1
.7
1.0
1.0
4.5
.4
R1
S
H3
H3
K
X
H3
H3
H3
H3
H3
H3
H3
R1
S
Bgl
Bgl
Bgl
p164
p180
p88
p163
p362
p363
p160
p159
p82
p162
p161
2.0 kb
DNA Sequence Analysis

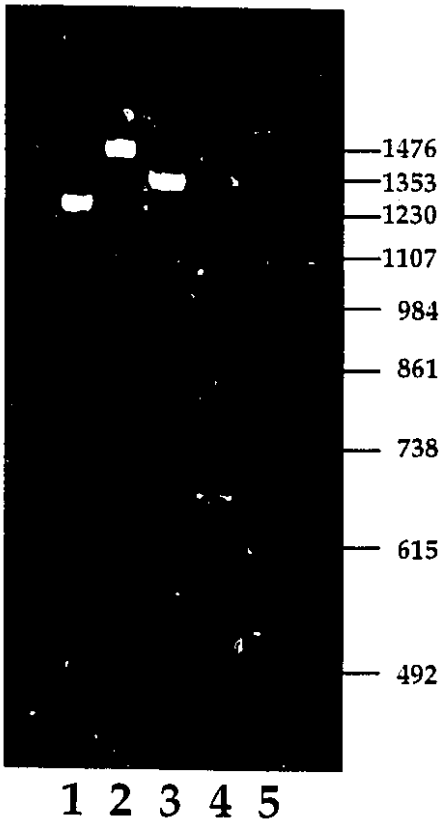
FIGURE 2. Helical stability for the 13 kb region derived from the GABRB3/GABRA5 intergenic region. In order to find potential DNA unwinding elements (DUEs) the computer program called Thermodyn (Natale et al., 1993) was used to calculate the free energy needed for DNA strand separation at 25°C in a 20 mM salt solution. The window (overlapping segments) size used in this study was 100 bp, with a two base pair increment (step=2). The output data were converted to graphic form using the INPLOT graphics program. Graphical analysis of A-T content was performed for a 100 bp window using the HIBIO MacDNASIS Pro system. DNA helical stability (ΔG) and A-T content are both plotted versus position.

Helical Stability ΔG
(kcal/ mol)
100
150
200
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
5500
6000
6500
7000
7500
8000
8500
9000
9500
10000
10500
11000
11500
12000
12500
13000
p88
p362
p363
p159
p160
p162
p82
Position (bp)
A-T Content
(%)
0
20
40
60
80
100

FIGURE 3A. Detection of bent DNA structure by two-dimensional gel analysis. Plasmid subclone inserts derived from λ84 were tested for migration anomalies by a two-dimensional gel technique (Anderson, 1986). Restriction digests of plasmid DNA subclones were fractionated in a 2.0% Nusieve:agarose (3:1, FMC inc) tube gel in 1X TBE buffer (89 mM Tris-borate, 2 mM EDTA, pH 8.2) at room temperature. The agarose tube gel was recovered and reoriented 90° relative to the first dimension on a 6% polyacrylamide (29:1, acrylamide:bisacrylamide) gel in 1X TBE. The second dimension gel was run at 4°C to enhance the DNA bending effects on fragment mobility. After electrophoresis, the DNA fragments were detected by staining the gel with 0.25 μg/ml ethidium bromide and photographed under UV light. Electrophoretic mobilities of the fragments analyzed were compared to the mobility of a 123 bp ladder molecular weight marker (BRL). The results obtained for the p82 (EcoRI-HindIII digest) and p162 (HindIII digest) in the presence of a 123 bp ladder internal standard (low-intensity fragments) are shown here. The arrow indicates the DNA fragment that migrates above the diagonal produced by the internal control, suggesting the presence of stably bent DNA elements. M, molecular weight marker (123 bp ladder).

p82 p162

FIGURES 3B AND 3C. Detection of bent DNA structure by one-dimensional gel analysis. All plasmid DNA samples were digested with HindIII with p180 and p82 also being cleaved with SalI and EcoRI, respectively. Samples were electrophoresed in a dual mini-vertical gel electrophoresis unit (Sigma-Aldrich model E4266) as described previously (von Kries et al., 1990) in 5% polyacrylamide (49:1, acrylamide:bisacrylamide), 2.5 X TBE buffer, 0.2% ammonium persulphate and 0.375% TEMED (N,N,N',N' tetramethylethylenediamine) with (B) or without (C) the addition of 1μg/ml EtBr in both the gel and running buffer. The EtBr-treated (B) samples were run at 7 V/cm at 4° for 14 hr while the untreated samples (C) were electrophoresed at 9 V/cm for 8 hr at room temperature. The 123 bp ladder was used as a molecular weight marker. The inserts of p180 (~1.0kb, lane 1), p159 (1161bp, lane 2), p160 (1148bp, lane 3) all display slightly reduced mobility in the presence of EtBr (B) relative to that in the absence of EtBr (C). The insert of p162 (669bp, lane 4) shows a slightly increased mobility in B relative to C. The smaller (480 bp) HindIII/EcoRI fragment of p82 (lane 5) displays the same mobility in B and C while the larger (530bp) fragment (lane 5) shows a larger apparent molecular weight when treated with EtBr (B). The 530 bp region of p82 thus contains bent DNA.


1476
1353
1230
1107
984
861
738
615
492
1 2 3 4 5

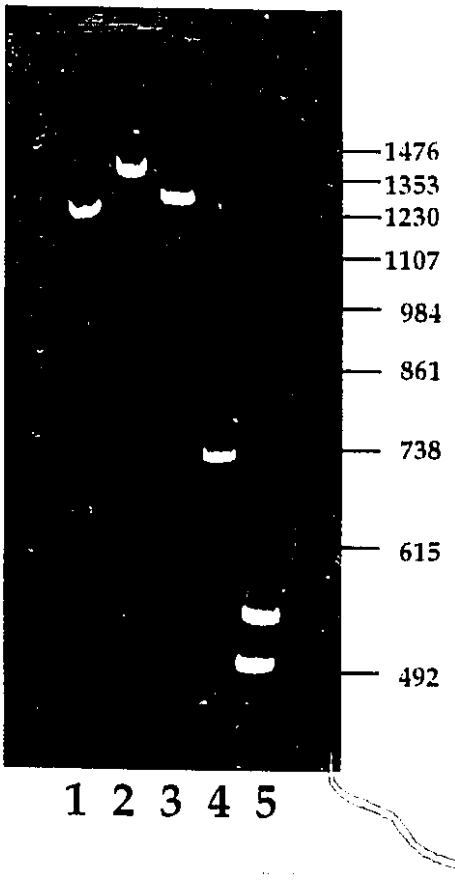

1476
1353
1230
1107
984
861
738
615
492
1 2 3 4 5

FIGURE 4. *In vitro* replication of clones from putative origin region between GABRB3 and GABRA5. pX24 (positive control), 30.4 (negative control), and a series of 10 plasmid subclones (cloned into pBluescript) spanning 19 kb of chromosome 15q11q13 were tested for autonomous replication by the *in vitro* replication assay. The positive control pX24 contains a 4.8 kb insert corresponding to the oriβ region of DHFR that has been described previously (Zannis-Hadjopoulos et al., 1994). The negative control, 30.4, is a pBluescript clone, randomly selected and containing a cDNA insert of 0.7 kb derived from human breast tumor. In the assay, equimolar quantities (23 fmol, for a final reaction concentration of 0.5 mM) of supercoiled plasmid templates were used in cell-free replication reactions (Pearson et al., 1991) composed of HeLa low salt cytoplasmic extract, HeLa nuclear extract, an ATP regenerating system, PEG and dNTP's including [α^{32}P]-dCTP and [α^{32}P]-TTP. One-half of the purified DNA product was digested with 1U DpnI, 90 min at 37°C, in the presence of 200ng λ DNA as a carrier. Uncut and digested samples were electrophoresed 16 h in an 1% agarose gel in 1X TAE buffer. To control the completeness and methylation specificity of DpnI digestion, 500 ng aliquots of λ DNA derived from both dam$^+$ and dam$^-$ bacteria were digested under identical conditions and monitored by agarose gel electrophoresis followed by ethidium bromide staining. The gel was then dried and exposed directly to a Fuji BAS-III imaging plate, and developed using the Fuji BAS 2000 phosphoimager Image Analysis program. Product DNA in the lanes on the left half of the figure shows the total radioactivity incorporation for each of the plasmids tested. DpnI-digested DNA product on the right half of the figure indicates replicated (DpnI-resistant) plasmid, together with fragments resulting from the cutting of DpnI-sensitive DNA. The positions of relaxed circular (form II) and linear (form III) DNA for p82 (4.1 kb) are shown on the right side of the figure. The total sizes of other plasmids tested are as follows: 30.4=3.7 kb, p164=5.2 kb, p88=4.3 kb, p163=10.5 kb, p362=6.2 kb, p363=7.0 kb, p159=4.2 kb, p160=4.1 kb, p162-3.7 kb, and p161=7.5 kb.

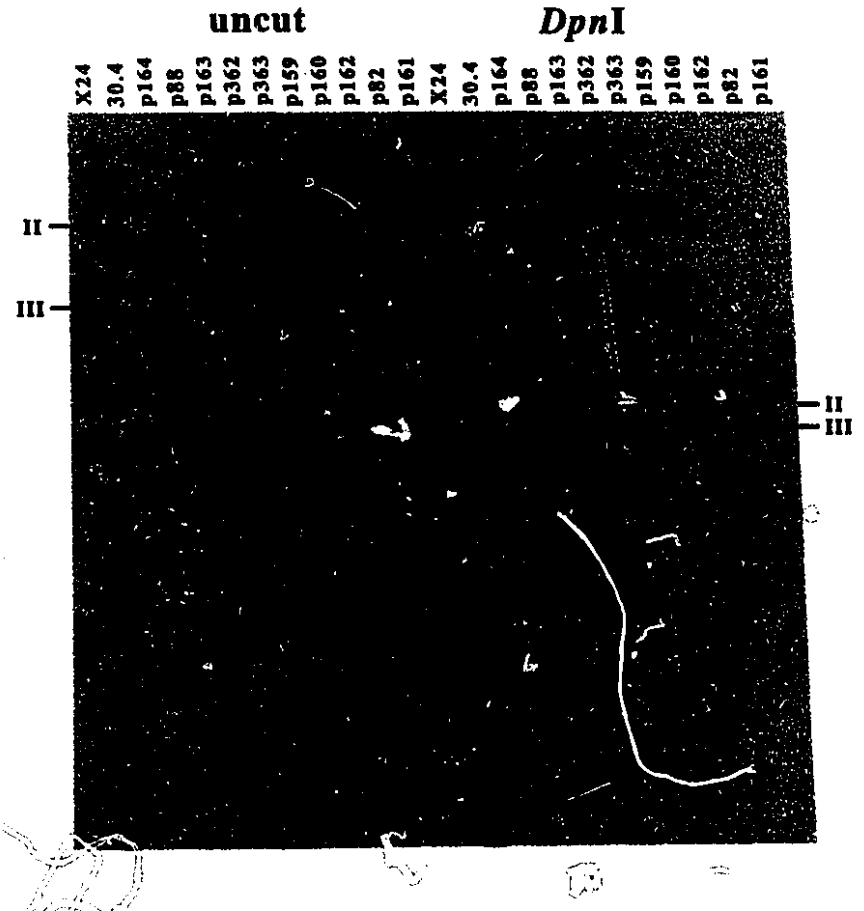

uncut
DpnI
X24
30.4
p164
p88
p163
p362
p363
p159
p160
p162
p82
p161
X24
30.4
p164
p88
p163
p362
p363
p159
p160
p162
p82
p161
II
III
II
III

FIGURE 5. Determination of the initiation site of p82 *in vitro* by earliest labeled fragment methodology. Triplicate *in vitro* replication reactions using p82 template were performed as described in figure 4, except that reactions were prematurely terminated with one volume of 1% SDS, 30 mM EDTA stop mix after 4, 8, and 12 minute incubation periods. The purified product was digested with 30 U *Ban*I and 20 U *Bst*BI in NEB buffer 4 (New England Biolabs), 100 minutes at 37°C and then 25 minutes at 65°C (the optimal temperature for BstBI). The resulting fragments and their positions relative to the pBluescript vector and genomic insert regions of p82 are indicated below the graph. Digested samples were separated by electrophoresis in 1.6 % agarose, 1 x TAE, which was subsequently dried and used to expose a phosphoimager plate. Densitometry was performed using the Fuji BAS 2000 image analysis software. Plotted values for nucleotide incorporation were obtained as follows: each band was quantified, and the background signal present in an equal area in the immediately adjacent portion of the same lane was subtracted. The resulting score was divided by the size of the fragment in question to calculate the radionucleotide incorporation per base pair, and the highest value so obtained was normalized to 100%. Error in the plotted values ranges from 1% at the earliest time points to a maximum of 7% for the smallest fragment at the latest time point. Incorporation was carried out for 4' (black bars), 8' (crosshatched bars), or 12' (open white bar). In the linearized plasmid restriction map below the graph, pBluescript vector sequences are indicated by the thin crosshatched line.

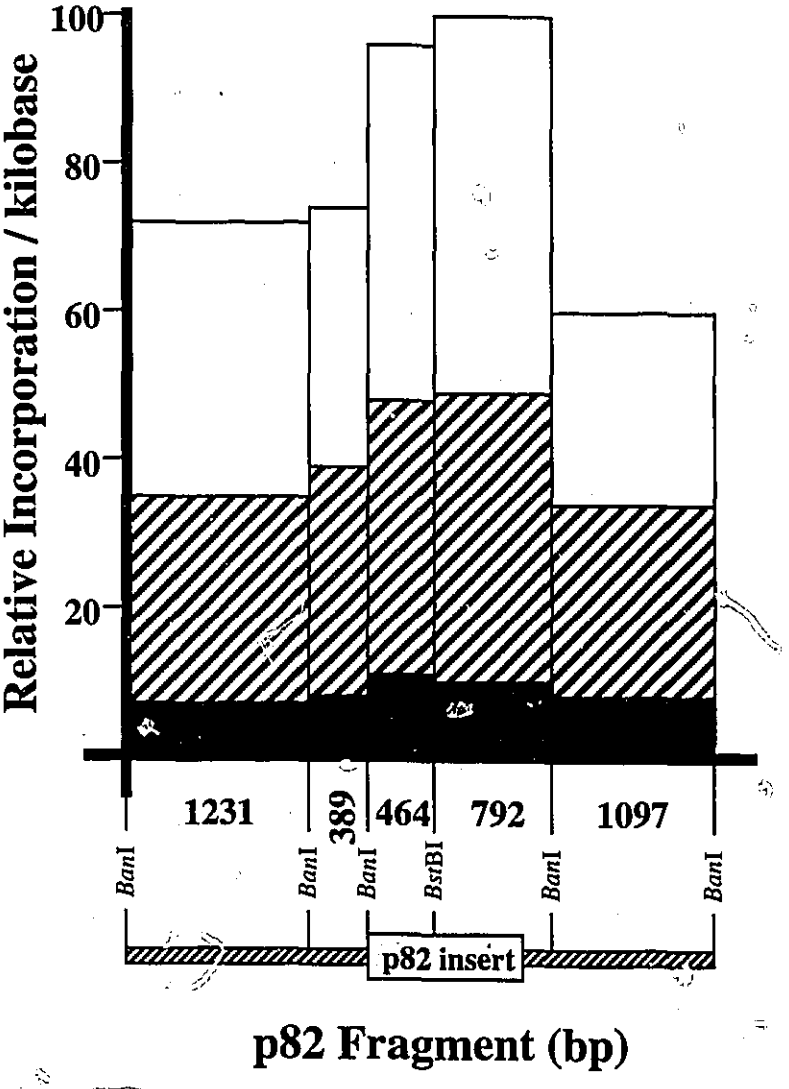
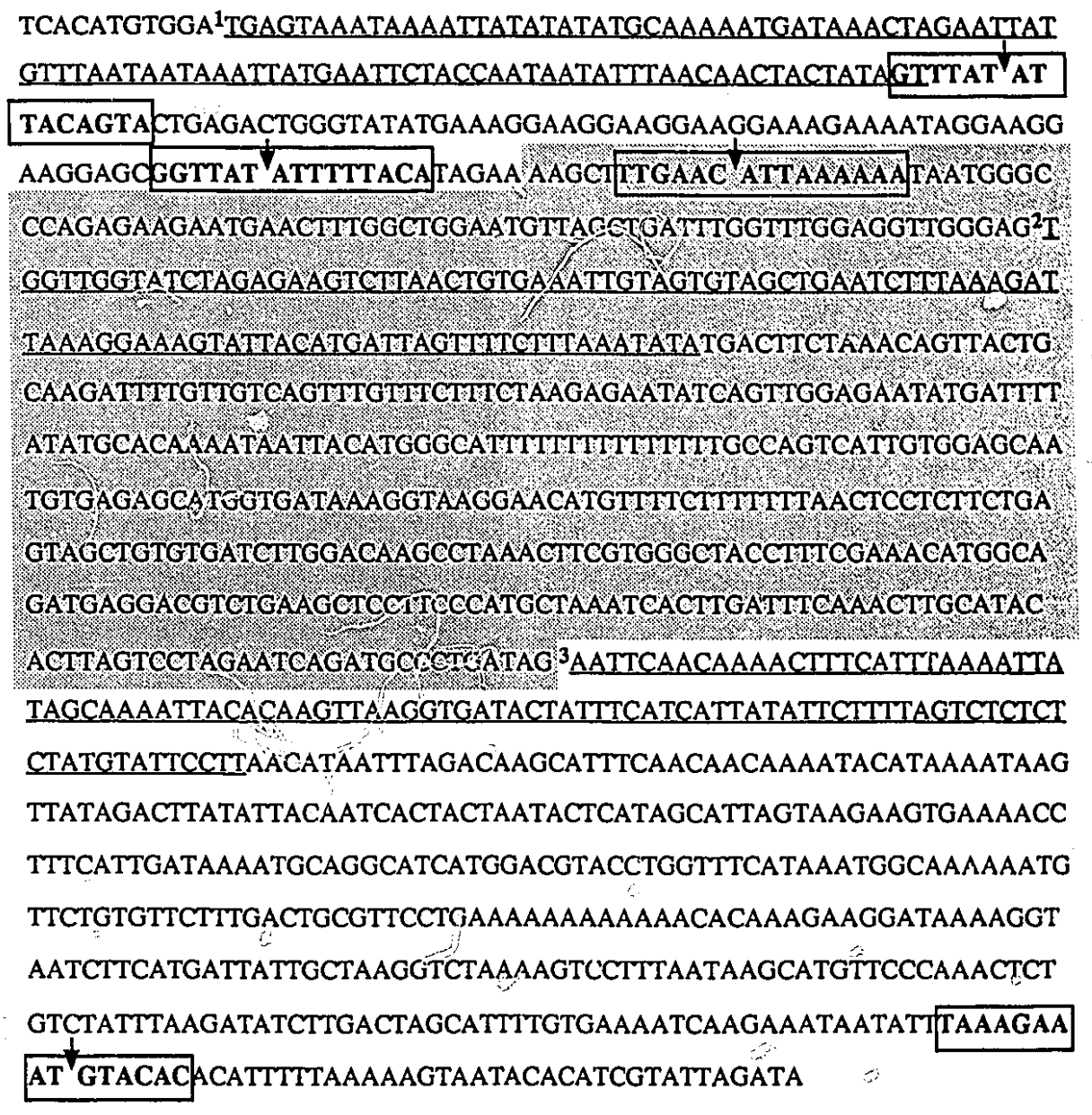

Relative Incorporation / kilobase
100
80
60
40
20
1231
389
464
792
1097
BanI
BanI
BanI
BstBI
BanI
BanI
p82 insert
p82 Fragment (bp)

FIGURE 6. Nucleotide sequence of the p82/p162 junction and the p82 fragment. The overlap (13/15 nucleotides) with the consensus sequence for topoisomerase II is indicated by a box. In all three cases, there is a perfect match (6/6) with the topoisomerase II core (cleavage) sequence (WAY▼ATT). Three potential DNA unwinding elements (DUEs) of 100 nt in length are underlined. The OLIGO 4.0 (National Biosciences) software package was used to identify these potential DUEs. For DUEs 1, 2 and 3, the melting temperatures are, respectively 12°, 9° and 9° lower than the average value (60°) for the entire 13 kb region. The %A-T content are 84, 73 and 76 for DUEs 1, 2 and 3, respectively. The shaded sequence is the portion of p82 containing bent DNA (figure 3).


TCACATGTGGA[1]TGAGTAAATAAAATTATATATATGCAAAAATGATAAACTAGAATTAT
GTTTAATAATAAATTATGAATTCTACCAATAATATTTAACAACTACTATAGTTTATAT
TACAGTACTGAGACTGGGTATATGAAAGGAAGGAAGGAAGGAAAGAAAATAGGAAGG
AAGGAGCGGTTATATTTTTACATAGAAAAGCTTTGAACATTAAAAAAATAATGGGC
CCAGAGAAGAATGAACTTTGGCTGGAATGTTACCTGATTTGGTTTGGAGGTTGGGAG[2]T
GGTTGGTATCTAGAGAAGTCTTAACTGTGAAATTGTAGTGTAGCTGAATCTTTAAAGAT
TAAAGGAAAGTATTACATGATTAGTTTTCTTTAAATATATGACTTCTAAACAGTTACTG
CAAGATTTTGTTGTCAGTTTGTTTCTTTCTAAGAGAATATCAGTTGGAGAATATGATTTT
ATATGCACAAAATAATTACATGGGCATTTTTTTTTTTTTTTGCCAGTCATTGTGGAGCAA
TGTGAGAGCATGGTGATAAAGGTAAGGAACATGTTTTCTTTTTTTAACTCCTCTTCTGA
GTAGCTGTGTGATCTTGGACAAGCCTAAACTTCGTGGGCTACCTTTCGAAACATGGCA
GATGAGGACGTCTGAAGCTCCTTCCCATGCTAAATCACTTGATTTCAAACTTGCATAC
ACTTAGTCCTAGAATCAGATGCCCTGATAG[3]AATTCAACAAAACTTTCATTTAAAATTA
TAGCAAAATTACACAAGTTAAGGTGATACTATTTCATCATTATATTCTTTTAGTCTCTCT
CTATGTATTCCTTAACATAATTTAGACAAGCATTTCAACAACAAAATACATAAAATAAG
TTATAGACTTATATTACAATCACTACTAATACTCATAGCATTAGTAAGAAGTGAAAACC
TTTCATTGATAAAATGCAGGCATCATGGACGTACCTGGTTTCATAAATGGCAAAAAATG
TTCTGTGTTCTTTGACTGCGTTCCTGAAAAAAAAAAAACACAAAGAAGGATAAAAGGT
AATCTTCATGATTATTGCTAAGGTCTAAAAGTCCTTTAATAAGCATGTTCCCAAACTCT
GTCTATTTAAGATATCTTGACTAGCATTTTGTGAAAATCAAGAAATAATATTTAAAGAA
ATGTACACACATTTTTAAAAAGTAATACACATCGTATTAGATA

ADDENDUM TO CHAPTER SIX

A) The PROB2.BAS program

I wrote this program in GW-BASIC, but it could easily be converted to other forms of BASIC or to another programming language; it is based on the methods described in Rao et al. (1990). The purpose of PROB2.BAS is to help in the interpretation of the importance of sequence analysis results, where short functional elements (eg. CCGCCC, the SP-1 transcription factor binding site) are matched to a genomic sequence, by determining the statistical significance of these observed matches. This is accomplished by calculating the number of matches such a consensus element would be expected to show when compared to a strictly random sequence of the same length as that used during the matching search. The user is prompted for this genomic sequence length, as well as the sequence's AT-richness. The AT level can be set at 50% to compare with totally random sequences of ATCG, or alternatively can be set to the AT level of organism's genome, or, as we usually do, to the AT% of the genomic sequence used for the matching procedure in the first place (this then takes into account the overall AT richness as an explanation of observed matches to highly AT-biased elements, like the *S. cerevisiae* ARS consensus WTTTATRTTTW).

The consensus element to be tested for expected random matches must be input in capitals (this could be easily changed in the program, but I haven't bothered; it's simpler just to hit CapsLock), and accepts all standard designations: A, T, C, G; W (A or T), S (G or C), Y (T or C), R (A or G), M (A or C), K (G or T); B (not A), D (not C), H (not G), V (not T); N or X (anybase). The most recently-added feature of the program allows it to take into account possible mismatches of 1 or 2 bp to the consensus sequence. It is assumed that the element can be matched independently of orientation -- that it could exist on either strand. If this assumption is wrong, answers must be divided by 2 and standard deviations by the square root of 2.

The program converts the input consensus sequence into individual bases, then assigns a probability for each base to occur considering the overall AT%. It then multiplies these probabilities to get the chance of the sequence of bases occurring in order, the overall probability of a consensus match in a random sequence of the same length as the consensus element. It then determines the number of opportunities to find a new consensus element sequence in a test genomic sequence of the input length (i.e. in 100 base pairs of linear DS DNA, there are 200 separate 1 bp sequences that could be checked for a match to a 1 bp

consensus, but only four 99 bp sequences). The two values are multiplied to give the number of expected matches, by random chance, for that element in a sequence of random DNA of defined length and AT%. If mismatches are allowed, the chance of perfect matches is also displayed.

The standard deviation of this value, giving the range of the number of matches that are expected by chance, is calculated from probability theory. To convert these values to 95% confidence intervals, it is approximately correct to multiply by two. Since DNA contains an integral number of bases, the "maximum number of matches expected by chance, at 95% confidence level" is rounded down to a whole number.

Limitations: uncertain gap lengths in the middle of the tested consensus sequence can only be taken into account by repeat analysis including different numbers of NNN's in the midst of each form of the consensus. To determine the total probability of any of a family of similar elements occurring (eg. any Scaffold Attachment Region, for which four consensus elements exist), add the expected number values, but note the P value for each single match as well, because the sum standard deviation must be calculated by hand as SD = Square root of (sum of expected numbers * (1 - sum of probabilities of single matches by each element)). The program assumes linear DNA; circular DNA has more possible matches, although the error is not significant if the length of the consensus element << the total sequence length. Finally, the program calculates each strand of DNA separately, not assuming complementarity (i.e. if the first strand randomly had a highly purine-rich element, the second must contain a pyrimidine-rich element). This only causes problems if the genomic sequence being searched for match elements is very short (10 bp); most genomic sequences will be hundreds or thousands of bp long and no error will be introduced.

LISTING OF THE PROB2.BAS PROGRAM

```
2 REM **  THE "PROB2.BAS" DNA SEQUENCE ANALYSIS PROGRAM, BY TORSTEN NIELSEN,
  MD/PhD PROGRAM, McGILL CANCER CENTRE, McGILL UNIVERSITY, MONTREAL, CANADA.
3 REM ** last modified 94.11.20
6 REM ** this program calculates the number of times that an input short
consensus element will be expected to match purely by random chance to a longer
sequence fragment of given AT richness
7 REM ** this modified upgraded deluxe version also can get that probability for
 consensus elements allowing one mismatch
8 REM ** assumes that sequence can be matched independently of orientation - ie
looks on BOTH strands.  If 5' --> 3' orientation of element is fixed, divide all
 match values by 2 and SD by SQRT(2).
9 REM ** Also does not assume complementarity, so inaccurate for short consensus
 elements in short fragments esp. if AT% close to 0 or 100.
10 PRINT : PRINT : PRINT
20 INPUT "Length of entire clone sequence in bp:"; L
30 INPUT "A + T content in %:"; AT: AT = AT / 100
40 INPUT "Type the matched consensus sequence, in capitals:"; CO$: CO = LEN(CO$)
: DIM C$(CO): DIM P(CO):DIM XP(CO):DIM XXP(CO):PRINT:TP=0
42 MI=0:INPUT"# mismatches allowed (max 2; default = 0):";MI
50 GOSUB 1000
60 GOSUB 2000
70 PRINT "comparisons = ("; L; "bases/strand - ("; CO; "bases/window - 1))(2 str
ands) ="; (L-(CO-1))*2
80 NP = (L - (CO - 1))*2*SP
90 PRINT "EXPECTED NUMBER OF PERFECT MATCHES: "; NP
92 SD=(NP*(1-SP))^.5
94 PRINT"STANDARD DEVIATION: ";SD:PRINT"Max. # perfect matches expected by rando
m chance at 95% confidence =";INT(NP+1.96*SD):PRINT
96 IF MI>0 GOTO 3500
100 INPUT "Type (1) to try a new consensus against the same genomic sequence,
               (2) to test a new genomic sequence,
            or (3) to quit"; CHOICE$: PRINT : PRINT
110 IF CHOICE$ = "1" THEN ERASE C$: ERASE P:ERASE XP:ERASE XXP: GOTO 40
120 IF CHOICE$ = "2" THEN ERASE C$: ERASE P:ERASE XP:ERASE XXP: GOTO 10
130 IF CHOICE$ = "3" THEN END
140 GOTO 100
1000 REM **CONSENSUS SEQUENCE ANALYZER**
1010 REM ** converts input sequence to individual bases
1030 FOR N = 1 TO CO
1040 C$(N) = MID$(CO$, N, 1)
1050 NEXT N
1060 RETURN
2000 REM **MATCH IN A SINGLE COMPARISON**
2005 REM ** assigns probability to each base, and to string of those bases
2010 FOR N = 1 TO CO
2020 IF C$(N) = "G" OR C$(N) = "C" THEN P(N) = (1 - AT) / 2
2030 IF C$(N) = "A" OR C$(N) = "T" THEN P(N) = AT / 2
2040 IF C$(N) = "W" THEN P(N) = AT
2050 IF C$(N) = "S" THEN P(N) = 1 - AT
2060 IF C$(N) = "R" OR C$(N) = "Y" OR C$(N) = "M" OR C$(N) = "K" THEN P(N) = .5
2070 IF C$(N) = "D" OR C$(N) = "H" THEN P(N) = 1 - ((1 - AT) / 2)
2080 IF C$(N) = "V" OR C$(N) = "B" THEN P(N) = 1 - AT / 2
2090 IF C$(N) = "N" OR C$(N) = "X" THEN P(N) = 1
2100 PRINT "BASE "; N; " = "; C$(N); ": P = "; P(N)
2110 NEXT N
2120 SP = 1
2130 FOR N = 1 TO CO: SP = SP * P(N): NEXT N
2140 PRINT "Sum probability for perfect match in single comparison: "; SP
2150 IF MI>0 THEN GOSUB 3000
2160 RETURN
```

```
3000 REM ** PROBABILITY ALLOWING ONE MISMATCH **
3005 REM ** algorithm to recalculate allowing one mismatch by adding probability
 for match of (COnsensus element length) sequences where probability for each
 base in turn is made up to 1.
3010 TP = SP
3020 FOR N = 1 TO CO
3030 P(N) = 1 - P(N)
3040 XP(N) = 1
3050 FOR NN = 1 TO CO
3060 XP(N) = XP(N) * P(NN)
3070 NEXT NN
3080 PRINT"Additional probability allowing mismatch of ";C$(N);" in position";N;
"=";XP(N)
3100 TP = TP + XP(N)
3110 IF MI=2 THEN GOSUB 4000
3115 P(N) = 1 - P(N)
3120 NEXT N
3130 RETURN
3500 REM ** ANSWERS WITH ONE MISMATCH **
3505 IF MI=2 THEN 4500
3510 PRINT"Total probability for perfect match & all one-mismatch possibilities
=";TP
3520 PRINT "comparisons = ("; L; "bases/strand - ("; CO; "bases/window - 1))(2 s
trands) ="; (L-(CO-1))*2
3530 NP = (L - (CO - 1))*2*TP
3540 PRINT "EXPECTED NUMBER OF AT LEAST";CO-1;"/";CO;"bp MATCHES: "; NP
3550 SD=(NP*(1-TP))^.5
3560 PRINT"STANDARD DEVIATION: ";SD:PRINT"Max. # matches expected by random chan
ce at 95% confidence =";INT(NP+1.96*SD):PRINT
3570 GOTO 100
4000 REM ** TWO MISMATCH CALCULATOR **
4010 IF N = CO THEN RETURN
4020 TTP = 0
4030 FOR TM = (N+1) TO CO
4040 P(TM) = 1 - P(TM)
4050 XXP(TM) = 1
4060 FOR TTM = 1 TO CO
4070 XXP(TM)=XXP(TM)*P(TTM)
4080 NEXT TTM
4090 PRINT" extra prob. if base";N;"(";C$(N);") & base";TM;"(";C$(TM);") mismat
ched =";XXP(TM)
4100 P(TM) = 1 - P(TM)
4110 TTP = TTP + XXP(TM)
4120 NEXT TM
4130 TP = TP + TTP
4140 RETURN
4500 REM ** ANSWERS WITH TWO MISMATCHES **
4510 PRINT"Total probability for perfect match & all one and two mismatch possib
ilities =";TP
4520 PRINT "comparisons = ("; L; "bases/strand - ("; CO; "bases/window - 1))(2 s
trands) ="; (L-(CO-1))*2
4530 NP = (L - (CO - 1))*2*TP
4540 PRINT "EXPECTED NUMBER OF AT LEAST";CO-2;"/";CO;"bp MATCHES: "; NP
4550 SD=(NP*(1-TP))^.5
4560 PRINT"STANDARD DEVIATION: ";SD:PRINT"Max. # matches expected by random chan
ce at 95% confidence =";INT(NP+1.96*SD):PRINT
4570 GOTO 100
```

B) Detailed results of statistical analysis

The following three pages show the results of the application of the PROB2.BAS program to the sequences described in the manuscript and to the observed numbers of short consensus elements, presented in Table 1. The significance of observed elements across the whole sequenced region was determined, as was the significance of elements occurring specifically in the autonomously replicating fragment, p82, and in the p82/p162 junction region presented in Figure 6.

While not mentioned in Table 1, the CP-1 CCAAT transcription factor consensus was present at levels significantly higher than expected by chance. The OTFI/NFIII transcription factor binding site, B1 intronic element and PUR factor consensus were significantly above the 95% confidence limit, while the AP1 and SP1 were also enriched above the expected number close to the 95% confidence limit. None of these matches occurred in the p82 region itself, but as discussed in Chapter 1, auxiliary replication factors need not lie within the minimal origin itself. p53 binding site occurrences are statistically significant in the p82 fragment itself. ARS, SAR, and polymerase start site matches are concentrated in regions flanking p82; the immediately neighbouring p162 region carries statistically-significant numbers of matches to the latter two elements. The telomere-like consensus matches can be explained by random chance and were not discussed in the manuscript.

Consensus elements present above the 95% confidence limit of the number expected by chance are suspected to serve some function *in situ*, although not necessarily in DNA replication.

	SUMMARY OF STATISTICAL ANALYSIS: LIKELIHOOD OF MATCHES OF TEST					
	ELEMENTS TO RANDOM SEQUENCE OF INDICATED SIZE AND A-T CONTENT					
	sheet 1 of 3: analysis of complete sequenced subregion of 15q11q13					
			matches to 13,108 base pairs, 61% AT			
		allowed # of	(comparison to total sequenced region)			
NAME	sequence	mismatches	expected #	Std. Dev.	95% CI max	observed #
CP-1 CAAT	SACCAATCAGG	0	0.008	0.087	0	3
OTFI/NFIII	ATTTGCAT	0	0.8	0.9	2	**3**
AP1	TGAGTCA	0	1.7	1.3	4	2
AP2	GCCTGGGG	0	0.09	0.29	0	
AP5	TGTGGAATG	0	0.1	0.32	0	
B1	AARRGGAA	0	2.2	1.5	5	**6**
SP1	CCGCCC	0	1.4	1.2	3	2
ARS-c	WTTTATRTTTW	0	0.36	0.6	1	**2**
		1	7.8	2.8	13	**18**
ARS-p	WRTTTATTTAW	0	0.36	0.6	1	0
		1	7.8	2.8	13	**12**
ARS-c or ARS-p		0	0.73	0.85	2	2
		1	16	4	24	**30**
SARs	AATAAAYAAA	0	0.3	0.55	1	1
		1	6.7	2.6	11	
	TTWTWTTWTT	0	1.5	1.2	3	**5**
		1	28	5	38	
	WADAWAYAWW	0	13	4	19	**23**
		1	180	10	200	
	TWWTDTTWWW	0	15	4	23	**43**
		1	210	10	230	
any of the four SARs		0	30	5	40	**72**
		1	420	20	460	367
PUR	GGNNGAGGGAGARRRR	2	0.38	0.62	1	**5**
topo II	GTNWAYATTNATNNR	0	0.19	0.44	1	
		1	4.5	2.1	8	
		2	49	7	63	
topo II core	WAYATT	0	69	8	85	
pol start (s)	YYYYYYYYCTTTYYYY	0	0.02	0.13	0	
		1	0.44	0.66	1	
pol start (w)	YYYYYYRCCCYYYYY	0	0.02	0.15	0	
		1	0.63	0.79	2	
either polymerase start site		1	1.1	1	3	**30**
telo-like	TGGGT	0	18	4	26	22
p53	RRRCWWGYYY	0	5.7	2.4	10	7

	SUMMARY OF STATISTICAL ANALYSIS: LIKELIHOOD OF MATCHES OF TEST					
	ELEMENTS TO RANDOM SEQUENCE OF INDICATED SIZE AND A-T CONTENT					
	sheet 2 of 3: analysis of p82 (the autonomously replicating fragment)					
			matches to 1012 base pairs, 68% AT			
		allowed # of	(comparison to p82 only)			
NAME	sequence	mismatches	expected #	Std. Dev.	95% CI max	observed #
CP-1 CAAT	SACCAATCAGG	0	0	0.017	0	
OTFI/NFIII	ATTTGCAT	0	0.08	0.28	0	0
AP1	TGAGTCA	0	0.11	0.33	0	0
AP2	GCCTGGGG	0	0.002	0.043	0	
AP5	TGTGGAATG	0	0.006	0.077	0	
B1	AARRGGAA	0	0.17	0.41	0	0
SP1	CCGCCC	0	0.03	0.18	0	0
ARS-c	WTTTATRTTTW	0	0.08	0.29	0	0
		1	1.5	1.2	3	1
ARS-p	WRTTTATTTAW	0	0.08	0.29	0	0
		1	1.5	1.2	3	1
ARS-c or ARS-p		0	0.17	0.41	0	0
		1	3.1	1.7	6	2
SARs	AATAAAYAAA	0	0.06	0.25	0	0
		1	1.2	1.1	3	
	TTWTWTTWTT	0	0.33	0.58	1	0
		1	5.3	2.3	9	
	WADAWAYAWW	0	2.4	1.6	5	4
		1	28	5	38	
	TWWTDTTWWW	0	3.3	1.8	6	4
		1	37	6	49	
any of the four SARs		0	6.1	2.5	10	8
		1				
PUR	GGNNGAGGGAGARRRR	2	0.01	0.12	0	0
topo II	GTNWAYATTNATNNR	0	0.03	0.17	0	
		1	0.64	0.8	2	
		2	6.3	2.5	11	
topo II core	WAYATT	0	9.2	3	15	
pol start (s)	YYYYYYYYCTTTYYYY	0	0.002	0.039	0	
		1	0.04	0.2	0	
pol start (w)	YYYYYYRCCCYYYYY	0	0.001	0.032	0	
		1	0.03	0.17	0	0
either polymerase start site		1			0	
telo-like	TGGGT	0	0.95	0.98	2	0
p53	RRRCWWGYYY	0	0.37	0.61	1	**3**

	SUMMARY OF STATISTICAL ANALYSIS: LIKELIHOOD OF MATCHES OF TEST						
	ELEMENTS TO RANDOM SEQUENCE OF INDICATED SIZE AND A-T CONTENT						
	sheet 3 of 3: analysis of p82/p162 junction region, as shown in Fig. 6						
				matches to 1218 base pairs, 69% AT			
		allowed # of		(comparison to p82/p162 junction, fig. 6)			
NAME	sequence	mismatches		expected #	Std. Dev.	95% CI max	observed #
CP-1 CAAT	SACCAATCAGG	0		0	0.018	0	
OTFI/NFIII	ATTTGCAT	0		0.1	0.31	0	0
AP1	TGAGTCA	0		0.13	0.36	0	0
AP2	GCCTGGGG	0		0.002	0.042	0	
AP5	TGTGGAATG	0		0.007	0.083	0	
B1	AARRGGAA	0		0.21	0.45	1	0
SP1	CCGCCC	0		0.03	0.18	0	0
ARS-c	WTTTATRTTTW	0		0.12	0.34	0	0
		1		2.1	1.4	4	2
ARS-p	WRTTTATTTAW	0		0.12	0.34	0	0
		1		2.1	1.4	4	1
ARS-c or ARS-p		0		0.23	0.48	1	0
		1		4.2	2	8	3
SARs	AATAAAYAAA	0		0.08	0.29	0	0
		1		1.6	1.3	4	
	TTWTWTTWTT	0		0.46	0.68	1	1
		1		7.2	2.7	12	
	WADAWAYAWW	0		3.3	1.8	6	6
		1		38	6	50	
	TWWTDTTWWW	0		4.5	2.1	8	9
		1		50	7	64	
any of the four SARs		0		8.4	2.9	14	16
		1					
PUR	GGNNGAGGGAGARRRR	2		0.01	0.12	0	0
topo II	GTNWAYATTNATNNR	0		0.04	0.19	0	
		1		0.83	0.91	2	
		2		8.2	2.9	14	
topo II core	WAYATT	0		12	3	18	
pol start (s)	YYYYYYYYCTTTYYYY	0		0.002	0.043	0	
		1		0.05	0.22	0	
pol start (w)	YYYYYYRCCCYYYYY	0		0.001	0.033	0	
		1		0.03	0.18	0	
either polymerase start site		1		0.08	0.28	0	8
telo-like	TGGGT	0		1.1	1	3	0
p53	RRRCWWGYYY	0		0.43	0.66	1	3

CHAPTER SEVEN

CIRCULAR YAC VECTORS CONTAINING SHORT MAMMALIAN ORIGIN SEQUENCES ARE MAINTAINED UNDER SELECTION AS HeLa EPISOMES

Torsten O. Nielsen, Maria Zannis-Hadjopoulos and Gerald B. Price

Preceding chapters have shown that short DNA fragments containing mammalian origins of DNA replication can replicate autonomously both *in vitro* and following transient transfection into human cells. In this chapter, some of the origins described previously are cloned into a vector containing a marker selectable in human cells, allowing long-term maintenance of transfected constructs. The vector and assay methods used are chosen with the hope that this research will help lead toward the development of a human artificial chromosome.

This chapter has been submitted for publication in the *American Journal of Human Genetics*. I designed and executed all experiments, assembled the results and wrote all portions of the paper. Maria Zannis-Hadjopoulos and Gerald B. Price helped in the planning of experiments and the editing of the manuscript, and Gerald B. Price assisted in the cell culturing work for the fluctuation assay (Figure 3 and Table 1).

Circular YAC vectors containing short mammalian origin sequences are maintained under selection as HeLa episomes

Running Title: *YACs containing cloned human origins*

Key words: artificial chromosome, autonomous replication, DNA replication, episome, human, replication origin.

Torsten O. Nielsen, Maria Zannis-Hadjopoulos and Gerald B. Price

McGill Cancer Centre, McGill University, 3655 Drummond Street, Montréal, Québec, Canada

Send all correspondence to:

Dr. Gerald B. Price
McGill Cancer Centre
3655 Drummond Street
Montréal, Québec H3G 1Y6
CANADA

Tel: (514) 398-3531
FAX: (514) 398-6769
E-mail: FACS1@medcor.mcgill.ca

SUMMARY

pYAC*neo*, a 15.8 kb plasmid, contains a bacterial origin, G418-resistance gene, and yeast ARS, CEN and TEL elements. Three mammalian origins have been cloned into this circular vector: 343, a 448 bp chromosomal origin from a transcribed region of human 6q; X24, a 4.3 kb element containing the hamster DHFR origin of bidirectional replication (oriβ), and S3, a 1.1 kb human anti-cruciform purified autonomously replicating sequence. The resulting constructs have been transfected into HeLa cells, and G418-resistant subcultures were isolated. The frequency of G418-resistant transformation was higher with origin-containing YAC*neo* than with vector alone. After ten or more weeks of G418 selection (> 45 generations), the presence of episomal versus integrated constructs was assessed by fluctuation assay, testing stability of the resistance trait after removal of selection, and by PCR of supercoiled circular and linear genomic cellular DNAs separated on EtBr-CsCl gradients. In stable G418-resistant subcultures which had been transfected with vector alone or with linearized constructs, as well as in some subcultures transfected with circular origin-containing constructs, resistance was conferred by integration into the host genome. However, two examples were found of G418-resistant transfectants maintaining the YAC.343 and the YAC.S3 circular constructs in a strictly episomal state after long-term culture in selective medium, with 70% stability per cell division, and with the PCR-detectable transfected construct present in the supercoiled episomal and not the linear genomic CsCl fraction.

These versatile constructs, containing mammalian origins, have the capacity for further modification with human telomere or large putative centromere elements, in an effort to move towards construction of a human artificial chromosome.

INTRODUCTION

Building a human artificial chromosome would not only provide a valuable tool for addressing difficult questions about chromosome biology, but would also create an all-human transfection vector with the capacity to carry large chromosomal regions including complete transcriptional units, from even the largest genes, for the purpose of complementation mapping, or for gene therapy (Huxley, 1994).

Artificial chromosomes require three *cis*-acting functional components: replication origins, telomeres, and a centromere. *S. cerevisiae* origin-containing yeast ARS plasmids (Strichcomb et al., 1979) provided the basis for the addition of TEL (Szostak and Blackburn, 1982) and CEN (Dani and Zakian, 1983) elements to complete the construction of stable yeast artificial chromosomes, or YACs (Murray and Szostak, 1983). A similar strategy proved successful for artificial chromosome assembly in the fission yeast *Schizosaccharomyces pombé*, in spite of the far more complicated structure of its centromeres (Hahnenberger et al., 1989).

The first component required in such a "ground up" strategy for the assembly of a prototype human artificial chromosome is a functional human replication origin. Different techniques have permitted the identification of a limited but rapidly-increasing number of putative and proven mammalian origins (DePamphilis, 1993; Hamlin et al., 1994). Our group has been able to isolate large numbers of putative origins using such techniques as nascent strand extrusion (Frappier and Zannis-Hadjopoulos, 1987) and anti-cruciform immunoaffinity purification (Bell et al., 1991); these sequences permit short-term autonomous replication of plasmids transfected into human cells, and can act as replication origins in their native chromosomal position (Wu et al., 1993b).

To use such isolated origin sequences for the construction of a first stage human artificial chromosome, they must be cloned into a circular vector which permits transfection into human cells and selection of transfected clonal subpopulations, and has the capacity for further modification to carry human-functional telomeres and, potentially, very large (hundreds of kilobases) putative centromere elements. In addition, methods must be developed both to demonstrate that such constructs are maintained in long-term culture as independent episomal elements, not integrated into a host chromosome, and to measure their mitotic stability.

MATERIALS AND METHODS

Molecular Cloning

The 15.8 kb vector pYAC*neo* (Clontech) was digested with *Eco*RI. The 1.1 kb *Eco*RI insert of plasmid S3 (Nielsen et al., 1994) was ligated directly into the dephosphorylated vector, whereas the 448 bp *Eco*RI/*Hin*cII insert of pURHc34 (Wu et al., 1993a) and the 4.3 kb *Xba*I fragment of pX24 (Burhans et al., 1990) were blunt-ended with T4 DNA polymerase before ligation into blunt-ended, dephosphorylated vector. The resulting circular constructs are designated YAC.S3 (16.9 kb), Y.343 (16.3 kb), and Y.X24 (20.1 kb). Each was used to transform competent *E. coli* and ampicillin-resistant colonies were grown for large scale plasmid preparation. The structures of the cloned constructs were then confirmed by restriction enzyme digestion.

Transfection and Culturing in Human Cells

HeLa cells, passaged once since resurrection from frozen stocks, were seeded in T-25 flasks at 1 x 10^4/cm^2, and grown for two days (in α-MEM + 10% FCS) before transfection with 20 µg pYAC*neo*, YAC.S3, Y.343, or Y.X24 DNA by calcium phosphate coprecipitation. Linearized forms of pYAC*neo* and Y.343, with the *Tetrahymena* telomeric ends of the YAC vector, were produced by *Bam*HI digestion of the circular constructs, and similarly transfected. Since YAC.S3 and Y.X24 contain an extra *Bam*HI site in their inserts, they have not yet been tested in linearized form. Two days post-transfection, cells were switched into medium containing 400 µg/mL G418, and a further two days later, T-25 flasks were trypsinized, counted, and 1 x 10^5 cells were seeded onto 60 mm dishes. The HeLa cells were cultured in G418 until 20 days post-transfection, when dishes were scored for visible growing drug-resistant colonies. Individual colonies were picked directly from the 60 mm dishes to isolate clonal subpopulations for further analysis. Cultures were maintained in 400 µg/mL G418 during this period, and had been actively growing for at least ten weeks (an estimated minimum of 45 doublings) between the initial transfection and the subsequent analysis of the episomal versus integrated state of the transfected constructs.

Fluctuation assay

For each cloned transfectant cell line to be tested, cells which had been maintained in G418 were counted and 4 x 10^5 cells were plated into two T-80 flasks, one containing

drug-free nonselective medium, and one used for maintaining the culture in G418-selective medium. In parallel, cells were seeded at a similar density onto two 24-well plates (200 mm^2/well), to allow their growth curves to be followed both in the presence and absence of drug selection. Triplicate wells were trypsinized and counted daily with a Coulter Counter ZM apparatus. After six days of growth, both T-80 flasks were trypsinized, counted, and diluted with either G418-containing or nonselective medium to final concentrations of 5 cells/mL. 200 µL aliquots were then distributed to each well of a 96 well plate (32 mm^2/well). Two plates were used for the case of cells which had been passaged in nonselective medium and were now being returned to G418. Eight days later, the number of wells containing a growing cell colony was scored under the microscope.

Southern Blots

DNA from 1 x 10^6 cells was isolated by the alkaline lysis method of Sun et al. (1994), using 25 µg glycogen as a carrier during precipitation steps. Isolated DNA was separated on a 0.7% agarose, 1x TBE gel (5 h, 7.1 V/cm, at room temperature) and blot-transferred onto a Nytran membrane (Schleicher & Schuell). The membrane was fixed, prehybridized at 63°C in 6x SSC, 1% SDS, 10x Denhardt's containing 0.15 mg/mL boiled, sheared herring sperm DNA, and hybridized overnight with 7.5 x 10^5 dpm/mL nick-translated pYAC*neo*. The washed membrane was used to expose a Fuji BAS 2000 phospho-imager screen, and densitometry was performed using the BAS 2000 Image Analysis software.

Demonstration of Episomal DNA

As a positive control, HeLa cells were transiently transfected with Y.X24, by calcium phosphate coprecipitation, and harvested 48 h later. Using standard methods (Strauss, 1989), total DNA from approximately 5 x 10^6 cells was isolated from untransfected HeLa (negative control), transiently transfected HeLa, and the cloned transfectant cell lines to be tested. DNA preparations where mixed with 1 mg ethidium bromide and 75 µg of a carrier plasmid, in this case F9, a pBluescript clone containing a 0.5 kb human genomic insert (Nielsen et al., 1994). CsCl solution was added to a final density of 1.56 g/mL before ultracentrifugation in a VTi80 rotor, 20 h at 67 500 rpm. Using the intact and nicked carrier plasmid bands as a visible guide to the position of supercoiled (lower band) and linear and relaxed circular (upper band) DNA, the two completely-resolved fractions were carefully removed by side puncture. Ethidium bromide was

removed by two washes with CsCl-saturated isopropanol, and CsCl was removed by ethanol precipitation with two 70% ethanol washing steps.

Aliquots representing one-eighth of the purified DNA were used as template in two separate 50 µL PCR reactions. To amplify a 131 bp region of the *neo* gene present in the transfected constructs but not in the native HeLa genomic DNA, primers 5'-TCA GGA CAT AGC GTT GGC T-3' and 5'-CGT CAA GAA GGC GAT AGA A-3', located in the *neo* gene, were used (at 0.4 µM) with a mixture of all four dNTP's (each at 0.2 mM), 1 x Taq buffer, and 1 U Taq polymerase (Pharmacia). 28 cycles were performed, each 94°C, 20 sec denaturation; 50°C, 90 sec annealing; and 72°C, 30 sec extension; the first denaturation and the final extension steps were carried out for 5 min. To amplify a 423 bp unique region on the long arm of human chromosome 6, primers 5'-TGT GTA TGG GAC GGT AGT CA-3' and 5'-GGA GCA AGG CAG AAC TAC TC-3' (Wu et al., 1993b) were used at 0.25 µM, with 1.5 U Taq, for 33 cycles (each 94°C, 60 sec; 50°C, 60 sec; 72°C, 60 sec) followed by a 5 min final extension. Products of both reactions were electrophoresed in a 1.6% agarose, 1x TBE gel.

RESULTS

Cloning Mammalian Origins into a YAC Vector

pYAC*neo* is a versatile plasmid shuttle vector (Traver et al., 1989). Because it includes the prokaryotic *Col*E1 origin and an ampicillin resistance marker from pBR322, pYAC*neo* (15,827 bp) can be grown in *E. coli* as a circular plasmid. Since this vector also contains the *S. cerevisiae* ARS1 replication origin and CEN4 centromere elements, it can alternatively be maintained as a circular yeast artificial chromosome, carrying the TRP1, URA3, and HIS3 selectable markers. Digestion of the circular plasmid with *Bam*HI removes the HIS3 gene, leaving a linear molecule capped by two 0.7 kb telomeric cassettes (originally derived from *Tetrahymena*) which are functional in budding yeast and allow the maintenance of the construct as a linear yeast artificial chromosome. In addition, pYAC*neo* carries a gene conferring resistance to the drug G418, a trait which is selectable in mammalian cells. Thus, this vector can replicate in both *E. coli* and *S. cerevisiae*, and contains markers for the selection of stable bacterial, yeast, or mammalian cell transfectants.

Three mammalian sequences previously shown to permit autonomous replication in human cells have been cloned into the *Eco*RI site of pYAC*neo* (Figure 1). 343 is a 0.45 kb cDNA clone derived from a transcribed region on the long arm of human chromosome 6 (Shihab-El-Deen et al., 1993), to which *in vivo* origin activity has been localized by nascent

strand PCR mapping (Wu et al., 1993b). S3 is a 1.1 kb human sequence isolated by anti-cruciform immunoaffinity purification of genomic DNA, followed by competitive selection for clones possessing strong autonomous replication activity by mass transfection and *in vitro* replication assays (Nielsen et al., 1994). Finally, X24 carries a 4.3 kb *Xba*I fragment from the hamster *DHFR* 3' region, and includes the predominant initiation site oriβ, as indicated by multiple techniques (Burhans et al., 1990; Zannis-Hadjopoulos et al., 1994).

HeLa Transfection Efficiency is Higher with Origin-Containing Constructs

Since the Y.343, YAC.S3, and Y.X24 constructs remain relatively small in size (16.3 kb, 16.9 kb, and 20.1 kb respectively), they can be grown in bacteria, and pure preparations can be transfected by the relatively high-efficiency calcium-phosphate coprecipitation method, unlike YACs in the 0.1 - 1.0 Mb range, which necessitate the use of techniques such as yeast spheroplast fusion for transfecting mammalian cells. Figure 2 shows that while calcium phosphate-treated, mock-transfected cells yield no G418-resistant colonies, pYAC*neo* vector devoid of any mammalian origin gave 45 colonies per 10^5 cells plated. Importantly, the test constructs YAC.S3, Y.343 and Y.X24 gave 50 - 100% more stably transfected G418-resistant colonies than the vector alone, suggesting that the presence of a short, cloned origin-containing insert is facilitating the maintenance of the transfected *neo* trait in human cells. Repeat experiments have further shown that linearized versions of pYAC*neo* and Y.343, produced by *Bam*HI digestion to free the T_2G_4 telomeric ends, are at least 100-fold less efficient in producing stable transfectants (data not shown).

Isolated colonies were cloned and grown in G418 to select for maintenance of the transfected constructs, and spent a minimum of 67 days (and in some cases up to 96 days) in culture before testing for the presence and stability of episomal *neo*-containing DNA.

Stability of Transfected Constructs

A protocol based on the classical Luria-Delbrück fluctuation assay permits calculation of the stability of the drug resistance marker during cell growth in non-selective medium. Stable transformants obtained in the usual fashion, through integration of the transfected marker into the host genome, maintain the drug resistance trait even in the absence of selective pressure. However, episomally-replicating DNA which lacks a functional centromere will not partition accurately at mitosis and will display a characteristic loss rate per generation.

Examples of G418-resistant clonal cell populations initially transfected with circular pYAC*neo*, YAC.S3, Y.343, or Y.X24, or with *Bam*HI-linearized pYAC*neo* or Y.343, were tested by fluctuation assay. In essence, cells were passaged from G418 into nonselective medium and allowed to grow, while in parallel, cells were also seeded onto 24 well plates to allow daily monitoring of their growth rate. Growth rates of individual clones tended to fall into one of the two patterns presented in Figure 3. Half of the tested cell lines (including YAC*neo* clone 1, YAC.S3 clone 2, and Y.343 clone 1) grew slightly more slowly in G418 than they did in nonselective medium, as did Y.X24 clone 1 (Fig. 3). However, growth rates in the remaining cell lines, including Y.343 clone 2 (Fig. 3), YAC.S3 clone 1, and YAC*neo* clones 2 and 3, were retarded in G418, with a doubling time in nonselective medium approximately two thirds of that observed in the presence of G418. A reduced population growth rate in G418 may reflect the loss of the *neo* marker during some cell divisions, or could instead reflect poor transcription of the *neo* gene in particular integrated contexts. From a regression analysis of the exponential growth rate in nonselective medium, the number of doublings which took place during the fluctuation period was calculated.

Six days after seeding into nonselective medium, while the cells were still in log phase, the test flasks were trypsinized, diluted, and replated to determine the proportion of cells that still retained the *neo* marker. Results are shown in Table 1, along with the calculated stability of the constructs in each of the subpopulations tested. All HeLa clones carrying circular pYAC*neo* vector maintained the *neo* marker with a stability of approximately 1, supporting the hypothesis that their transfection is the result of integration events. The same is true for several of the clones transfected with origin-containing constructs. However, in one of two YAC.S3 and one of two Y.343 HeLa cell clones, a significant proportion of the cells in each population lost the G418 resistance trait during the nonselective fluctuation period, strongly suggesting that these clonal lines carried only episomal forms of the transfected constructs. Indeed, their calculated stability of 0.7 per generation is similar to that obtained by others during construction of *S. cerevisiae* and *S. pombé* yeast artificial chromosomes, using equivalent plasmids which carried a functional origin but lacked a centromere.

Tested clones which were transfected with linearized constructs have so far only displayed fluctuation results compatible with integration, in line with observations that the YAC vector's T_2G_4 telomeres are non-functional in human cells (Hanish et al., 1994). In contrast, the ultimate instability of episomally-maintained *neo* marker during long-term nonselective growth has been confirmed, in the case of Y.343 clone 2, by repeating the

fluctuation assay using a 46-day period of nonselective growth, after which in excess of 99% of cells have lost their drug resistant phenotype.

HeLa Cells Maintain Human Origin-Containing YACs as Episomes

Initial attempts to differentiate episomal from integrated DNA by Southern blots of low and high molecular weight Hirt lysate fractions were hampered by a poor detection threshold, inadequate separation of 20-kilobase episomes from genomic DNA, and the inability, following restriction digests, to differentiate episomal DNA from the product of head-to-head multimeric integration events. Nevertheless, results suggested that in Y.343 clone 1, and in YAC*neo* clones 1 and 2, transfected constructs were integrated into the genome, whereas Y.343 clone 2 carried intact Y.343 at a total copy number of approximately 30 per cell (data not shown). Using an alkaline cell lysis technique, which selects for covalently closed circular DNA, we found that Y.343 is present in clone 2 at 119 days after transfection, but only in the case where G418 selection was constantly maintained (Figure 4, lane A). Consistent with the results of the fluctuation assay, no band was visible from a cell population that had been passaged in nonselective medium between day 89 and lysis on day 119 (Figure 4, lane B). Most of the DNA migrated in the same position as the form I supercoiled monomer in the Y.343 standard (Figure 4, lane C). The second band observed in lane A may represent linearized monomer generated by the introduction of double-strand breaks during the alkaline preparation procedure, since it migrates at the same position as 16.1 kb unit length linearized Y.343, or, alternatively, it could represent a episomal multimeric recombined form of Y.343.

In an effort to confirm that the modified YAC plasmids containing human origins were indeed being maintained in episomal form, total DNA was isolated from HeLa subclones YAC.S3 clone 1 and Y.343 clone 2, which had demonstrated instability of the drug resistance marker during fluctuation assay. CsCl / ethidium bromide density gradient ultracentrifugation was, in fact, originally developed as a means for separating episomal from genomic DNA in HeLa cells (Radloff et al., 1967), and conveniently resolves covalently-closed circular DNA from linear DNA and nicked (form II) circular DNA, which intercalate more of the buoyant dye. Any old, unrelated plasmid preparation which has accumulated 50% nicked and linear forms can serve as a visible guide to the position of the lower (supercoiled episomal) and upper (predominantly linear genomic) bands in the density gradient. DNA fractions prepared in this fashion from the HeLa subclones, as well as from positive and negative controls, were analyzed by PCR, using primers from the *neo* gene (Figure 5).

DNA from normal, untransfected HeLa cells did not yield a PCR product (only the carrier plasmid is visible) since these cells do not carry a *neo* gene. A polyclonal population of HeLa cells, transfected with Y.X24 48 h before DNA preparation, contained large quantities of PCR template in both the lower (intact circular) and upper (nicked or damaged circular, or integrated) CsCl gradient bands. However, the test clones yielded a product from only the lower, episomal DNA fraction, indicating that the transfected YAC.S3 and Y.343 constructs are being maintained, in these subcultures, as covalently closed circular episomes during long-term culture in selective medium (81 days for YAC.S3 clone 1; 96 days for Y.343 clone 2). The rightmost lanes show, through the use of primers directed at a unique genomic locus, that the linear genomic DNA segregates exclusively to the upper CsCl band; thus, the lower band is free from contaminating genomic DNA, and if a copy of the construct had integrated, it should have been detected in the upper band fraction. In fact, a 131 bp PCR product does become visible in the upper band fraction of YAC.S3 clone 1 and Y.343 clone 2, but only after 40 cycles of PCR (data not shown), and not after 35 or 28 cycles (when an equal or lesser amount of lower band template yields a detectable product); this likely represents the small amount of episomal DNA (less than 0.5%) damaged during preparation and now present in a relaxed circular or linear form. In theory, however, it could also represent a very small subpopulation of cells in which an integration event has occurred, perhaps even without any expression from integrated *neo*.

DISCUSSION

Mammalian origin sequences have been cloned into a versatile shuttle vector, capable of further modification and of growth in several host systems. These constructs remain small enough to be transfected into human cells by calcium phosphate coprecipitation, and are selected for using the *neo* marker.

The efficiency of stable transfection is doubled by the presence of an origin sequence, in line with previous observations using a cloned portion of the 5' c-*myc* origin (McWhinney and Leffak, 1990), but remains far less impressive than the 1000-fold increased yeast transfection efficiency conferred onto plasmids by the presence of an ARS element (Stinchcomb et al., 1979). Stable transfection of the *neo* gene, using standard vectors, requires integration of the plasmid into mammalian genomic DNA (Southern and Berg, 1982). Recently, attempts have been made to use YAC vectors to transfect large genomic regions (hundreds of kb) into mammalian cells (Forget, 1993), but again stable transfection required integration into the host genome. Nonet and Wahl (1993) cloned a 70 kb region including the mouse 5' adenosine deaminase origin (Carroll et al., 1993) into a

YAC and transfected mouse cells by yeast spheroplast fusion. In this case, examples were found of stable transfectants harbouring episomally-replicating circular constructs, but only in the presence of coexisting integrated copies. The modified YACs used here to transfect human cells are completely defined, can be grown in bacteria or yeast, can be transfected by calcium phosphate coprecipitation (which gives higher transformation efficiencies and does not cotransfect yeast genomic DNA), and YAC.S3 and Y.343 have been demonstrated to be capable, in a G418-selected human cell system, of persisting for several months in a purely episomal form. However, cells which carry episomal *neo* grow more slowly in G418 than in nonselective medium, and, while this is also true for a subset of the integrated transfectants, there may remain a selective pressure for eventual integration of the episomes, or for recombination into multimeric forms which may replicate more efficiently (Kelly et al., 1995).

While the oriβ from the hamster *DHFR* locus can permit autonomous replication in human cells (Zannis-Hadjopoulos et al., 1994), only two Y.X24 subpopulations have been analyzed for the presence of episomal elements, and neither showed fluctuation instability or an exclusively episomal PCR product (one clone did, however, yield PCR product in both the genomic and episomal fractions). If the rate of stable transfection with vector alone defines the frequency of integration, and Y.X24 has a 50% higher transfection efficiency than pYAC*neo* (Figure 2), an estimated one in three Y.X24 transfectant subpopulations might be expected to contain episomes; thus, analysis of more transfected colonies would be required before eliminating the possibility that Y.X24 can also be maintained as an episome in a long-term assay.

The overall stability of YAC.S3 and Y.343 episome maintenance during growth in nonselective medium is comparable to that obtained with autonomously replicating yeast plasmids (Table 1), which constituted the first step in the construction of yeast artificial chromosomes. To complete a human artificial chromosome, the T_2G_4 telomeres will likely have to be replaced by mammalian-functional T_2AG_3 telomere cassettes. The resulting linear acentromeric chromosome would represent an appropriate cloning vector for the isolation of a human centromere, which might be hundreds of kilobases or more in length (Earnshaw and Tomkiel, 1992). The modified pYAC*neo* constructs described here are sufficiently versatile to allow such additions to be made. Stability in nonselective medium would be expected to be enhanced, from about 70% to more than 99.99%, in larger constructs (>100 kb) containing a centromere.

Vectors capable of long-term persistence in mammalian cells have been constructed by others, but these rely on viral origins of DNA replication. Examples include a murine plasmid based on a defective polyoma virus (Gassman et al., 1995), and a "human artificial

episomal chromosome" which carries the EBV latent origin, *oriP* (Sun et al., 1994). While capable of acting as excellent gene vectors *in vitro*, such constructs are not helpful for studying mammalian chromosomal origin biology, and require the presence of viral transactivating proteins for their replication. Attempts to create artificial chromosomes by deleting large blocks from native chromosomes still leave huge undefined regions. The potential of a mammalian artificial chromosome, not only as a gene vector, but also as a model for addressing questions about chromosomal biology, may be best reached through a "ground-up" assembly of the functional *cis*-acting components, origins, telomeres, and centromeres. Origins represent the logical first step.

ACKNOWLEDGEMENTS

We thank Drs. M.L. DePamphilis (Roche Institute of Molecular Biology) and J.L. Hamlin (University of Virginia) for providing cloned fragments from the *DHFR* region, and Dr. DePamphilis for giving us plasmid pX24. We also thank Drs. C. Wu (Biotechnology Research Institute, Montréal) and M. Leffak (Wright State University) for helpful advice in PCR primer design, and Antoine Jothy for technical assistance. T.O.N. is a recipient of a studentship from the Medical Research Council of Canada. This work is supported by a grant from the Canadian Genome Analysis and Technology program.

REFERENCES

Bell D, Sabloff M, Zannis-Hadjopoulos M, Price GB (1991) Anti-cruciform DNA affinity purification of active mammalian origins of replication. *Biochim. Biophys. Acta* 1089:299-308

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

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

Dani GM, Zakian VA (1983) Mitotic and meiotic stability of linear plasmids in yeast. *Proc. Natl. Acad. Sci. USA* 80:3406-3410

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

Earnshaw WC, Tomkiel JE (1992) Centromere and kinetochore structure. *Curr. Opin. Cell Biol.* 4:86-93

Forget BG (1993) YAC transgenes: bigger is probably better. *Proc. Natl. Acad. Sci. USA* 90:7909-7911

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

Gassman M, Donoho G, Berg P (1995) Maintenance of an extrachromosomal plasmid vector in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. USA* 92:1292-1296

Hahnenberger KM, Baum MP, Polizzi CM, Carbon J, Clarke L (1989) Construction of artificial minichromosomes in the fission yeast *Schizosaccharomyces pombe*. *Proc. Natl. Acad. Sci. USA* 86:577-581

Hamlin JL, Mosca PJ, Levenson VV (1994) Defining origins of replication in mammalian cells. *Biochim. Biophys. Acta* 1198:85-111

Hanish JP, Yanowitz JL, de Lange T (1994) Stringent sequence requirements for the formation of human telomeres. *Proc. Natl. Acad. Sci. USA* 91:8861-8865

Huxley C (1994) Mammalian artificial chromosomes: a new tool for gene therapy. *Gene Ther.* 1:7-12

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

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

Murray AW, Szostak JW (1983) Construction of artificial chromosomes in yeast. *Nature* 305:189-193

Nielsen T, Bell D, Lamoureux C, Zannis-Hadjopoulos M, Price GB (1994) A reproducible method for identification of human genomic DNA autonomously replicating sequences. *Mol. Gen. Genet.* 242:280-288

Nonet GH, Wahl GM (1993) Introduction of YACs containing a putative mammalian replication origin into mammalian cells can generate structures that replicate autonomously. *Somat. Cell Mol. Genet.* 19:171-192

Radloff R, Bauer W, Vinograd J (1967) A dye-buoyant density method for the detection and isolation of closed circular duplex DNA: the closed circular DNA in HeLa cells. *Proc. Natl. Acad. Sci. USA* 57:1514-1521

Shihab-El-Deen A, Wu C, Hamerton JL, Zannis-Hadjopoulos M, Price GB (1993) Chromosomal localization of a sequence with in vivo activity for initiation of DNA replication. *Somat. Cell Mol. Genet.* 19:103-109

Southern PJ, Berg P (1982) Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. *J. Mol. Appl. Genet.* 1:327-341

Strauss WM (1989) Preparation of genomic DNA from mammalian tissue. In: Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (eds) *Short protocols in molecular biology*. John Wiley & Sons, New York, p 59

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

Sun TQ, Fenstermacher DA, Vos JMH (1994) Human artificial episomal chromosomes for cloning large DNA fragments in human cells. *Nature Genet.* 8:33-41

Szostak JW, Blackburn EH (1982) Cloning yeast telomeres on linear plasmid vectors. *Cell* 29:245-255

Traver CN, Klapholz S, Hyman RW, Davis RW (1989) Rapid screening of a human genomic library in yeast artificial chromosomes for single-copy sequences. *Proc. Natl. Acad. Sci. USA* 86:5898-5902

Wu C, Friedlander P, Lamoureux C, Zannis-Hadjopoulos M, Price GB (1993a) cDNA clones contain autonomous replication activity. *Biochim. Biophys. Acta* 1174:241-257

Wu C, Zannis-Hadjopoulos M, Price GB (1993b) *In vivo* activity for initiation of DNA replication resides in a transcribed region of the human genome. *Biochim. Biophys. Acta* 1174:258-266

Zannis-Hadjopoulos M, Nielsen TO, Todd A, Price GB (1994) Autonomous replication in vivo and in vitro of clones spanning the region of the *DHFR* origin of bidirectional replication *(oriβ)*. *Gene* 151:273-277

TABLE 1: *In Vivo* Stability of Transfected Constructs by Fluctuation Assay

Host Cell	Transfection	# growing colonies on 96 well plates after fluctuation		number of generations	STABILITY per division[d]
		G418[a]	nonselective[b]	#gen [c]	
HeLa	YAC*neo* clone 1	57	44	3.2	1.0
HeLa	YAC*neo* clone 2	42	47	2.8	0.9
HeLa	YAC*neo* clone 3	58	69	2.8	0.9
HeLa	YAC.S3 clone 1	36	64	2.9	**0.7**
HeLa	YAC.S3 clone 2	26	28	3.0	1.0
HeLa	Y.343 clone 1	20	14	3.2	1.0
HeLa	Y.343 clone 2	26	56	3.4	**0.7**
HeLa	Y.X24 clone 1	18	28	3.5	0.9
HeLa	linear YAC*neo* clone 1	12	14	3.3	1.0
HeLa	linear Y.343 clone 1	37	38	3.3	1.0
S. cerevisiae	circular ARS plasmid				0.7 [e]
S. cerevisiae	linear ARS plasmid				0.8 [e]
S. cerevisiae	CEN-containing YAC				0.9-0.999 [e]
S. pombé	circular *ars* plasmid				0.7 [e]
S. pombé	*cen*-containing YAC				0.98-0.999 [e]
Any host cell	integrated DNA				1.0

Footnotes:

a. **G418** = number of wells containing growing colonies, after test clone cells, grown in non-selective medium for six days, were *returned to G418 selective medium* while plating at one cell per well onto 96 well plates.

b. **nonselective** = number of wells containing growing colonies, after test clone cells, grown in non-selective medium for six days, were *kept in nonselective medium* while plating at one cell per well onto 96 well plates.

c. **#gen** = the number of cell divisions which took place during the six day nonselective fluctuation period, as assessed by daily counts of parallel cultures.

d. STABILITY refers to the chance, following each cell division, that a daughter cell will inherit the selectable marker, and is calculated from the following relation:

(**#gen**)(ln STABILITY) = ln{ln[(96-**G418**)/96]/ln[(96-**nonselective**)/96]}

Results are shown to one significant figure, where 1.0 is the maximum stability possible.

e. Data from Murray and Szostak (1983), and Hahnenberger et al. (1989). Stability of centromere-containing yeast artificial chromosomes increases as a function of size.

FIGURE 1. Diagram of pYAC*neo* and cloned mammalian origin sequences. In the vector, prokaryotic sequences are indicated by a thin grey line. Functional yeast *cis*-acting chromosomal components are shown as open boxes, while yeast marker genes are indicated by shaded boxes. The mammalian-selectable G418 resistance marker is in black. Two human (343, S3) and one hamster (X24) origin-containing, autonomously replicating sequences were cloned into the *Eco*RI site of pYAC*neo*; restriction sites indicating the orientation of the inserts are presented. All components are drawn to scale.

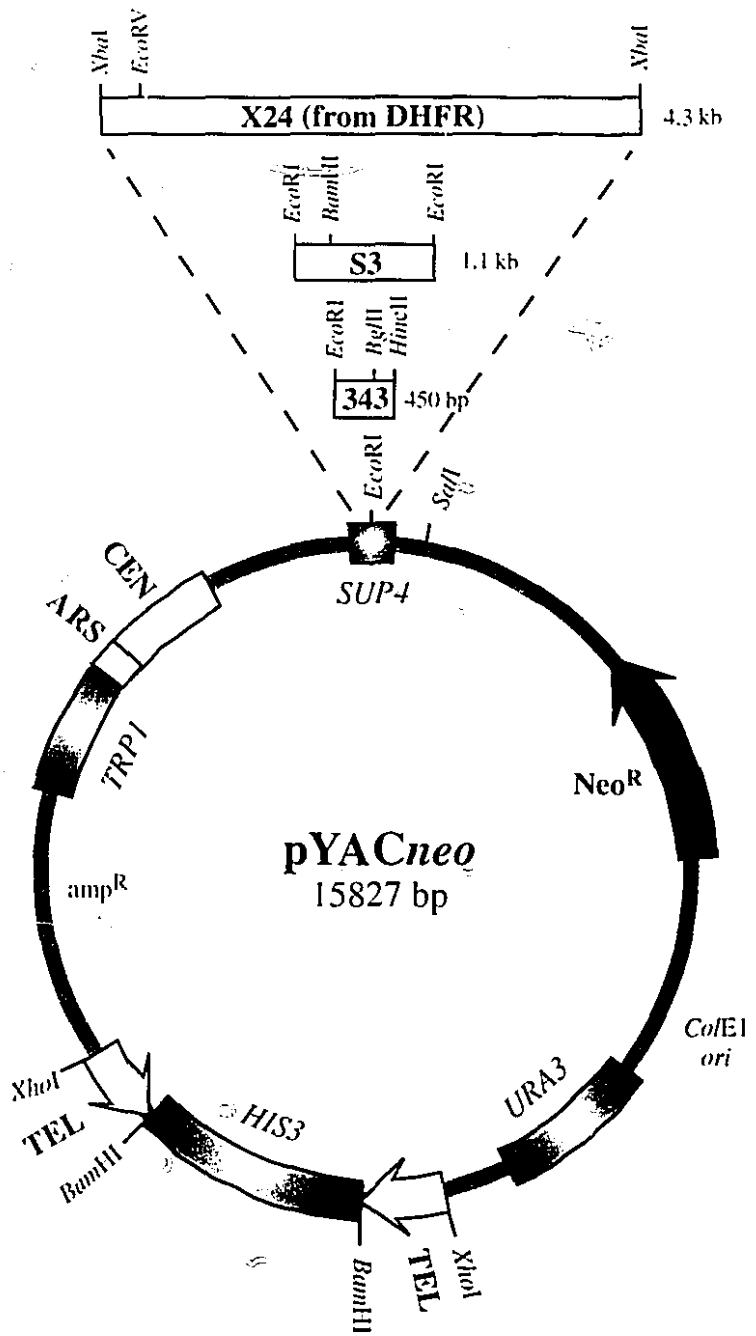

XbaI
EcoRV
XbaI
X24 (from DHFR)
4.3 kb
EcoRI
BamHI
EcoRI
S3
1.1 kb
EcoRI
BglII
HincII
343
450 bp
EcoRI
SalI
SUP4
CEN
ARS
TRP1
ampR
pYACneo
15827 bp
NeoR
ColE1
ori
URA3
HIS3
XhoI
TEL
BamHI
BamHI
TEL
XhoI

FIGURE 2. Colony assay for efficiency of stable *neo*-trait transfection with origin-containing circular YAC constructs. The number of G418-resistant colonies formed per 100 000 transfected cells is plotted for each construct, for vector alone, and for a "dummy"-transfected negative control, as scored 20 days following calcium phosphate coprecipitation transfections of HeLa host cells.

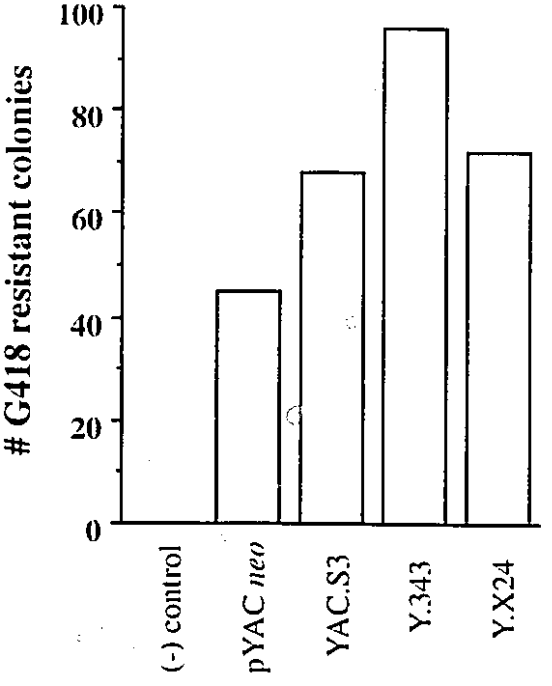
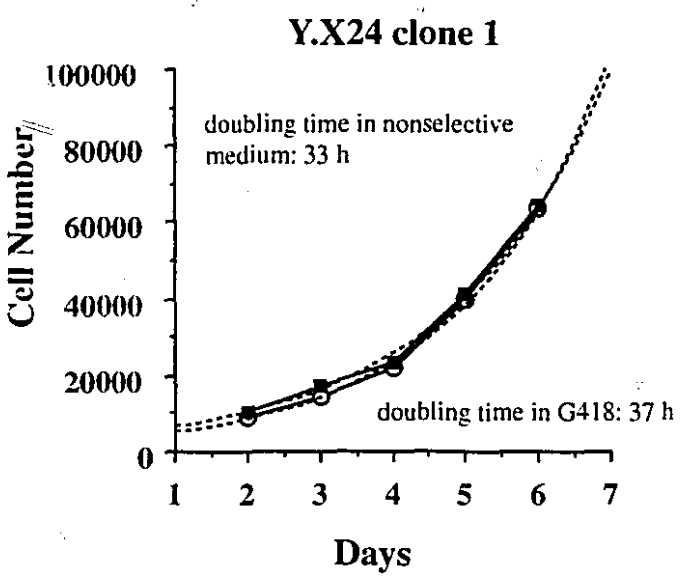
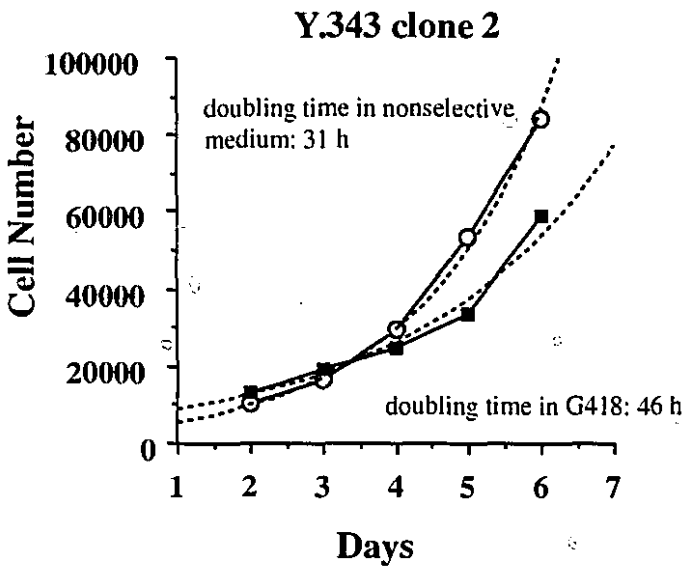

G418 resistant colonies
100
80
60
40
20
0
(-) control
pYAC neo
YAC.S3
Y.343
Y.X24

FIGURE 3. Growth curves of two representative HeLa clonal subpopulations, derived from transfections with mammalian origin-containing, modified YAC vectors. After plating in 200 mm^2 wells, cell number was determined by Coulter count; each plotted point represents the average of triplicate wells for each day during log-phase growth. Open circles denote cell number in nonselective medium, while filled squares track growth in the presence of 400 μg/mL G418. Dashed lines represent the exponential equation derived from regression analysis of each series of points, from which the population doubling time is determined; in all cases the curve fit correlation (R^2) value was better than 0.98.


Y.X24 clone 1
100000
80000
60000
40000
20000
0
Cell Number
doubling time in nonselective medium: 33 h
doubling time in G418: 37 h
1
2
3
4
5
6
7
Days


Y.343 clone 2
100000
80000
60000
40000
20000
0
Cell Number
doubling time in nonselective medium: 31 h
doubling time in G418: 46 h
1
2
3
4
5
6
7
Days

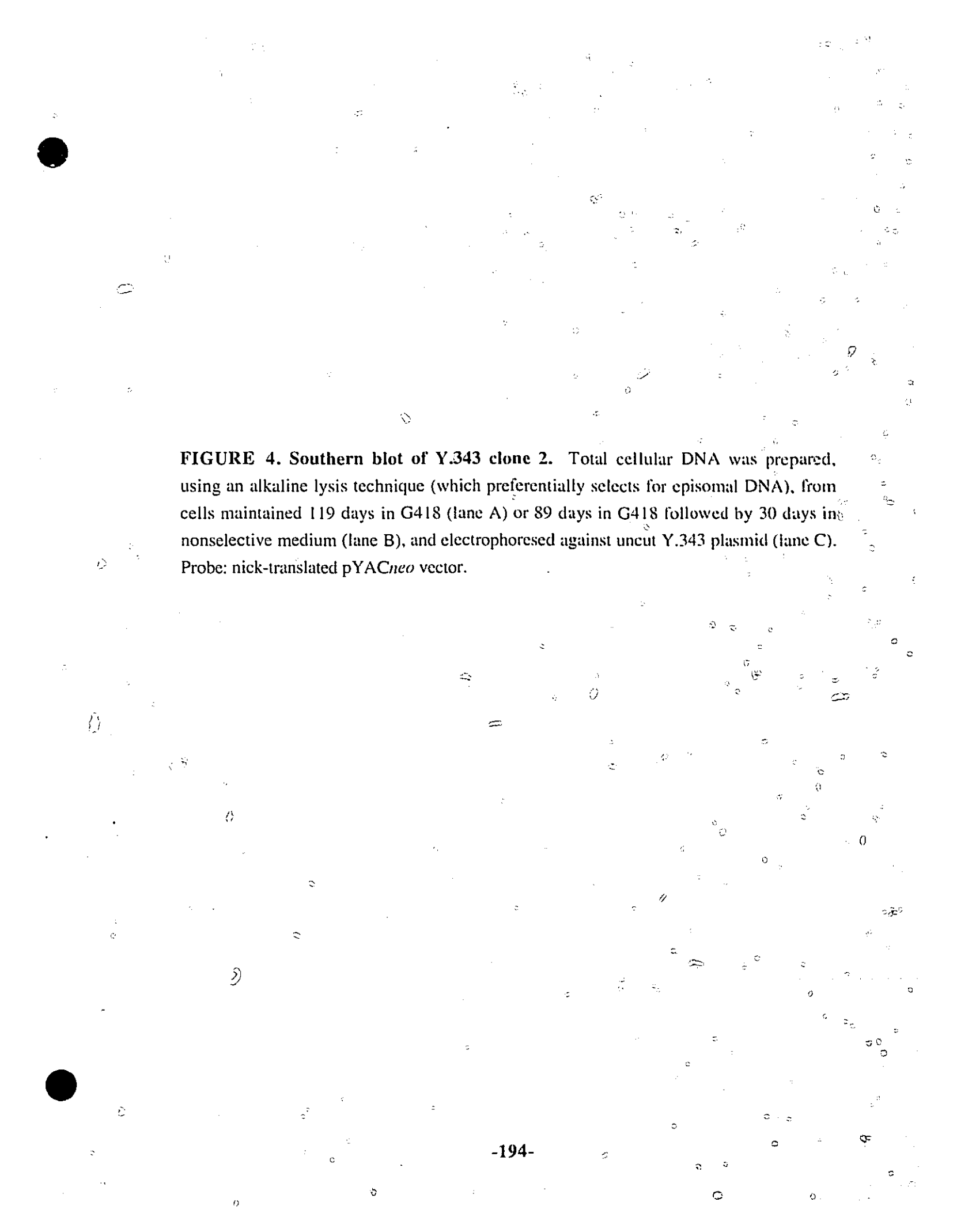


FIGURE 4. Southern blot of Y.343 clone 2. Total cellular DNA was prepared, using an alkaline lysis technique (which preferentially selects for episomal DNA), from cells maintained 119 days in G418 (lane A) or 89 days in G418 followed by 30 days in nonselective medium (lane B), and electrophoresed against uncut Y.343 plasmid (lane C). Probe: nick-translated pYAC*neo* vector.

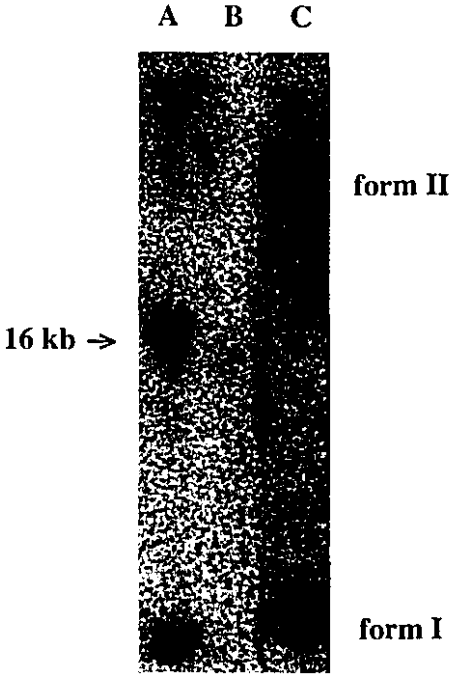
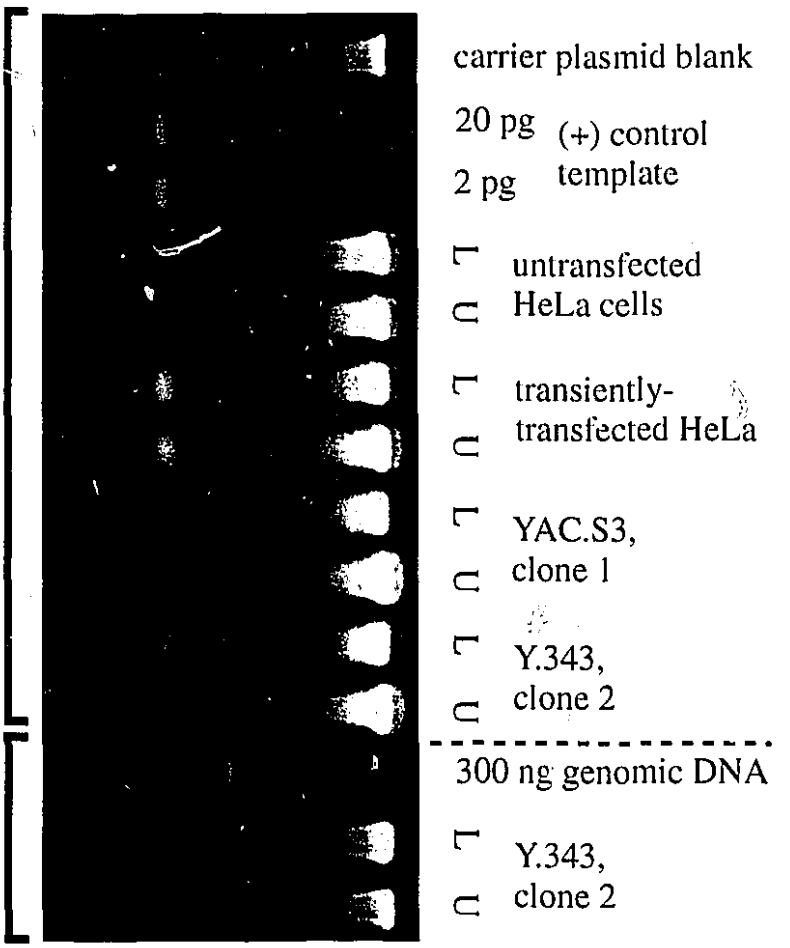

A
B
C
form II
16 kb →
form I

FIGURE 5. PCR analysis of supercoiled episomal and linear genomic DNA fractions. Total DNA from HeLa subcultures was fractionated by CsCl / EtBr ultracentifugation, using a carrier plasmid as a guide to the position of the lower band (L) containing supercoiled circular DNA, and the upper band (U) containing linear and relaxed circular forms of DNA. Fractions were tested for the presence of the transfected constructs by PCR amplification using a primer pair directed against the *neo* marker (positive control template is a pure Y.X24 plasmid sample). The position of linear genomic chromosomal DNA was confirmed in a separate amplification, shown in the three rightmost lanes, using primers that recognize a unique region on human chromosome 6 (positive control is total genomic DNA isolated from the ME180 human cell line, ATCC HTB 33).


carrier plasmid blank
20 pg (+) control
2 pg template
L untransfected
U HeLa cells
L transiently-
U transfected HeLa
L YAC.S3,
U clone 1
L Y.343,
U clone 2
300 ng genomic DNA
L Y.343,
U clone 2
neo primers, to detect fractions
unique genomic DNA primers
423 bp genomic product
131 bp neo product
leftover

CHAPTER EIGHT

GENERAL DISCUSSION

A consistent theme among all the experiments presented in this dissertation is that short fragments derived from mammalian genomic DNA can replicate autonomously in human cells and cell extracts. Results using transfection and *in vitro* assay systems are consistent both with each other, and with results obtained using different mapping methodologies. Consequently, mammalian cell replication, like the simpler model systems discussed in Chapter 1, fits the replicon model, inasmuch as initiation requires the presence of specific, cloneable, *cis*-active sequences. Since origin function is cloneable, autonomous replication assays can be used to isolate new origins, by mass screening of origin-enriched libraries, or by the systematic testing of contiguous fragments from larger genomic regions. Identification and analysis of new origins could then help in the determination of consensus sequences, the isolation of an initiator protein, and the examination of temporal controls and the chromosomal domain organization of human replication origins. In addition, origins can be cloned into larger constructs containing other *cis*-acting chromosomal components, in an effort to build a human artificial chromosome.

Short, cloned mammalian DNA sequences that have been demonstrated to replicate autonomously in these studies include 4.8 kb fragments from the oriβ region 3' to the hamster *DHFR* gene (Chapter 3), a subset of human fibroblast and adenocarcinoma anti-cruciform purified genomic fragments (Chapter 4), ranging in size from 110 bp to 3.4 kb, pHL-1 (1942 bp) and a 957 bp subclone containing portions of coxIII cDNA (Chapter 5), and a 1012 bp segment, p82, from the asynchronously-replicating imprinted human chromosomal region 15q11q13 (Chapter 6). In addition, the autonomous replication of plasmids *ors*8 and *ors*12 (Frappier and Zannis-Hadjopoulos, 1987; Pearson et al., 1991), originally isolated by the nascent strand extrusion method, was confirmed. A 448 bp fragment (Wu et al., 1993a) derived from the long arm of human chromosome 6 (Shihab-El-Deen et al., 1993) and mapped to a chromosomal initiation site by nascent strand PCR (Wu et al., 1993b), was shown to permit long-term episomal replication in the new context of the pYAC*neo* vector (Chapter 7). In cases where the same plasmids were individually tested by both transfection and *in vitro* replication assays, including *DHFR* fragments (Chapter 3), 13 autonomously replicating anticruciform clones (Chapter 4 Addendum), and pHL-1 (Chapter 5), results from the two assays were entirely consistent. Finally, mapping of the initiation site, as the earliest labeled fragment *in vitro*, confirmed that autonomous replication initiates within the inserted putative origin sequence in both pHL-1 (Chapter 5) and p82 (Chapter 6).

The validity of short fragment mammalian autonomous replication assays in relation to physical initiation site mapping techniques has also been confirmed by other investigators. Pearson et al. (1994) showed that initiation bubbles are seen, by electron microscopy, to centre upon the cloned insert of the monkey *ors*8 sequence, replicated using the same *in vitro* assay system employed in chapters 3 through 6. The human 5' c-*myc* promoter region has been demonstrated to contain a chromosomal origin of DNA replication by several methods, including earliest labeled fragment, nascent strand PCR, and Okazaki strand switching (reviewed in Gilbert et al., 1993), and has also been shown to direct autonomous replication in transfection (McWhinney and Leffak, 1988; Iguchi-Ariga et al., 1988) and *in vitro* replication assays (Berberich et al., 1995). The initiation site *in vitro* maps, by neutral-neutral 2DGE, electron microscopy, and Okazaki strand switching, to a 2.4 kb c-*myc* promoter fragment (Berberich et al., 1995) which contains the chromosomal initiation site. Recent work shows that *in vivo* chromosomal origins, mapped by nascent strand PCR or Okazaki fragment strand switch assay, are also autonomously replicating sequences, permitting episomal replication of the mouse IgH enhancer origin (Ariizumi et al., 1993), mouse 5' (Carroll et al., 1993) and 3' (Virta-Pearlman et al., 1993) adenosine deaminase origins, the hamster *CAD* gene internal origin (Kelly et al., 1995), and the human 5' *hsp*70 origin (Taira et al., 1994).

A chromosomal initiation site, oriβ, has been mapped 3' to the hamster *DHFR* gene by many methods, including earliest labeled fragment, imbalanced DNA synthesis, nascent strand PCR, Okazaki strand switching (reviewed in Burhans et al., 1990), and both types of 2DGE (Vaughn et al., 1990). While fragments containing oriβ have been shown to function as chromosomal origins after ectopic insertion into other genomic sites (Handeli et al., 1989), at least two groups were unable to show significant and reproducible autonomous replication of short (4 - 6 kb) fragments incorporating oriβ (Burhans et al., 1990; Caddle and Calos, 1992). However, in these experiments the assay conditions employed were not sufficiently sensitive to detect autonomous replication of plasmids containing c-*myc* origin or *ors* element inserts that had been shown to replicate by other groups. Autonomous replication assays in mammalian cells are considerably less sensitive than yeast ARS assays (Hamlin et al., 1994), and the *DHFR* 3' origin may not be used in every cell cycle (Dijkwel and Hamlin, 1995). Under more sensitive transfection and *in vitro* assay conditions, with *myc* and *ors* elements as positive controls, fragments from the hamster 3' DHFR region do replicate autonomously (Chapter 3). To date, then, all mammalian chromosomal origins tested by cloning into plasmids can direct autonomous replication, much as all tested yeast chromosomal origins can function as plasmid ARS elements.

These experiments substantiate the notion that the replication origins in human and other mammalian chromosomes are structurally similar to those present in well-characterized prokaryotic and eukaryotic model systems, such as *E. coli*, λ phage, SV40, Epstein-Barr virus, and yeast (Chapter 1). The replicon model (Jacob and Brenner, 1963) predicts that *cis*-acting genetic control elements (replicators) are closely related to actual initiation sites for DNA replication, and can be cloned as a functional unit. A complete origin contains multiple components, including an initiator protein binding site and a DNA unwinding element, with auxiliary proteins such as transcription factors facilitating or modulating origin function; particular DNA secondary structures and interactions with the nuclear matrix may also be necessary for recognition and activation of the origin (Chapter 1). In support of this model, genetic replicator activity was found in DNA fragments derived from a known mammalian chromosomal initiation site (Chapter 3), large numbers of autonomously replicating putative human origins of DNA replication were isolated from libraries of sequences which form cruciforms *in vivo* (Chapter 4), and a human autonomously replicating sequence was found to coincide with a genomic sequence containing a potential DNA unwinding element, bent and cruciform secondary structures, and associated matches to nuclear matrix and transcription factor binding site consensus sequences (Chapter 6).

Taking advantage of the capacity to clone origins and screen for autonomous replication activity, this thesis describes two methods for isolating new human origins of DNA replication. The first requires the availability of long sequenced blocks of genomic DNA, cloned contiguously from a chromosomal region suspected, on the basis of replication timing studies, to contain an origin (Chapter 6). The *in vitro* replication assay, in particular, presents a rapid method for screening a series of such clones for putative origin activity. A strong case can be made for the existence of a chromosomal replication origin when a clone containing a DNA unwinding element and other suspected origin components replicates autonomously *in vitro*, with initiation occurring in the genomic insert. This approach could become increasingly useful as more and more regions of human genomic DNA are sequenced in conjunction with the Human Genome Project. Secondly, genomic libraries enriched for the presence of origins can be screened by competitive mass assays for autonomous replication activity, using a pool of clones in each experiment (Chapter 4). Shotgun libraries do not contain a sufficient density of origin-containing clones to permit human autonomous replication assays to detect and isolate genomic origins successfully (Biamonti et al., 1985). However, enriched libraries can be created through the isolation of nascent strand DNA (Kaufman et al., 1985; Dimitrova et al., 1993), or by immunoaffinity purification of fragments which bind origin auxiliary

factors (Iguchi-Ariga et al., 1987) or adopt a DNA cruciform secondary structure (Bell et al., 1991). In these experiments, some isolated clones were tested individually for autonomous replication activity, in an effort to screen for putative origins. While standard single-clone assays have been successful in isolating a handful of autonomously replicating sequences, including examples proven to function as chromosomal origins *in vivo* (Wu et al., 1993b), libraries can be screened more rapidly and efficiently by mass assay of clone pools, using the method described in Chapter 4. Concurrently, a Japanese group worked out a mass screening protocol based on multiple cycles of bromodeoxyuridine semiconservative replication assay, with isolation and cloning of heavy-heavy plasmids from transfected 293S cells (Masukata et al., 1993). Unfortunately, this group presented very little data on the composition of the chromosomal DNA fragment panels used as the starting point for their screening protocol, making it impossible to assess the sensitivity and specificity of their origin purification procedure. Nevertheless, their results are consistent with some of the observations made in Chapter 4; human autonomously replicating sequences function with differing efficiencies, and activity depends on the presence of specific sequence components rather than on simple fragment size (Masukata et al., 1993).

Minimal, core origin function is contained within cloneable sequence modules, and can be shuttled between different plasmid vectors (Chapter 7). Known autonomously replicating sequences can be subcloned and tested (Chapter 5), to determine the minimal sequences necessary for activity. This opens the door to careful mutational analysis of mammalian origins to determine the nature and arrangement of functional sequence modules, much as was done with yeast ARS elements (Marahrens and Stillman, 1992).

Isolation of significant numbers of replication origins (Chapters 4 and 6) from one system (the human) would be a valuable aid in the recognition of any minimal consensus DNA element that may be required for origin function, particularly if such a consensus is, like the yeast ARS core, relatively degenerate, or if there exist multiple subtypes of replication origins with different core elements. Subclones which retain the capacity for autonomous replication (Chapter 5) can help delimit the required comparisons; sequence and statistical analysis for the presence of known consensus elements (Chapter 6) may facilitate alignment of otherwise disparate sequences, and mutational analysis could identify potentially critical areas. Thus, techniques described in this thesis can be applied and extended to overcome some of the obstacles (Chapter 1) that may be impeding the identification of a human replication origin core consensus sequence.

In *S. cerevisiae*, the origin core consensus sequence provided a tool for the isolation of the ORC initiator complex (Bell and Stillman, 1992), and the same strategy may eventually be applied in humans. However, even in the absence of an available core

consensus as a starting point, work directed at identifying a human initiator protein can begin, employing some of the strategies described in this thesis. A 186 bp subclone of the autonomously replicating sequence *ors*8 (Todd et al., 1995) has been used as a probe in bandshift assays to identify proteins, present in the cell extracts used for *in vitro* replication (Chapters 3, 4, 5 and 6), which specifically bind to this putative minimal origin (Ruiz et al., 1995). An *ors* binding activity, composed of a 146 kD and 154 kD protein doublet, cofractionates with the known replication proteins RP-A, topoisomerase II, and DNA polymerases α and δ. In this work, I demonstrated that the HeLa cell extract fraction most highly enriched for the *ors* binding activity still retains its capacity to permit *in vitro* replication of autonomously replicating plasmids, indicating that an initiator activity is present (Ruiz et al., 1995). Thus, mammalian autonomous replication assays are providing a tool not only for the identification and analysis of replication origins, but also for the isolation of a potential human initiator protein. DNA footprinting techniques can then be applied to establish the nature and pattern of initiator-origin interactions. An initiator can be used as a basis for examining other *trans*-acting factors responsible for activation of DNA replication, and could establish connections between the fundamental process of DNA replication, cell cycle controls and checkpoints, and oncogenesis.

Cis-acting sequences may contribute to the control of replication timing within S phase. New data from the Fangman group suggests that, in yeast, cloneable elements associated with ARS sequences may function to delay activation of an origin to a point late in S (Brewer et al., 1993; Huberman, 1995). In the human β-globin locus, replication timing is influenced by the choice of which origins are activated (Kitsberg et al., 1993), and new results from the Wahl group suggest that the transcriptional locus control region, 60 kb away from the region's internal replication origin, is required for origin activation and early replication of the gene cluster (Huberman, 1995). Replication timing is also regulated at imprinted genetic loci, which display the property of allele-specific replication timing asynchrony (Chapter 1). While the predominant pattern is that the expressed imprinted allele replicates earlier than the non-expressed allele, the opposite is true at the murine *H19*, murine *Igf2r* (Nicholls, 1994), and possibly the human *XIST* (Hansen et al., 1995) loci; thus, transcriptional status is not the sole determinant of origin activation and timing in imprinted chromosomal regions. Until the identification of putative origin p82 (Chapter 6), derived from a 60 kb maternal-early replication timing subdomain in the human GABA receptor gene cluster of the imprinted 15q11q13 chromosomal region (LaSalle and Lalande, 1995), no replication origin had been characterized involving an imprinted locus. Thus, p82 provides a valuable focus for studies on replication timing, genomic imprinting, and origin function. Subsequent experiments can be directed toward *in vivo* mapping (by

nascent strand PCR) of chromosomal initiation sites near p82, and analysis of possible differential origin usage on maternal versus paternal alleles in normal and Angelman syndrome cells. Eventually, characterization of differences in methylation and protein binding footprints associated with p82 may help identify factors important in creating or maintaining the imprinted, asynchronously-replicating state. Allele-specific replication timing control colocalizes with a putative imprinting control element in the 5' *SNRPN* gene, from the Prader-Willi syndrome critical region in the proximal portion of 15q11q13 (Gunaratne et al., 1995), 1000-1400 kilobases away from p82. While the Angelman syndrome gene may in fact lie 200 to 400 kb proximal to p82, if a replication origin can function as part of an imprinting control element, the biology of a regional replication origin may influence the etiology of Angelman syndrome (Lalande, 1994).

Origins function in temporally-synchronous clusters to replicate entire chromosomal subdomains (Chapter 1), and the existence of p82 in a region of transition between different allele-specific timing patterns suggests that it may lie near the border of such a subdomain. Few methods exist for the study of replicon cluster regulation, even though disruption of replicon function may contribute to aging or to malignant transformation (Chapter 5). Chapter 7 demonstrates that circular YAC vectors containing cloned origins can be maintained as episomes. Considering the versatility of YAC vectors, it may well be possible to clone in larger genomic regions and to analyze replication timing patterns in entire domains and at domain borders. In this fashion, even early stage construction of a human artificial chromosome can provide tools to help in the understanding of the higher order organization and function of chromosomal replication origins.

The cloning and assay methods described in Chapter 7 provide a basis for constructing a human artificial chromosome. As in the construction of yeast artificial chromosomes (Chapter 1), the next step would be the addition of human telomere ends to create a linear acentromeric chromosome, the perfect vector for cloning a human centromere, although large putative centromere sequence blocks can also be cloned directly into the Y.343, YAC.S3, and Y.X24 constructs (this work is currently being attempted by a collaborating group at the University of Calgary). In addition to its role as a tool for the study of origin and centromere biology, a human artificial chromosome could function as a gene vector with no practical limit on the size of inserted DNA, no need for viral components, and no requirement for integration into the host genome. If efficient methods for transfecting such constructs into human cells are available, a human artificial chromosome could well see use in somatic cell gene therapy.

REFERENCES

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

Bell D, Sabloff M, Zannis-Hadjopoulos M, Price GB (1991) Anti-cruciform DNA affinity purification of active mammalian origins of replication. *Biochim. Biophys. Acta* 1089:299-308

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

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

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

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

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

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

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

Dijkwel PA, Hamlin JL (1995) The Chinese hamster dihydrofolate reductase origin consists of multiple potential nascent-strand start sites. *Mol. Cell. Biol.* 15:3023-3031

Dimitrova D, Vassilev L, Anachkova B, Russev G (1993) Isolation and cloning of putative mouse DNA replication initiation sites: binding to nuclear protein factors. *Nucl. Acids Res.* 21:5554-5560

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

Gilbert DM, Miyazawa H, Nallaseth FS, Ortega JM, Blow JJ, DePamphilis ML (1993) Site-specific initiation of DNA replication in metazoan chromosomes and the role of nuclear organization. *Cold Spring Harbor Symp. Quant. Biol.* LVIII:475-485

Gunaratne PH, Nakao M, Ledbetter DH, Sutcliffe JS, Chinault AC (1995) Tissue-specific and allele-specific replication timing control in the imprinted human Prader-Willi syndrome region. *Genes Dev.* 9:808-820

Hamlin JL, Mosca PJ, Levenson VV (1994) Defining origins of replication in mammalian cells. *Biochim. Biophys. Acta* 1198:85-111

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

Hansen RS, Canfield TK, Gartler SM (1995) Reverse replication timing for the XIST gene in human fibroblasts. *Hum. Mol. Genet.* 4:813-820

Huberman JA (1995) Prokaryotic and eukaryotic replicons. Cell 82:535-542

Iguchi-Ariga SMM, Itani T, Kiji Y, Ariga H (1987) Possible function of the c-myc product: promotion of cellular DNA replication. *EMBO J.* 6:2365-2371

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

Jacob F, Brenner S (1963) Sur la régulation de la synthèse du DNA chez les bactéries: l'hypothèse du réplicon. *Comp. rend. acad. sci.* 256:298-300

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

Kelly R, DeRose M, Draper BW, Wahl GM (1995) Identification of an origin of bidirectional replication in the ubiquitously expressed mammalian CAD gene. *Mol. Cell. Biol.* 15:4136-4148

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

Lalande M (1994) In and around SNRPN. *Nature Genet.* 8:5-7

LaSalle JM, Lalande M (1995) Domain organization of allele-specific replication within the GABRB3 gene cluster requires a biparental chromosome 15q11-q13 contribution. *Nature Genet.* 9:386-394

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

Masukata H, Satoh H, Obuse C, Okazaki T (1993) Autonomous replication of human chromosomal DNA fragments in human cells. *Mol. Biol. Cell* 4:1121-1132

McWhinney C, Leffak M (1988) Episomal persistence of a plasmid containing human c-myc DNA. *Cancer Cells* 6:467-471

Nicholls RD (1994) New insights reveal complex mechanisms involved in genomic imprinting. *Am. J. Hum. Genet.* 54:733-740

Pearson CE, Frappier L, Zannis-Hadjopoulos M (1991) Plasmids bearing mammalian DNA-replication origin-enriched (ors) fragments initiate semiconservative replication in a cell-free system. *Biochim. Biophys. Acta* 1090:156-166

Pearson CE, Shihab-El-Deen A, Price GB, Zannis-Hadjopoulos M (1994) Electron microscopic analysis of in vitro replication products of ors 8, a mammalian origin enriched sequence. *Somat. Cell Mol. Genet.* 20:147-152

Ruiz MT, Pearson CE, Nielsen T, Price GB, Zannis-Hadjopoulos M (1995) Cofractionation of HeLa cell replication proteins with ors-binding activity. *J. Cell. Biochem.* 58:221-236

Shihab-El-Deen A, Wu C, Hamerton JL, Zannis-Hadjopoulos M, Price GB (1993) Chromosomal localization of a sequence with in vivo activity for initiation of DNA replication. *Somat. Cell Mol. Genet.* 19:103-109

Taira T, Iguchi-Ariga SMM, Ariga H (1994) A novel DNA replication origin identified in the human heat shock protein 70 gene promoter. *Mol. Cell. Biol.* 14:6386-6397

Todd A, Landry S, Pearson CE, Khoury V, Zannis-Hadjopoulos M (1995) Deletion analysis of minimal sequence requirements for autonomous replication of ors8, a monkey early-replicating DNA sequence. *J. Cell. Biochem.* 57:280-289

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

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

Wu C, Friedlander P, Lamoureux C, Zannis-Hadjopoulos M, Price GB (1993a) cDNA clones contain autonomous replication activity. *Biochim. Biophys. Acta* 1174:241-257

Wu C, Zannis-Hadjopoulos M, Price GB (1993b) In vivo activity for initiation of DNA replication resides in a transcribed region of the human genome. *Biochim. Biophys. Acta* 1174:258-266