Gene Ther Mol Biol Vol 6, 33-46, 2001

The Hitchhiking principle: Optimizing episomal vectors for the use in gene therapy and biotechnology

Review Article

Jürgen Bode[1*], Christian P. Fetzer[2], Kristina Nehlsen[1], Monica Scinteie[2], Bok-Hee Hinrichsen[2], Armin Baiker[2#], Christoph Piechazcek[2¨], Craig Benham[3] and Hans J. Lipps[2]

[1] German Reasearch Institute for Biotechnology, Braunschweig, Germany

[2] Institute of Cell Biology, University Witten/Herdecke, Witten

[3] Department of Biomathematical Sciences Mount Sinai School of Medicince New York, NY

***Correspondence**: Jürgen Bode, Ph.D., German Reasearch Institute for Biotechnology, D- 38124 Braunschweig, FRG, Mascheroder Weg 1; Tel./Fax: +49 531 6181 251/262; email: jbo@gbf.de

#present address: Stanford University Department of Pediatrics, 300 Pasteur Drive,Grant Bldg. Room S 356, Stanford, CA 94305-5208.

¨present address: AmCell D-51429 Bergisch Gladbach, Technologiepark H 13

Key words: Hitchhiking principle, gene therapy, biotechnology, replication origins, mammalian viruses, Maintenance elements, Replication–transcription coupling, nonviral episomes

Abbreviations: auxiliary elements, (Aux); base-unpaired region, (BUR); cyclin-dependent class, (CDKs); DNA unwinding element, (DUE); duplex destabilization, (SIDD); dyad symmetry, (DS); Epstein-Barr viral nuclear antigen 1, (EBNA-1); Epstein-Barr virus, (EBV); family of repeats, (FR); inverted repeats, (IRs); lyric origins, (ori Lyt); minichromosome maintenance proteins, (Mcms); origin recognition complex, (ORC); origin recognition element, (ORE); origins, (ORIs); papilloma virus, (BPV); parent vector, (pC1); polyoma virus, (PyV); Proliferating-cell nuclear antigen, (PCNA); replication fork barrier, (RFB); replication protein A, (RPA); scaffold/matrix attached region, (S/MAR)

Received: 24 January 2001; accepted: 29 January 2001; electronically published: February 2004

Summary

We have recently introduced an episomally replicating vector the function of which depends on the combination of SV40 origin of replication with a human scaffold/matrix attached region (S/MAR). The episomal status of this vector is maintained in several cell lines for an extended period of time in the absence of a virally encoded protein and in the absence of selection conditions. In this article we start to identify the elements required for recruiting this type of episome to the endogenous cellular replication machinery and we discuss aspects of replication-transcription coupling. We try to establish a catalogue of parameters which should be considered for the design of functional episomes.

I. Introduction

Considerable efforts have been undertaken in the past years to construct vectors which replicate episomally in higher eukaryotic cells. While in bacteria and yeast the structure and regulation of replication origins (ORIs) is rather well understood and origin functions can clearly be assigned to certain sequence elements this is not true for chromosomal origins of mammalian cells (De Pamphilis 1993, Boulikas, 1996; Kelly and Brown, 2000). Understanding the control of replication is relevant both for the purpose of academic and applied research. Failure in the control of replication may result in cell death and is one of the primary reasons leading to cellular transformation and uncontrolled growth. From a practical point of view episomal replicating vectors provide several advantages over the classical systems (Calos 1996).

With the present technology targeted integration is still no routine (Bode et al, 2000b) and the conventional alternative, random integration, may lead to insertional mutagenesis with unpredictable consequences. For the same reasons expression of the transgene can not be controlled since it is dependent on the chromatin context of the integration site (Baer et al, 2000). High level expression can only be achieved at favorable genomic loci but the danger exists that integration into highly expressed sites interferes with vital cellular functions. In addition, there is increasing evidence for the existence of cellular defense mechanisms against foreign DNA which operate by down-regulating transgenes in a process that is accompanied by DNA methylation (Bingham 1997; Garrick et al, 1998). As discussed below, episomal vectors are intrinsically free from all these disadvantages.

A number of DNA viruses, such as SV40, BPV or EBV replicate episomally in mammalian cells. These processes depend both on viral trans-acting factors and on accessory activities recruited from the host cell replication machinery. For their segregation DNA viruses apply a "hitchhiking principle" (Calos, 1998), i.e. they acquire centromere functions by associating with host chromosomes. Since the replication origins of these viruses are well characterized, they represent convenient tools for the study of the associated elements and the relation between ongoing transcription and replication. Still, vectors derived from these minimal systems require at least one viral protein, the large T-antigen in case of SV40, E1/E2 for BPV or EBNA-1 for EBV which usually restrict their replication to a narrow host range. Since transformation of recipient cells is an inevitable consequence of these factors the development of a new vector generation is desirable in which their function is replaced by components of the endogenous cellular replication machinery.

Such a vector, pEPI-1, which is based on the SV40 origin of replication but is independent on the virally encoded large T-antigen for replication has recently been introduced (Piechaczek et al, 1999, Baiker et al, 2000). pEPI-1 contains a well characterized S/MAR from the human interferon-β gene (Element I in Mielke et al, 1990), it replicates episomally in a variety of mammalian cell types and is stably maintained and segregated over several hundred generations even in the absence of selection. Further facets of potential relevance are the two transcription units present on this construct (**Figure 3**). All these factors may contribute to the properties of this vector but the exact functions and interactions of the participating elements still have to be determined..

In this review we will describe the role of modules in minimal eukaryotic origins of replication (ORIs) as well as possible interrelations between transcription and replication. We will discuss how we can take advantage of this knowledge to design novel non-viral episomal vector systems of wide applicability.

II. Common structural features of replication origins

It seems that among all terrestrial life forms the activation of replication origins shares some common fundamental characteristics. First, an initiator protein, such as dnaA in E. coli, T antigen in SV40 or polyoma, and EBNA1 in Epstein-Barr virus, directs the formation of a large complex containing several initiator protein molecules. As a prerequisite, most ORI sequences possess AT-rich stretches, so called base-unpairing regions (BURs), which undergo strand separation under superhelical tension (see chapter IIB below). Most ORIs also harbor a significant number of nearby tri, tetra and higher nucleotide repeats which can recruit the energy stored in a BUR in order to induce a defined secondary structure (hairpin, stem-loop) which in turn is the recognition feature for the initiator protein complex. Frequently, these repeats are parts of an element which has been termed a "DUE" (DNA unwinding element) and which can be characterized by its localized reactivity towards single strand specific agents. While the function of a base unpairing region may certainly overlap that of a DUE, this is not necessarily the case (**Figure 1**): in the context of the present article, a BUR is a box that is composed of multiple destabilized sites and which is hence defined by its duplex destabilization (SIDD) parameters as introduced in Benham et al, (1997) (**Figure 4**). In contrast, a DUE depends on repetitive sequence features. Once formed, ssDNA binding proteins have to interact with the induced secondary structures before the DNA polymerase/primase complex is able to initiate DNA replication. In this view a continuously reciprocating transition may exist between extended single strands and DNA forms which invest the energy stored in BURs in the formation of alternative structures.

For eukaryotes, competence to initiate DNA replication in G1 depends on the ordered assembly of multiprotein complexes at the ORI. Once assembled, the competent complexes initiate DNA synthesis, triggered by protein kinases of the cyclin-dependent class (CDKs) and the Cdc 7 family. In addition to initiating replication, CDKs have a role in preventing re-initiation within a given cell cycle. Competence is restored by the elimination of CDK activity during metaphase (review: Kelly and Brown, 2000). This CDK-driven switch explains why the default state of an eukaryotic cell is the one in which DNA is replicated once and only once per cell cycle (**Figure 2**).

A. Yeast: ARS

A glimpse of the complexity of eukaryotic origins arises from studies on ARS activation in the budding yeast Saccharomyces cerevisiae. This unicellular lower eukaryote does not have to cope with the complexity of multicellular organisms that need to develop a specific program of silencing sets of ORIs during embryogenesis. Nevertheless, its ORIs appear to be differentially activated

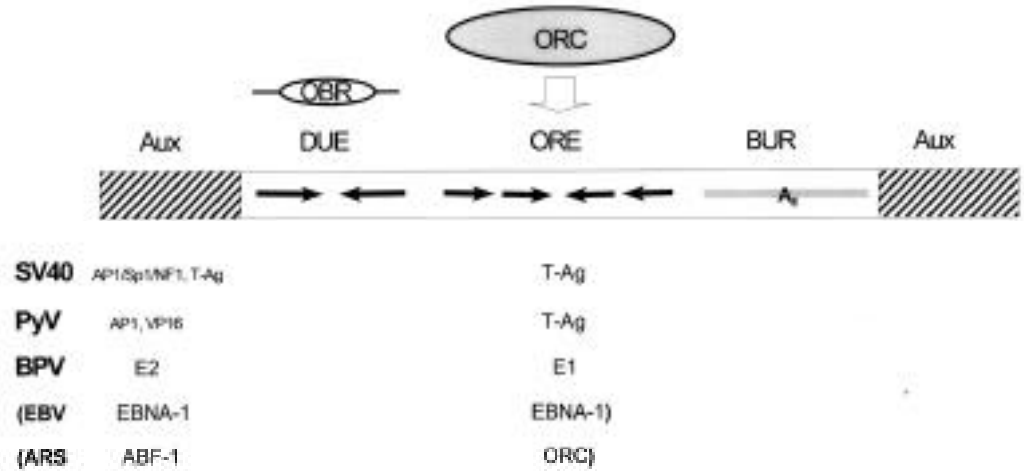

Figure 1. Anatomy of an eukaryotic minimal origin of replication, exemplified by SV40. Replication initiation is usually supported by an easily melting DNA tract, the so-called base-unpairing region (BUR; cf. Figure 4). The energy absorbed by base-unpairing can be delivered to a DNA unwinding element (DUE) and used to establish secondary structures such as a hairpins or a stem-loop (cf. Figure 2). These are the prerequisites for the origin recognition complex (ORC) to associate with the origin recognition element (ORE) and to initiate replication. The process gains additional support from transcription factors which may associate with auxiliary elements (Aux). The polyoma (PyV) and papilloma virus (BPV) origins conform to the same general design whereas related features are found in case of Epstein-Barr virus (EBV, Figure 2, top) and the yeast autonomously replicating sequences (ARSs, Figure 2, bottom).

during S-phase intervals (Ferguson et al, 1991). Yeast ORIs could be cloned using an in vivo selection system based on autonomous replication. The responsible sequences were called autonomously replicating sequences or ARS elements as they permit replication of the selected gene as an extra-chromosomal plasmid. The properties of ARS-containing plasmids were consistent with the function of genomic ORIs, in that their replication occurs in the nucleus, only once during the S-phase, and in that it requires the same genes as does chromosomal replication.

ARS elements contain a 15-35 bp binding site for origin recognition proteins, the ORE (origin recognition element) which comprises a conserved 11 bp AT-rich tract (ARS consensus sequence, ACS, $^{T}/_{A}TTTA^{T}/_{C}{}^{A}/_{G}TTT^{T}/_{A}$). This tract represents the binding site for ACS binding proteins (parts of the origin recognition complex, ORC), which are constituents of the nuclear matrix from which they can be recovered by extraction. The ACS forms the core of a larger functional sequence called element "A". In addition to A, a region called "B", located downstream of the T-rich strand of the ACS, is also required for ARS function. In several or all ARSs, domain B is easily unwound. A functional role for a BUR/DUE system localized within domain B is supported by mutations which reduce the ease of unwinding and thereby replication activity. It is also suggested by the substitution of domain B-sequences with dissimilar sequences that are easily unwound (Huang and Kowalski 1996). While the A element plays the major role in specifying a sequence as a replication origin, the B region elements serve to enhance the efficiency of origin utilization (**Figures 1, 2**).

Usually, the B-region can be further subdivided into 2 or 3 essential subregions which may overlap in function: while single mutations in any subregion reduce replication activity, replication is still observed as long as a a single B-element is left. In contrast to the A element, the B-region is not conserved among different ARSs. In case of ARS1, element B3 contains a characterized binding site for the ABF1 transcriptional regulatory protein. While binding of ABF1 to several ARS elements stimulates initiation of replication, the factor also contributes to either repression or activation of transcription, (Diffley and Stillman, 1988, 1990). Its function can be substituted by other transcriptional regulators such as RAP1 or GAL4 if they are provided with both, DNA binding and activation domains to stimulate DNA synthesis, and if their recognition sites are adapted to region B sequences.

The protein complex specifically recognizing the ARS core region was originally identified by DNA footprinting. This complete "origin recognition complex" (ORC), consists of 6 polypeptides which protect the ACS throughout the cell cycle. During Gl, the protection over the ACS extends to adjacent nucleotides, a result suggesting that either the ORC undergoes significant conformation alteration prior to S phase, or - more plausibly - that additional proteins become associated with it. Most likely, these alterations are an imprint by Cdc6 and the Mcm (minichromosome maintenance) proteins which 'reset' chromatin for another round of DNA replication early in the cell cycle. Mcm2-7 family members are highly conserved in the eukaryotic kingdom. They are nuclear proteins which form several types of oligomeric complexes some of which have ATPase and

weak helicase activity. Chromatin, thus 'licensed' for replication, is guided into the S phase by the activation of cell cycle-regulated protein kinases. Upon entry into S phase, the pre-replication complex is partially dissolved, first by the dissociation of Cdc6 and then by a gradual release of Mcm proteins. This process appears to be accompanied by a recruitment of chain elongation factors and the establishment of replication forks (Kelly and Brown, 2000).

While ARS plasmids replicate efficiently, some authors state that they still lack the sequences needed for proper and efficient segregation into daughter cells during mitosis (dePamphilis, 1993). A more or less random segregation of the replicated plasmid at division is thought to produce cells with multiple copies and some with none. Since cells with multiple copies have a selective advantage they would accumulate with time. On the other hand, S/MAR-characteristic of yeast ARSs have been noted which could well serve a maintenance function (Amati and Gasser, 1990).

Metazoan ORIs may in fact differ from those of yeast because their nuclei become disassembled during mitosis whereas they remain largely intact in fungi. This puts emphasis on nuclear retention activities which are strictly needed to establish and maintain extrachromosomal DNA in higher eukaryotes.

B. Minimal ORIs in mammalian viruses

Activation of yeast ARS sequences seems to be more complicated than structural distortions that are caused by a simple complex formed by a unique protein such as the large T antigen in SV40, or the E1/E2 complex in Bovine papilloma. It is now believed that an eukaryotic cell can

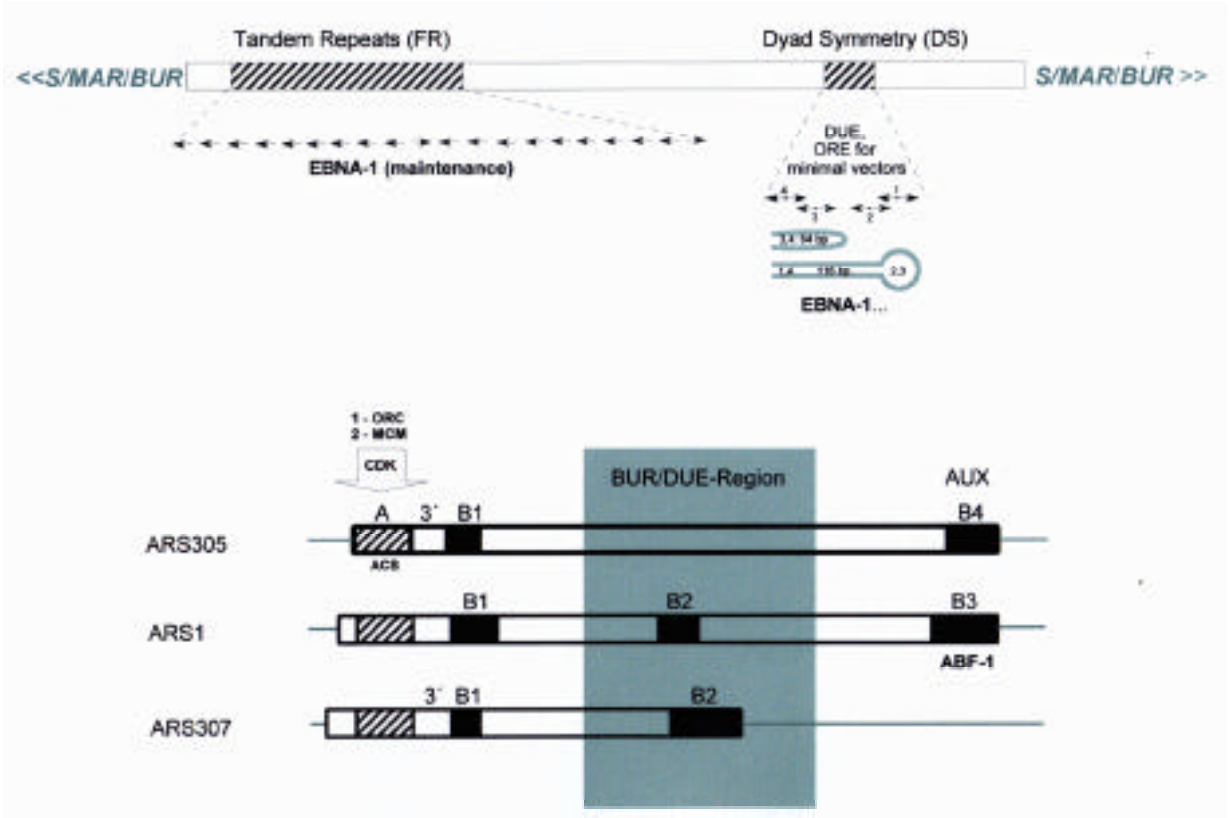

Figure 2: Some particular features of the Epstein Barr virus-oriP (top) and of yeast ARSs (bottom) For EBV a dyad symmetry element (DS) serves the function of a DUE and (sometimes) ORE. Only in case of artificial BPV-ori constructs EBNA-1 has a function at the DUE which otherwise seems to attract ORCs from the host cell (Norio et al, 2000). On the other hand, there is a massive association of EBNA-1 with the family of repeats (FR) segment which, in its presence, serves the function of a maintenance element. Unwinding potential is probably provided by the surrounding S/MAR sequences (Jankelevich et al, 1992) which can serve as highly efficient BURs (Benham et al, 1997). All ARS elements are characterized by an element "A" harboring the 11 bp AT rich ARS consensus sequence (ACS). ACS is the principal binding site of the ORC. ORC is followed by Cdc6 (a protein necessary for origin firing (Cocker et al, 1996) and the minichromosome maintenance proteins (Mcms) which together form the initiation complex (review: Kelly and Brown, 2000). Element "A" is supported by one or more B-elements of various composition which mark segments with DNA unwinding potential (Huang and Kowalski, 1996). Whether or not replication in yeast requires maintenance functions has remained a controversial matter.

replicate any sequence which contains the appropriate signals for initiation in a once per cell cycle manner. Therefore, an unexpected complication arose from the observation that some DNA viruses seem to have developed strategies to overcome this stringent copy number regulation. An notable exception is the latent Epstein-Barr virus.

1. SV40

As any origin of replication, the SV40 ORI consists of multiple modules (**Figure 1**). Although flanking regulatory sequences of the early promoter facilitate replication, a 64 bp core is sufficient for initiation. A 20 bp motif, 5′TGCATAAATAAAAAAAATTA-3′, forms one T-antigen binding site. Its continuos tract of 8 adenines is highly conserved also among polyoma viruses as it is the prerequisite for DNA bending. Bending has been shown to facilitate strand separation of the AT tract and it is thought that DNA supercoiling and the 27 bp T-antigen binding palindrome cooperate with the A-tract to destabilize the origin. SV40 large T antigen (T-Ag) is the only viral protein required for replication. The host provides all other replication proteins, among these the ssDNA binding protein (SSB), DNA polymerase-primase complex, topoisomerases I and II, RNAseH, a 5′-> 3′exonuclease, Proliferating-cell nuclear antigen (PCNA) and PCNA-dependent DNA polymerase.

An initial event in the replication pathway is ATP-dependent binding of T-Ag molecules to four GAGGC repeats which are part of a palindrome within the ORE (**Fig**. **1**). In the presence of ATP, 12 molecules of large T become assembled as two hexamers on the SV40 core ORI (Valle et al, 2000). These hexamers melt the early and untwist the late half of the core ORI. These process releases T-Ag from the pentanucleotides and permits its action as a helicase at the flanking AT tract.

The SV40 ORI region overlaps promoter regions for the SV40 early genes and mutational studies demonstrated that transcription factor binding greatly enhances the efficiency of replication. While the SV40 replicon may be unusual in that it is a highly compact genetic element it has been observed that this and other ORIs are in close proximity to transcription factor binding sites. The stimulation of replication afforded by these auxiliary elements is generally ascribed to several contributing factors:

(i) An altered chromatin structure: Factor NF1 binds adjacent to the SV40 origin and prevents nucleosome formation in the ori region thereby allowing more efficient binding of T-Ag and replication proteins. This contribution stimulates replication at least 20-fold.

(ii) Direct protein-protein contact: transcription factors are thought to interact with essential replication proteins and to enhance their activity. Both DNA binding and acidic activation domains are required for these stimulatory effects.

(iii) Local DNA unwinding or distortion: transcription factors stabilize weak interactions with replication proteins by causing limited unwinding or altering local structure of the DNA to favor replication protein-DNA interactions.

2. BPV

Papillomaviruses are members of the small DNA tumor virus family. Their mode of replication is closely coupled to the differentiation status of the infected epithelium. In the basal and parabasal cells, the virus is maintained as a low copy number extra-chromosomal episome and undergoes regulated DNA replication modulated by both viral and host proteins. As cells undergo progressive differentiation, vegetative viral replication is triggered, "late" viral genes are expressed, and progeny virions are produced in a fraction of the terminally differentiated cells in papillomas.

The latent stage of papillomaviral replication provides an ideal system for the study of regulated eukaryotic DNA replication. The BPV ORI appears similar in organization to that of SV40, with a 12 bp AT-rich sequence at one end and a palindrome at the other that constitutes the primary binding site for the BPV E1 protein. E1 is a functional homolog of SV40 large T antigen, with origin binding potential as well as ATPase and helicase activity. E1associates as a trimer or a hexamer on its cognate E1-binding site in the viral origin with relatively low initial affinity and sequence specificity. In the replication competent form of BPV, E1 forms a dihexameric complex which may be stabilized by the Hsp40 chaperone, mirroring the structure of SV40 T antigen on the SV40 ORI. In addition to its role in initiation, HPV-11 E1 is also required during elongation in vitro, suggesting that its helicase activity may be critical at the replicating forks.

E2 is a viral transcriptional transactivator that is also essential for viral DNA replication in vivo. It binds as a dimer with high affinity to its conserved binding sites in the viral genome, including several sites in the viral origin of replication. One of the critical functions of E2 in replication is to interact with and recruit E1 to the viral ORI by virtue of its stronger DNA binding affinity and specificity for E1. Based on these data the following model of E2/E1 interaction during initiation of bovine papillomaviral DNA replication has been proposed: once the first molecule of E1 is loaded onto the origin by E2, E2 is released from the origin, allowing E1 to multimerize into a replication-competent form. The role of E2 may extend beyond the recruitment of E1 as it was found to contribute to the formation of the entire pre-initiation complex, but it is dispensable during elongation.

3. EBV

Within latently infected human B-lymphocytes the circular 165 kb chromosome of Epstein-Barr virus (EBV) is maintained as a large episome. Synthesis and

maintenance of this episome is mediated by a viral cis-acting sequence, oriP, and a single viral protein, the Epstein-Barr viral nuclear antigen 1 (EBNA-1). On latent EBV chromosomes replication initiates at multiple sites including a 1.8-kb region called oriP, which is sufficient for both replication and stabilization of recombinant plasmids in the presence of EBNA- 1. In such a minimal system replication depends on multiple EBNA-1 binding sites and it is initiated at or near the dyad symmetry component which contains two 46 bp protected regions each encompassing two paired core binding sites.

Altogether, there are 24 binding sites for EBNA-1 within ORIP. These sites are organized into two clusters, 1000 bp apart, referred to as the dyad symmetry element (DS) and the family of repeats (FR-element). The relative orientation of DS and FR can be altered without affecting oriP function which is also barely affected by yet another locus, BamHI-Q, with two additional low-affinity EBNA-1 bindings sites (Rawlins et al, 1985)

FR consists of 20 copies of a 30 bp repeat unit each of which represents a high affinity site for EBNA-1. This interaction is able to prevent plasmids from being lost from mitotically active cells, it serves as replication enhancer and also as a potent EBNA1-dependent transcriptional enhancer. In addition to sequence-specific DNA binding, EBNA-1 molecules interact efficiently with each other by a DNA looping mechanism (Laine and Frappier, 1995), by which they link the various binding sites. These interactions lend EBNA-1 properties which are otherwise typical of nuclear matrix proteins such as SAF-A and RAP-1 and, as a consequence, it might stabilize binding of EBNA-1 to the DS element or it might attract the cellular replication machinery (Frappier and O'Donnell, 199l).

Named for the dyad symmetry it contains, the DS element, has four overlapping, palindromic binding sites of intermediate affinity for EBNA-1. *In vivo* footprinting studies have proven that EBNA1 is the only protein interacting with these sites. Considering the dimeric nature of EBNA-1, and the importance of the precise spacing between the palindromic halves, it is likely that the dimer is the associating species. EBNA-1 is closely associated with cellular chromatin as it is uniformly distributed over all chromosomes in metaphase spreads (Reedman and Klein, 1973). By itself, it does not appear to melt origin DNA but it induces localized changes, visible as sensitivity to permanganate oxidation, at two of the four DS sites, which may be the prerequisite for origin DNA melting.

The core of the EBNA-1 DNA binding region is virtually identical in structure and positioning on DNA with the DNA-binding domain of BPV-E2 protein. A distinguishing feature is an N-terminal extension which mediates several sequence-specific DNA contacts along the minor groove. While this extended chain is not immediately required for DNA binding, it is responsible for permanganate sensitivity: this structure appears to produce DNA overwinding in two ratchet-like events enforcing a distorted B-helix that is bent around the protein core. EBNA-1 appears to lack the enzymatic activities that are present in ORI-binding proteins of other mammalian viruses and it does not seem to interact with human cellular proteins that provide the equivalent functions. It has even been demonstrated that it is dispensable for the synthesis of oriP plasmids. However, in its absence, newly synthesized oriP plasmids are lost rapidly from proliferating cells indicating the existence of elements that permit retention of replicated DNA in daughter cells. Although the precise role of EBNA-1 in this process has remained elusive, it is now agreed that its main function occurs post-synthetically to ensure plasmid maintenance and segregation in dividing cells.

Unlike plasmids of other viruses but akin to human chromosomes, ORIP plasmids are synthesized once per cell cycle and are partitioned faithfully to daughter chromosomes during mitosis. Aiyar et al, (1998) have found that oriP is recognized directly by the human DNA synthesizing machinery, indicating why, unlike most other viral origins, oriP is replicated once per cell cycle and in synchrony with cellular chromosomes. By all these criteria, replication of the oriP replicon differs substantially from the replication of other viral replicons and it has therefore been used as a paradigm for the function of chromosomal ORIs. Its action is clearly different from large the T-Ag of SV40 and the E1 of BPV described above which possess an ATP-hydrolysis-dependent DNA helicase activity and which interact directly with cellular proteins involved in initiation of DNA synthesis to recruit them to the viral origin. These observations suggest a mechanism by which the SV40 and BPV but not EBV replicons bypass the cellular mechanisms that restrict chromosomal and ORIP DNA synthesis to a single round per cell cycle.

As with initiation zones on human chromosomes, in EBV initiation occurs more often at some regions than at others. A major zone exists adjacent to the terminal repeats resembling the zones at chromosomal mammalian loci. Initiation within the zone is likely to be determined by interactions with cellular proteins. It is not known what determines extended initiation zones but this property is certainly contributed by the host cell rather than the viral protein. Therefore, a functional redundancy of sites with the potential to serve origin functions is a common feature of genomic DNA replication in the mammalian nucleus.

III. Conserved properties in higher origins

A. AT-Rich stretches in ORIs

AT-rich regions of varying size are omnipresent components of origins of replication. They may flank the core ORI, as, for example, in EBV where the core is defined as the short sequence where the initiator protein causes local distortion in the double helix to initiate DNA

unwinding. More often the AT-rich tract is situated between two binding sites of the initiator protein.

Five different functions have been assigned to AT-rich stretches (Boulikas, 1996). The most conspicuous role is their property to facilitate DNA unwinding catalyzed by helicases. A DNA unwinding element is already present in the origin of E. coli in form of a $GATCT_nTT_nTTTT$ tract which is thermodynamically unstable, as evidenced by its sensitivity to the single-strand-specific mung bean nuclease (Kowalski and Eddy, 1989). Therefore, AT tracts become unwound caused by the binding of the initiator proteins to the flanking region(s). ssDNA-binding proteins then interact with the melted portion and attract the DNA polymerase-primase complex. Second, AT tracts are typical constituents of S/MARs which in higher eukaryotes guide the ORI to the replication factories which are organized by the nuclear matrix. A DNA unwinding potential has been ascribed to rather short AT-rich motifs, for instance the AATATATTT tract which is present within the S/MARs of both IgH gene and interferon-γ genes where it becomes the nucleation site for DNA base unpairing under torsional strain (Bode et al, 1992). Third, AT-rich stretches in ORIs might interact with HMG 1 and 2 as in the 50 bp AT-rich stretches of the amplification origins located within the nontranscribed spacer of the murine rDNA (Wegner et al, 1990). Fourth, AT-rich stretches represent the binding sites of a special class of regulatory proteins; for example, yeast ARS elements include the mentioned $^{T}/_{A}TTTA^{T}/_{C}{}^{A}/_{G}TTT^{T}/_{A}$ sequence which here and in ORIs of other species forms the binding site of the ORC protein complex (Diffley and Cocker, 1992). Fifth, if not the actual binding sites of the initiator protein, AT tracts may be the principal place of local distortion of the double helix caused by the binding of the initiator protein to the immediate flanks.

B. BURs, DUEs and IRs

In general, origins of replication require a DNA unwinding element. Adenovirus is only an apparent exception since replication begins at the end of the linear genome where unwinding requires less energy. We have shown above that a DUE is determined by base-stacking interactions rather than AT content. Frequently, DUEs comprise inverted repeats (IRs) which have found their perfection in viral ORIs. They easily convert into cruciform structures when DNA is torsionally strained due to the action of a tracking protein for instance during transcription (RNA polymerase) or replication (DNA polymerase). This process can be supported by retrieving the energy stored in a nearby AT rich base-unpaired region (BUR).

Inverted repeats are of two kinds. They can be quite short (5-20 bp), usually representing the binding sites of initiator proteins. Long IRs are exemplified by a 144 bp perfect inverted repeat in HSV-1 ori_L which is believed to convert into cruciform structures and to act like sinks of torsional strain, to facilitate unwinding of the double helix at the core origin. The role of a special class of inverted repeat-binding proteins and their function in stabilizing DNA in its cruciform structure has been anticipated by Pearson et al, (1994).

Several lines of evidence suggest cruciform formation at the time of activation of an origin of replication (Boulikas, 1996), monoclonal antibodies, directed against cruciforms occurring in ors sequences supposed to represent monkey origins of replication from unknown genes were shown to enhance DNA replication in permeabilized monkey cells. Using monoclonal antibodies directed against cruciform and quantitative fluorescence flow cytometry, 3-5E5 cruciforms/nucleus were estimated for monkey CV-1 and human colon adenocarcinoma SW48 cells throughout S phase while no cruciform-like structures could be detected during G0, G2M or in metaphase chromosomes. Sl nuclease sensitive sites appear as rodent cells move through Gl phase (Collins et al, l982).

A number of studies on ORIs in viral and in higher genomes support the idea that either the origin possesses intrinsically curved DNA or that a severe bent is produced at the origin fragment as a result of its interaction with replication initiator proteins. The fact that origins of replication coincide or colocalize with S/MARs and that S/MARs have been proven to possess intrinsically curved DNA from the retardation in mobility on agarose gels is one additional argument in favor of curved DNA occurring in ORIs (Boulikas and Kong, 1993; Boulikas, 1996).

The EBV-origin does not appear to be an exception to this rule. Based on the known functions of oriP, this region was likely to be situated within or adjacent to a S/MAR which has been verified in an elegant study by Jankelevich et al, (1992). More recently, Mattia et al, (1999) have demonstrated that both the latent (oriP) and one of the lyric origins (ori Lyt) become attached to the nuclear matrix, oriP during the latent cycle of infection and ori Lyt after induction of the lytic cycle.

C. Maintenance elements

The role of nuclear retention functions for the authentic segregation of episomes has been emphasized above for EBV-based plasmids. In that example, retention is supported to a significant extent by the EBNA-protein which binds to its cognate sequences in the FR region. In other examples a strategically positioned S/MAR may mediate at least some of the required interactions. In case of BPV the E1/E2 proteins are believed to be the main contributors to chromosome attachment and episome maintenance (Calos, 1998). Although the participation of a S/MAR has not directly been demonstrated in this case, such a role has become evident during the construction of artificial episomes: the potential of BPV-derived vectors was increased dramatically when a hybrid plasmid (BPV-BV1) was constructed which could be shuttled between E. coli and mouse cells (Di Maio et al, 1981). For this

function it had to contain a 69% subfragment of BPV and a minimum of 2.7 kb eukaryotic "stabilizing sequence" which had been found by trial and error in the large ß-globin intron. Later on we have demonstrated that this sequence coincides with a S/MAR (Klehr and Bode, 1988).

In mammalian cells, ORIs colocalize with S/MARs and become DNAse I hypersensitive during their activation. ORIs are in close proximity to even within the nuclear matrix attachment sites of chromatin loops and a number of studies has conclusively demonstrated that initiation of DNA replication takes place on the nuclear matrix (review: Boulikas, 1996). In addition, elongation of new DNA proceeds by reeling of the old strands through the matrix where the replication forks are anchored.

IV. Replication–transcription coupling

There is evidence that transcription and replication may be antagonistic: transcription appears to prevent replication from initiating within transcribed regions (Haase et al, 1994). We will show below that such a generalization cannot be true and under which conditions transcription and replication may affect each other either in a positive or a negative sense.

A. Contribution of transcription factors

Exemplified by SV40, a single protein may simultaneously regulate two entirely different processes: transcription and replication. Spl stimulates SV40 DNA replication (Guo and DePamphilis, 1992) as does NF-I (also called CTF, or C/EBP). The CCAAT element recognized by NF-I is an important promoter element for a significant number of eukaryotic genes. NF-I as Oct-1 which is involved in the regulation of the histone H2B and immunoglobulin genes can also stimulate initiation of adenovirus DNA replication in vitro. In addition, proteins implicated in the control of DNA replication may include p53, a sequence-specific DNA-binding factor with a GC-rich sequence preference that might interact with the Spl site of SV40 ORI (Boulikas, 1994).

The tight coupling between replication and transcription might simply arise from the fact that most replication factors interacting with the core ORI are actually at the same time transcription factors. Along the same lines, transcriptional enhancers effective in replication may exert this effect by stabilizing replication initiation complexes at the origin core, even in the cases where they are found at a large distance. This is particularly pronounced for the ORIs in multicellular organisms which comprise binding sites for a high number of transcription factors in addition to the binding sites for the replication-specific initiator protein. This level of complexity directs their programming during embryogenesis and their differential replication during S phase which is tightly coupled to gene expression - sometimes in a negative sense (see below). It has also been reasoned that the high number of initiation sites in embryos may reflect a relative transcriptional quiescence of embryonic cells.

At a lower level of complexity, replication in yeast is enhanced by the transcription factor Abf1p (ARS binding factor 1) which associates with a region 3′ to the T-rich strand (B-domain). In the tandemly organized rDNA repeats mentioned above, the RNA Pol I enhancer also contains a weak binding site for Abf1. While binding of Abf1 is inhibited at the nucleosome-packed enhancers 3′ from inactive gene copies, enhancers downstream from active genes are always organized in a nucleosome free structure that is accessible to transcription factors. In this situation Abf1 could attract replication protein A (RPA) which would assist the unwinding process thereby mediating activation of a nearby A domain.

B. Role of transcription direction

In bacteria the majority of strong promoters on the chromosome are oriented such that transcription complexes move away from *oriC* (Brewer 1988) which is plausible in the framework of the twin- domain model (Wu et al, 1988) which postulates that a tracking protein like RNA polymerase leaves behind underwound DNA in its wake (which may be stabilized by negative supercoiling) and causes overwinding in front of it (which may revert to B-type DNA by positive superhelicity). Therefore, an initiation site would become single stranded more easily if transcription would move away from it. For eukaryotes, the scenario is clearly more complicated, possibly because of the rules according to which nucleosomes associate with superhelical DNA. Since DNA is wrapped in a left-handed sense around a nucleosome, this packaging represents a repository of underwound DNA which is released once the protein-DNA interaction is weakened, for instance by nucleosome hyperacetylation (Norton et al, 1989, 1990). As a consequence, nucleosomes will tend to associate with the negative superhelical part of the twin-domain (Wang et al, 1993) while they will be driven off by the approach of a tracking protein (Studitsky et al, 1994).

It has been proposed that transcription through a yeast ARS element affects ARS function in a negative sense (Kipling and Kearsey, 1989). A common model to demonstrate this property is the yeast rDNA locus which consists of a tandem array of 9.1 kb units that are repeated 100-200 times. Although each rARS constitutes a potential ORI, less then one third of rARSs are actually used in a given S-phase (Muller et al, 2000). Recently, replication initiation has been demonstrated exclusively for those rARSs placed immediately downstream of actually transcribed genes. Once an rARS has fired, replication proceeds bidirectionally. The leftward moving fork is stopped at the replication fork barrier (RFB) at the 3´ end of transcriptionally active genes whereas the rightward-moving fork proceeds through about 5 repeats. In contrast

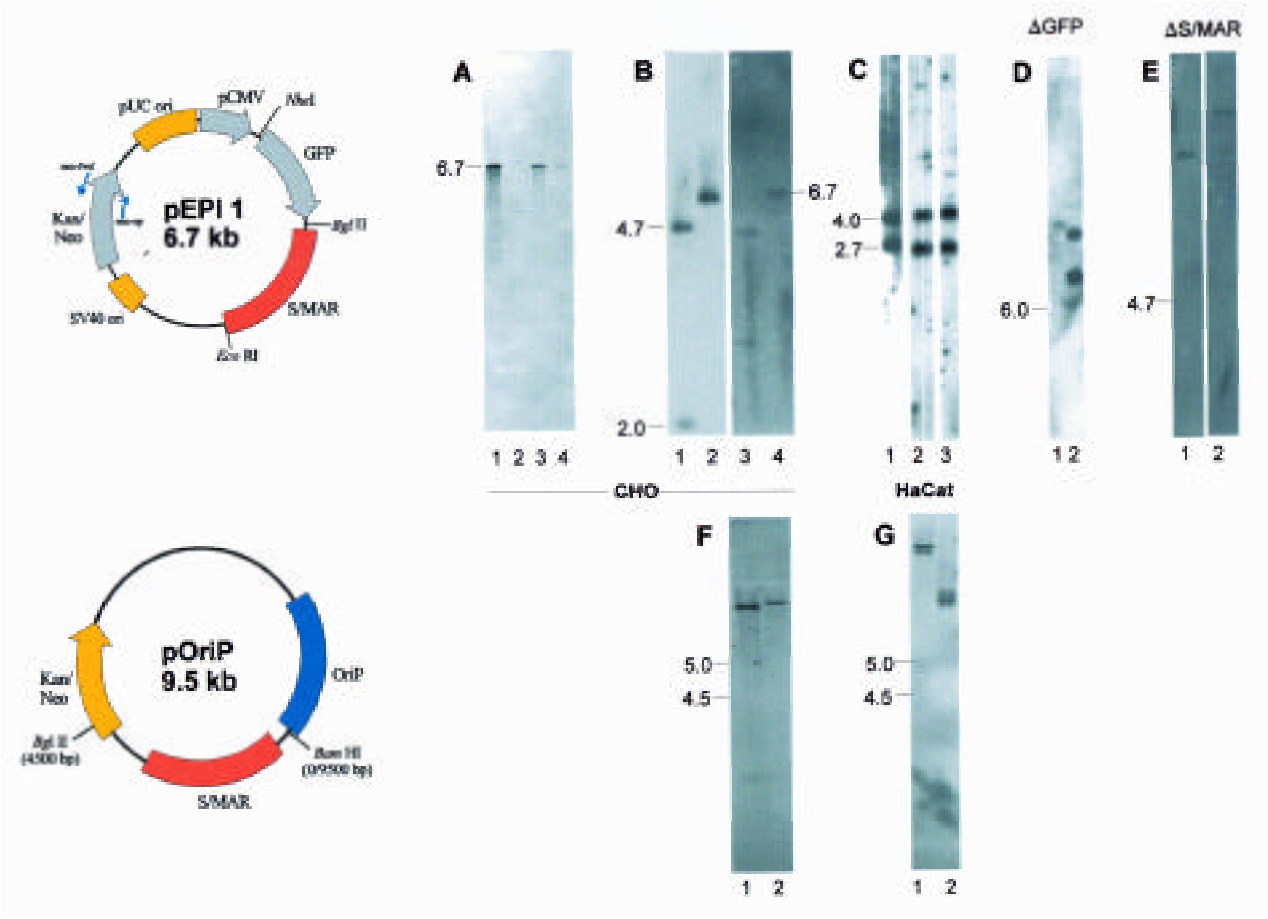

Figure 3. Verification and disproval of the episomal status. Upper row: The S/MAR-ORI^{SV40} plasmid pEPI-1 and its derivatives; Bottom row: A S/MAR-ORI^{P} plasmid. **A** - PCR tests using primers neo-up and neo-fwd; lane 1: pEPI-1 plasmid; lane 2 control CHO cells; lanes 3, 4 pEPI-1 transfected CHO cells after extended culture. **B** - Southern blots; lanes 1, 2, pEPI1 restricted by *Eco*RI plus *Bgl* II or *Eco*RI, resp.; lanes 3,4 same for pEPI-1-transfected CHO cells after extended culture. **C** - lane 1, Southern blot for pEPI-1 restricted by *Eco*RI plus *Nhe*I; lanes 2, 3 same for two independent pEPI-1-transfected HaCat clones. **D** - lanes 1,2 Southern blots for two clones transfected by pEPI1 GFP. 6.0 kb mark shows localization of the signal expected for episomal status (after linearization by *Eco*RI). **E** - lanes 1,2 Southern blots for two clones transfected by pEPI1 S/MAR. 4.7 kb mark shows localization of the signal expected for episomal status (after linearization by *Eco*RI). **F** - lanes 1,2 Southern blots for two pOriP-transfected CHO clones. 5.0 and 4.5 kb marks show localization of the signals expected for episomal status (after restriction with *Bam*HI/*Bgl* II). G - same for HaCat cells.

to the upstream gene, the transcriptional activity of the downstream gene does not influence ARS activation.

Unlike yeast, genomic DNA replication in metazoa and particularly in mammalians does not initiate at fixed ORI sequences. In some cases, active genes appear to possess an actively used ORI in their 5´ flanking region, whereas inactive genes are replicated from an origin in their 3´ flanking region (Boulikas, 1996). Thus, replication of the transcriptionally active c-myc and histone H5 genes occurs from origins in the 5´ flanking region, whereas in cell types where these genes are in a transcriptionally inactive conformation, these genes are replicated from a 3´ flanking ORI. However, replication of the active DHFR locus in CHO cells which occurs from a position in the 3´ flank underlines that we may still have to deal with isolated observations rather than with a firm rule.

C. Importance of terminators

We have shown above, that replication is tuned by transcription in a rather unpredictable fashion which in several examples is correlated with the direction of RNA polymerase movement. In case of plasmids a more general statement seems to hold that replication is only possible as long as transcription is restricted to small regions and sufficient non-transcribed regions are available for replication initiation (Haase et al 1994). Possibly, transcription through extended regions of a small circular entity affects supercoiling in a rather global fashion which

might interfere with binding of the DNA polymerase complex . While it appears conceivable that inhibition of plasmid replication is a result of interference with replication fork progression by opposing transcription complexes, other examples seem to indicate that replication inhibition can also be the consequence of an interference with replication initiation functions rather than fork progression. These properties may depend on the nature of the participating transcription factors and on their precise constellation .

In the quoted work (Haase et al, 1994) the authors have described the antagonistic effects of transcription and replication for a 24 kb episome in short-term culture. Here, unspecified origin functions were contributed by a 17 kb piece of human DNA whereas the nuclear retention functions were provided by the FR (family of repeats) from EBV (cf. **Figure 2**). The plasmid contained a complete active transcription unit composed of the Hyg gene which was driven by a HSV-*tk* promoter and flanked downstream by a *tk* transcriptional terminator. In addition, it contained a CMV promoter/enhancer the orientation of which could be changed to transcribe either a short pBR322 sequence in the (-) direction or some of the human insert in the (+)-direction. A severe inhibition of replication activity was monitored only for the (+)-case and it was reasoned that the precise termination of the (-) transcript might have prevented such an interference. This assumption was verified by adding the UMS element, a known terminator sequence, directly downstream of the promoter in the (+) construct which raised its replication activity sevenfold.

For yeast and for many eukaryotic viruses transcription units are arranged such that RNA polymerase proceeds in the same direction as the replication fork whereby it is restricted from moving through an ORI. These paradigms, however cannot necessarily be extended to the present system in which replication does not start a fixed position but rather at any of several sites throughout an initiation zone. Since there is no single region required for replication initiation, it is unlikely that transcription through a specific site will interfere with replication. It is therefore conceivable that plasmids of this type replicate as long as active transcription is confined to small regions leaving sufficient non-transcribed sequences for replication initiation. Interestingly, such a situation can be enforced by the appropriate positioning of termination sequences.

V. Lessons for the design of nonviral episomes

We have recently demonstrated that a prototype scaffold/matrix attachment element (S/MAR) is capable of recruiting the cellular replication machinery to an SV40 origin of replication which is flanked by two transcription units (**Figure 3**). This vector contains an Kan/Neo transcription gene which is transcribed in the direction of fork progression and terminated by a HSV-*tk* polyadenylation site. In addition, it harbors a reporter, i.e. a GFP gene cassette consisting of the CMV promoter, the S/MAR and an SV40 polyadenylation sequence which again shields the ORI from being approached by RNA polymerase. This vector replicates as an episome in the absence of a viral protein (T-Ag) and it is maintained for several hundred generations at a copy number around 10. Although the design of this vector has considered several of the above mentioned criteria, it is felt that its performance is not a trivial consequence of appropriately inserting transcription units, terminators and a potential nuclear retention signal in form of a S/MAR. We are in the process of unraveling the relative role these factors play for the function of the episome, its retention and the regulation of its replication, during the cell cycle. To this end we apply the entire scope of strategies offered by site-specific recombinase, to excise elements, flip their orientation or supplement them by inserting accessory elements (Bode et al 2000). Experiments so far have shown that any major alteration of this constellation is detrimental in the sense that it leads to integration (**Figure 3**).

Figure 3 (top row) show some criteria that have been used for establishing the episomal status of pEPI 1 before this property could unambiguously be demonstrated by FISH analysis (Baiker et al, 2000). As an initial test, part A shows PCR signals that are generated by closely apposed primers which have been positioned in the Kan/Neo termination region. The 6.7 kb signal shows that no random opening of pEPI 1 has occurred which would inevitably have led to its integration. Although this signal has been derived from a Hirt extract which is commonly used to enrich for episomal (non-integrated) DNA, this does not strictly exclude the presence of a precise tandem head-to-tail integration event consisting of two or more copies. Although such an integration mode is a common consequence of the classical Ca-phosphate mediated transfection of transgenes it is very unlikely if electroporation is used as in the present case (Baer et al, 2000).

A more stringent criterion are Southern blots which demonstrate the presence of the restriction fragments expected for a circular vector. This analysis is valid if presence of any additional signal can be excluded which would indicate a “bordering fragment” as a consequence of an integration event. Such a “bordering fragment“ would comprise parts of the vector and a stretch of host cell DNA terminating at the respective restriction site. Panels D and E demonstrate that no major alterations are tolerated by the system. Unexpectedly, excision of the GFP coding region causes integration which indicates an intricate interplay of transcription and replication which will have to be unraveled.

Other modifications that have been performed are exemplified by **Figure 1E** which demonstrates the consequence of deleting the S/MAR. It is noted that this deletion causes integration which suggests that the S/MAR

not only serves a retention function but also is the element enabling the formation a functional replication initiation complex. It could be argued that it is the S/MAR which contains an endogenous human origin of replication since a cohabitation of ORIs and S/MARs has frequently been documented. This is most likely not the case however, as the complementary experiment (deletion of the SV40-ORI but maintenance of the S/MAR) leads to the same outcome: integration (Baiker et al, 2000).

Besides SV40 we have also studied the function of alternative minimal viral origins of replication. Integration of a construct pORIP into the genome of either CHO or HaCat cells is demonstrated in panels F and G. We have not looked into the molecular reasons of this results which may be manifold: there is only a single (appropriately terminated) transcription unit and the S/MAR lies in the underwound part of the twin-domain (which would reinforce its strand-separation properties). Whether the underwound and overwound parts of the plasmid can compensate out each other is hard to decide due to the interposition of the matrix attachment region.

The function of the S/MAR in pEPI-1 is probably not a trivial one which could be provided by any AT rich sequence. The AT rich NTS-1 and NTS-2 sequences are associated with an endogenous origin of replication in mice (Wegner et al, 1990). When these sequences were cloned in place of the S/MAR, they did not mediate episomal replication nor did they prevent the construct_s integration. We have recently performed biomathematical analyses on multiple S/MAR elements and have demonstrated that S/MAR activities can be derived from stress-induced duplex destabilization (SIDD-) profiles in which the double strand stability of each nucleotide in the context of a given sequence is plotted (Benham et al, 1997 and in preparation). S/MARs are characterized by a regular distribution of unwinding elements which can be visualized as destabilized regions in the SIDD diagram. If these minima reach a certain level, if a certain spacing requirement is fulfilled between them and if the base-unpairing region (BUR) as a whole exceeds a threshold extension, these parameters indicate matrix attachment potential. This type of analyses is exemplified in **Figure**. **4** for pEPI-1 (Piechaczek et al, 1999), the S/MAR-free parent vector (pC1) and its NTS-1 and -2 derivatives. For pC1 the AT rich element is clearly seen as a base unpairing element. If AT sequences are cloned next to the ORI, unwinding at this site is efficiently competed for and this is most pronounced for the particular constellation which is present in the S/MAR construct pEPI-1. This finding can be interpreted as follows: the S/MAR is a major sink for superhelical strain and will be single-stranded in the absence of dsDNA binding proteins. Once the replication initiation complex has formed over the ORE, the energy stored in this sink can be retrieved by the DUE and later on by the ORE which will use it for the formation of secondary structures as a prelude to helicase action and replication initiation. Since the process is independent of viral proteins, it is possible though unproven that such a vector becomes subject to the “once per cell cycle” replication control that is typical for endogenous replication initiation zones and for EBV.

VI. Outlook and perspective

By now, it has become obvious that viruses depend on host cell functions for coming alive. Cellular functions that require a structural organization are used by the virus for transcription and replication. Replication of the small DNA tumor virus SV40 is an excellent example, because in addition to virion proteins, it encodes only a few regulatory proteins. Deppert and Schirmbeck, (1995) have summarized the evidence that all major viral processes during the life cycle from viral DNA replication to virion formation occur within the structural systems of the nucleus, in particular the chromatin and the nuclear matrix. Large T antigen itself becomes a member of the nuclear matrix where it binds to the ORI and starts the assembly of an initiation complex in concert with cellular factors. It might also mediate the known matrix association of SV40 minichromosomes which grants their replication and maintenance as episomes. Interestingly, the SV40 genome contains a S/MAR which is part of the large T coding region (Pommier et al, 1990).

DNA viruses from several families start their transcription and replication adjacent to a specific nuclear compartment which has been termed ND10, PML body or POD. Association of SV40 with ND10 appears to be a prerequisite for replication (Tang et al, 2000) where also SV40 transcription is noted. Apparently, transcription would also occur at other nuclear locations but might be concentrated at ND10 as a consequence of replication. A possible role of a S/MAR in mediating these contacts remains to be documented.

Besides their established function in replication and in the establishment of a transcriptionally active methylation-free DNA status in the genome (Dang et al, 2000) there is at least one convincing demonstration that S/MARs maintain central activities also in replicating episomes: to define the elements of the Ig- gene involved in deregulation of the c-myc gene after translocation, Hörtnagel et al (1995) have assembled different parts of the Ig-locus in an EBV-derived episomal vector. These experiments clearly showed that the S/MAR is required for the maximum c-myc activation observed in Burkitt lymphoma cells. In order to differentiate between S/MAR and enhancer functions, both elements were also tested in transient transfection experiments where the enhancers provided a 30 fold activation while in the presence of the S/MAR transcription was reduced to the background level. This work suggests that episomally replicating constructs allow to study the role of S/MARs in transcription and these systems should therefore be useful for their detailed analysis. From a practical viewpoint it is hoped that a systematic exploitation of the “hitchhiking” strategy

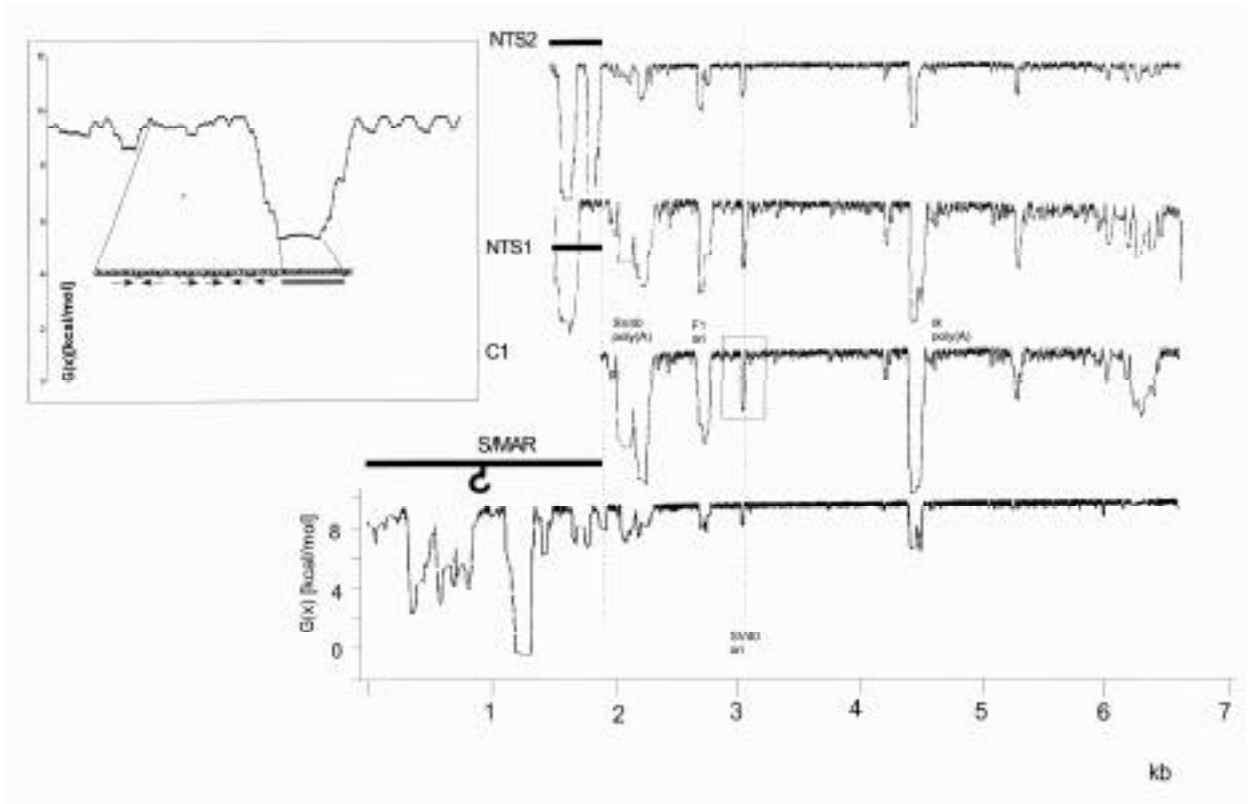

Figure 4. Stress-induced duplex destabilization (SIDD) profiles for the episome, pEPI-1, the non-episomal basic construct pC1 and its derivatives pNTS1 and pNTS2. Insert: Anatomy of a BUR. High resolution analysis for the boxed part in the C1-SIDD profile demonstrates that the A_8-tract forms the BUR (see also Fig. 1). The destabilization of this BUR (marked by the dashed line labelled "SV40 ori" is seen to be strongly modulated by the nature of elements which are present on the left of map position 1.8 (NTS2, NTS1 and S/MAR). Thereby it can be used as a gauge for estimating the relative base-unpairing potential of the surroundings which may be one parameter responsible for ORI support

invented by viruses and the positive effects of a S/MAR on both replication and transcription activities will lead to a new generation of vectors with wide applications in gene therapy and biotechnology.

Acknowledgements

We want to thank all colleagues who contributed concepts, references and outlines for the Figures. The help of Angela Knopp (GBF) in this respect is particularly acknowledged. Special thanks are due to Teni Boulikas for his encouragement and for setting the path in this evolving field. Work in the authors laboratories was supported by grants from Deutsche Forschungsgemeinschaft (Bo 419/10-1/-2 and Li 231/18-1/-2).

References

Aiyar A, Tyree C, Sugden B (**1998**) The plasmid replicon of EBV consists of multiple cis-acting elements that facilitate DNA synthesis by the cell and a viral maintenance element. **EMBO J** 21, 6394-6403.

Amati B and Gasser SM (**1990**) Drosophila scaffold-attached regions bind nuclear scaffolds and can function as ARS elements in both budding and fission yeasts. **Mol Cell Biol** 10, 5442- 5454.

Baer A, Schubeler D, Bode J (**2000**) Transcriptional properties of genomic transgenes integration sites marked by electroporation or retroviral infection. **Biochemistry** 39, 7041-7049.

Baiker A, Maercker C, Piechaczek C, Schmidt SB, Bode J, Benham C, Lipps HJ (**2000)** Mitotic stability of an episomal vector containing a human scaffold/matrix-attached region is provided by association with nuclear matrix. **Nat Cell Biol** 2, 182-184.

Benham C, Kohwi-Shigematsu T, Bode J (**1997**) Stress- induced duplex DNA destabilization in scaffold/matrix attachment regions. J Mol Biol **274**, 181-196.

Bingham PM (**1997)** Cosuppression comes to the animals. **Cell** 90, 385-387.

Bode J, Benham C, Knopp A and Mielke C (**2000a**) Transcriptional Augmentation, Modulation of Gene Expression by Scaffold/Matrix Attached Regions (S/MAR Elements) **Crit. Rev. Eukaryot. Gene Expr** 10, 73-90.

Bode J, Kohwi Y, Dickinson L, Joh T, Klehr D, Mielke C, Kohwi-Shigematsu T (**1992**) Biological significance of

unwinding capability of nuclear matrix-associating DNAs. **Science** 255, 195-197.

Bode J, Schlake T, Iber M, Schubeler D, Seibler J, Snezhkov E, Nikolaev L (**2000b**) The transgeneticist s toolbox: Novel methods for the targeted modification of eukaryotic genomes. **Biol. Chem.** 381, 801-813.

Boulikas T (**1994)** A compilation and classification of DNA binding sites for protein transcriptionfactors from vertebrates. **Crit. Rev. Eukaryot. Gene Expression** 4, 117-321.

Boulikas T (**1996)** Common structural features of replication origins in all life forms. **J Cell Biochem** 60, 297-316.

Boulikas T. and Kong CF (**1993)** A novel class of matrix attached regions (MARs) identified by random cloning and their implications in differentiation and carcinogenesis. **Int. J. Oncology** 2, 325-330.

Brewer BJ (**1988)** When polymerases collide , Replication and the transcriptional organization of the E. coli chromosome. **Cell** 53, 679-686.

Calos MP (**1996)** The potential of extrachromosomal replicating vectors for gene therapy. Trends Genet 12, 463-466.

Calos MP (**1998**) Stability without a centromere. **Proc. Natl. Acad. Sci. USA** 95, 4084-4085.

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

Collins JM, Glock MS, Chu AK (**1982**) Nuclease S1 sensitive sites in parental deoxyribonucleic acid of cold- and temperature-sensitive mammalian cells. **Biochemistry** 21, 3414- 3419.

Dang Q, Auten J, Plavec I (**2000**) Human beta interferon scaffold attachment region inhibits de novo methylation and confers long-term, copy number- dependent expression to a retroviral vector. **J Virol** 74, 2671-2678.

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

Deppert W and Schirmbeck R (**1995)** The nuclear matrix and virus function. **Int Rev Cytol** 162A, 485-537.

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

Diffley JF and Stillman B (**1988)** Protein-DNA interactions at a yeast origin. **Nature (London)** 357, 169-172.

Diffley JF and Stillman B (**1990**) The initiation of chromosomal DNA replication in eukaryotes. **Trends Genet.** 6, 427- 432.

DiMaio D, Treisman R and Maniatis T (**1982**) Bovine papilloma vector that propagates as a plasmid in both mouse and bacterial cells. **Proc Natl Acad Sci USA** 79, 4030-4034.

Ferguson BM, Brewer BJ, Reynolds AE, Fangman WL (**1991**) A yeast origin of replication is activated late in S phase. **Cell** 65, 507-515.

Frappier L and Donnell MO (**1991**) Overproduction, purifuication, and characterization of EBNA-1, the origin binding protein of Epstein-Barr virus. **J Biol Chem** 266, 7819- 7826.

Garrick D, Fiering S, Martin DI, Whitelaw E (**1998**) Repeat-induced gene silencing in mammals. **Nature Genetics** 18, 56- 59.

Guo ZS and DePamphilis ML (**1992**) Specific transcription factors stimulate simian virus 40 and polyomavirus origins of DNA replication. **Mol Cell Biol** 12, 2514-2524.

Haase SB, Heinzel SS, Calos MP (**1994**) Transcription Inhibits the Replication of Autonomously Replicating Plasmids in Human Cells. **Mol Cell Biol** 14, 2516- 2524.

Hörtnagel K, Mautner J, Strobl LJ, Wolf DA, Christoph B, Geltinger C, Polack A (**1995**) The role of immunoglobulin kappa elements in c-myc activation. **Oncogene** 10, 1393-1401.

Huang RY and Kowalski D (**1996**) Multiple DNA elements in ARS305 determine replication origin activity in a yeast chromosome. **Nucleic Acids Res**. 24 ,816-823.

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.

Kelly TJ and Brown GW (**2000**) Regulation of chromosome replication. **Annu Rev Biochem** 69, 829-880.

Kipling D and Kearsey SE (**1989**) Analysis of expression of hybrid yeast genes containing ARS elements. **Mol. Gen. Genet.** 218, 531-535.

Klehr D and Bode J (**1988**) Comparative evaluation of bovine papilloma virus (BPV) vectors for the study of gene expression in mammalian cells. **Mol Gen (Life Sci Adv) 7**, 47-52.

Kowalski D and Eddy MJ (**1989**) The DNA unwinding element, A novel, cis-acting component that facilitates opening of the E. coli replication origin. **EMBO J** 8, 4335-4344.

Laine A and Frappier L (**1995**) Identification of Epstein-Barr virus nuclear antigen 1 protein domains that direct interactions at a distance between DNA-bound proteins. **J Biol Chem** 270, 30914-30918.

Mattia E, Ceridono M, Chichiarelli S, D'Erme M (**1999**) Interactions of Epstein-Barr virus origins of replication with nuclear matrix in the latent and in the lytic phases of viral infection. **Virology** 262, 9-17.

Mielke, C Kohwi, Y Kohwi-Shigematsu, T. and Bode, J (**1990**) Hierarchical binding of DNA fragments derived from scaffold-attached regions: Correlations of properties in vitro and function in vivo. **Biochemistry** 29, 7475-7485.

Muller M, Lucchini R, Sogo JM (**2000**) Replication of yeast rDNA initiates downstream of transcriptionally active genes. **Mol Cell** 5, 767-777.

Norio P, Schildkraut CL, Yates JL (**2000**) Initiation of DNA replication within oriP is dispensable for stable replication of the latent Epstein-Barr virus chromosome after infection of established cell lines. **J Virol** 74, 8563-8574.

Norton VG, Imai BS, Yau P, Bradbury EM (**1989**) Histone acetylation reduces nucleosome core particle linking number change. **Cell** 57, 449-457.

Norton VG, Marvin KW, Yau P, Bradbury EM (**1990**) Nucleosome linking number change controlled by acetylation of histones H3 and H4. **J Biol Chem** 265, 19848-19852.

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

Piechaczek C, Fetzer C, Baiker A, Bode J, Lipps HJ (**1999**) A vector based on the SV40 origin of replication and chromosomal S/MARs replicates episomally in CHO cells. **Nucl Acid Res 27**, 426-428.

Pommier Y, Cockerill PN, Kohn KW, Garrard WT (**1990**) Identification within the simian virus 40 genome of a chromosomal loop attachment site that contains topoisomerase II cleavage sites. **J Virol** 64, 419-423.

Rawlins DR, Milman G, Hayward SD, Hayward GS (**1985)** Sequence-specific DNA binding of the Epstein-Barr virus nuclear antigen (EBNA-1) to clustered sites in the plasmid maintenance region. **Cell** 42, 859-868.

Reedman BM. and Klein G (**1973**) Cellular localization of an Epstein-Barr virus (EBV-) associated complement-fixing antigen in producer and non-producer lymphoblastoid cell lines. **Int. J. Cancer** 11, 499-520.

Studitsky VM, Clark DJ, Felsenfeld G (**1994**) A histone octamer can step around a transcribing polymerase without leaving the template. **Cell** 76, 371-382.

Tang Q, Bell P, Tegtmeyer P, Maul GG (**2000**) Replication but not transcription of simian virus 40 DNA is dependent on nuclear domain 10. **J Virol** 74, 9694-9700.

Valle M, Gruss C, Halmer L, Carazo JM, Donate LE (**2000**) Large T-antigen double hexamers imaged at the simian virus 40 origin of replication. **Mol Cell Biol** 20, 34-41.

Wang JC and Lynch AS (**1993**) Transcription and DNA supercoiling. **Curr Opin Genet Dev** 3, 764-768.

Wegner M, Schwender S, Dinkl E, Grummt F (**1990**) Interaction of a protein with a palindromic sequence from murine rDNA increases the occurence of amplification-dependent transformation in mouse cells. **J Biol Chem** 265, 13925-13932.

Wu HY, Shyy SH, Wang JC, Liu LF (**1988**) Transcription generates positively and negatively supercoiled domains in the template. **Cell** 53, 433-440.