

Targeting Regulatory Factors to Intranuclear Replication Sites

Heinrich Leonhardt,* Anje Sporbert, and M. Cristina Cardoso

Max Delbrück Center for Molecular Medicine, 13125 Berlin, Germany

* Address all correspondence to Franz Volhard Clinic, Wiltbergstr. 50, 13125 Berlin, Germany.

ABSTRACT: Plenty of evidence exists that mammalian nuclei are highly organized. Complex biochemical processes like DNA replication take place at specialized subnuclear sites and proteins directly or indirectly involved are concentrated at these sites. DNA replication is being used as a paradigm to study this functional organization of the nucleus, its underlying principles, and its potential regulatory consequences. In this review we discuss which factors were shown to be localized at nuclear replication sites, how they get there, and what role this might play in the precise, genome-wide regulation and coordination of complex biochemical processes.

KEY WORDS: functional organization of the nucleus, protein targeting, random diffusion, chromatin structure, DNA replication.

I. FROM THE TEST TUBE TO THE LIVING CELL: HOW IT ALL STARTED

A good part of what we know about DNA replication and other nuclear processes like transcription and RNA splicing we owe to biochemical analyses. Elegant and sophisticated biochemical experiments lead to the identification, isolation, and characterization of the key players. As successful as this approach was, it also hampered the appreciation/recognition of the functional organization of the nucleus. Biochemical analyses typically start with a soluble extract, discard the insoluble fraction, and still can reproduce some sort of DNA replication in the test tube. Therefore it was hard to accept that it is exactly this insoluble fraction that is carrying out DNA replication *in vivo* (reviewed in Leonhardt and Cardoso, 1995). Like other new ideas before, the concept of the functional organization of the nucleus met fierce opposition and was long disregarded as an artifact. It was only after evidence of several independent lines of research, scores of well-controlled experiments, and the advent of new live-

cell imaging techniques that this new concept gained credibility. Now it is widely accepted that DNA replication in the test tube with isolated soluble factors can explain the basic mechanisms but cannot explain the high efficiency and complex, coordinated regulation observed *in vivo*. It is now believed but not yet conclusively proven that the higher order structures identified in living cells may close the gap between rates measured in test tubes on one side and in living cells on the other side.

II. PROTEINS CAUGHT AT REPLICATION SITES AND FURTHER SUSPECTS

The field was started with three key observations: (1) the bulk of DNA synthesis activity was found to be associated with insoluble nuclear structures termed “nuclear matrix” (Berezney and Coffey, 1975); (2) DNA replication within the mammalian nucleus was found to occur in discrete subnuclear foci (Nakamura et al., 1986);

(3) the DNA polymerase delta auxiliary factor (proliferating cell nuclear antigen, PCNA) was found during S-phase not evenly distributed in the nucleus but rather tightly associated with subnuclear sites (Bravo and Macdonald-Bravo, 1987) where DNA synthesis takes place (Kill et al., 1991). These findings implied that factors involved in DNA replication were not soluble in the nucleus as in the typical biochemical assay but rather highly organized in functional subnuclear domains. Several studies using single- and double-pulse labeling with thymidine analogs suggested that different patterns of labeled chromatin exist during S-phase (reviewed in Berezney et al., 1995; Jackson and Cook, 1995; van Driel et al., 1995) with early replicating chromatin generally corresponding to transcriptionally active domains and late replicating chromatin, including heterochromatic regions (Goldman et al., 1984).

Over the last decade, attention turned also onto the other part of the equation, that is, the proteins involved in DNA replication and its control. The mammalian proteins identified to date at subnuclear replication foci, their function, and the corresponding references are summarized in the Table 1. With one exception (replication factor C, RFC), the major criteria in these studies have been colocalization with bromodeoxyuridine (BrdU) incorporation sites and with PCNA or other previously identified replication foci components. These studies have identified many expected proteins such as enzymes and auxiliary factors known to be required for DNA replication from *in vitro* biochemical studies. Interestingly, several factors "shown" *in vitro* to participate in DNA replication have not yet been localized at replication foci during S-phase. In some cases clones and antibodies are only now becoming available. In any case, negative colocalization results can be caused by the fixation method or by epitope masking, which is particularly frequent with monoclonal antibodies (the most famous example being PCNA; see Waseem and Lane, 1990). To circumvent these problems some researchers have generated epitope-tagged versions and more recently green fluorescent protein (GFP) translational fusions to study the localization of their protein of interest. But this approach can lead to erroneous conclusions because fusion proteins often show an aberrant

subcellular localization. Therefore, one should whenever possible use several independent approaches to study their subcellular distribution.

The most surprising outcome of these analyses of replication sites has been the identification of factors known to be involved in other biochemical processes. This has provided a spatio-temporal link to cell cycle regulation (cdk2 and cyclin A), DNA repair (uracyl-DNA glycosylase), nuclear structure (lamin B), DNA methylation (Dnmt1), and chromatin structure (CAF-1; for references see Table 1). The redistribution of these proteins to replication sites during S-phase connects these previously independent processes and it likely ensures the accurate genome duplication, that is, with its genetic and epigenetic features. In the last years, the list of factors "caught" at replication sites has increased but is far from being complete, and further suspects in line are minimally all the components of the DNA replication machinery identified thus far. Equally interesting will be to determine which ones are not present at these sites because they shall be unlikely to play a role in genomic replication. Finally, just having been seen at the scene of the crime does not suffice for a conviction. Colocalization at nuclear replication sites does not immediately prove any direct functional involvement; the possibility of just being an innocent bystander remains and has to be tested.

III. HOW PROTEINS GET TO REPLICATION SITES AND STAY THERE

The discovery of these higher order structures in the nucleus raised questions about how they are assembled and maintained. Because there are no subdividing membranes within the nucleus that could keep them together, other mechanisms had to be considered. The first factor to be analyzed in more detail was actually not even a real replication factor but rather the postreplicative DNA methyltransferase (Dnmt1) that was also localized at nuclear replication sites. A deletion analysis led to the identification of a targeting sequence that is necessary and sufficient for localization at replication foci (Leonhardt et al.,

TABLE 1
Mammalian Proteins Identified at Subnuclear Replication Foci

Protein	Function	Reference
DNA polymerase alpha RPA (replication protein A)	RNA primer and DNA polymerization Single-stranded DNA binding protein (SSB)	Hozak et al., 1993 Cardoso et al., 1993; Dimitrova and Gilbert, 2000
PCNA (proliferating cell nuclear antigen)	Processivity factor for DNA polymerase	Kill et al., 1991; Cardoso et al., 1993; Leonhardt et al., 2000
RFC (replication factor C)	PCNA loading onto DNA	Montecucco et al., 1998 ^a
FEN 1 (flap endonuclease)	Okazaki fragment maturation	Warbrick et al., 1998
DNA ligase I	Okazaki fragment ligation	Cardoso et al., 1997; Montecucco et al., 1995, 1998
UNG2 (uracyl-DNA glycosylase)	Uracyl removal during base excision repair	Otterlei et al., 1999
Lamin B	Nuclear lamina	Moir et al., 1994
Dnmt1 (DNA methyltransferase)	DNA methylation	Leonhardt et al., 1992; Chuang et al., 1997; Liu et al., 1998
CAF-1 (chromatin assembly factor)	Chromatin assembly	Krude, 1995; Shibahara and Stillman, 1999
Cdk 2 (cyclin-dependent kinase)	Ser/Thr kinase	Cardoso et al., 1993
Cyclin A	Cdk regulatory subunit	Cardoso et al., 1993; Sobczak- Thepot et al., 1993
PARP (poly ADP ribose polymerase)	Poly (ADP-ribosyl)ation	Simbulan-Rosenthal et al., 1996

^aOnly transiently expressed recombinant fusion protein containing the first 24 amino acids of RFC fused to the green fluorescent protein shown.

1992). This targeting sequence is located in the N-terminal, regulatory domain of the protein and could by itself direct unrelated proteins like the *Escherichia coli* derived beta-galactosidase to nuclear replication sites. Later a functionally similar sequence was identified at the N-terminus of the DNA ligase I (Cardoso et al., 1997). Both targeting sequences, from Dnmt1 and DNA ligase I, are functionally similar because they can bring unrelated proteins to replication sites but they do not share any obvious sequence similarity (see further discussion below). They are not part of the enzymatic core structure because they are dispensible for enzymatic activity *in vitro* (Kodama et al., 1991; Margot et al., 2000). A comparison of DNA ligases and methyltransferases from different species revealed a strong evolutionary conservation of the catalytic domain but not of the targeting sequence. These results suggest that the targeting sequences were added to conserved, enzymatic core structures only in the

course of the evolution of complex mammalian cells (reviewed in Cardoso and Leonhardt, 1998).

The identification and characterization of targeting sequences, however, does not yet reveal the underlying mechanism. In principle two mechanisms can be envisioned, an active transport and a diffusion-based mechanism. These mechanisms have been further investigated by studying the mobility and dynamics of splicing factors in living cells. On one hand, splicing factors seem to be specifically recruited to transcriptionally active genes and seem to move along distinct tracks corresponding to perichromatin fibrils (Misteli et al., 1997). On the other hand, splicing factors were shown to move rapidly within the nucleus using an energy-independent mechanism (Phair and Misteli, 2000). Basic principles of thermodynamics, however, require energy for the establishment of higher order structures. Therefore, the functional organization of the nucleus is likely to involve both components, an energy-indepen-

dent rapid diffusion and an energy-dependent assembly and disassembly of higher order structures. In this random diffusion step, molecules constantly collide with each other (Figure 1). Most of these collisions, however, are unproductive and do not lead to the formation of stable complexes. It is their molecular fit in terms of electrostatic and Van der Waals forces that decides whether the collision partners are bouncing off or remain associated. Energy-dependent posttranslational modifications such as phosphorylation and dephosphorylation, may modify intermolecular affinities and may thus drive the assembly and disassembly of these higher order structures.

Targeting sequences seem to be specialized structures that promote the assembly of nuclear

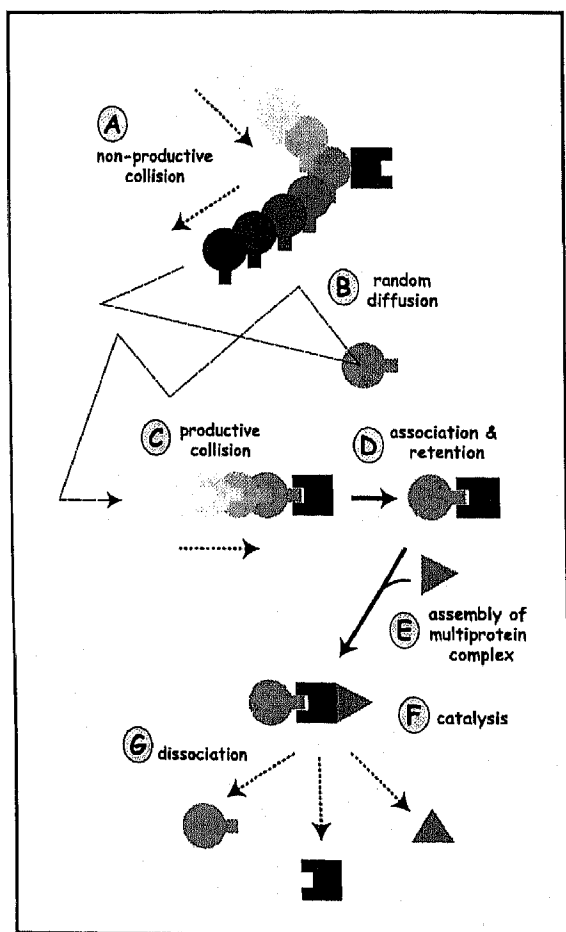


FIGURE 1. Scheme of a potential targeting mechanism. For explanations and discussion see text.

structures to coordinate multiple steps of complex biological processes like the duplication of the genome in an assembly line type of arrangement. Targeting sequences are usually defined as peptide sequences that are necessary and sufficient for localization at particular cellular structures. They are separate and independent of the

can target unrelated proteins to their particular target structures. In other words, targeting sequences make sure that enzymes are at the right place at the right time.

However, many or most replication proteins do not seem to have distinct targeting sequence but are localized at replication sites anyway. In these cases, complex assembly is likely to proceed through simple old-fashioned protein-protein interactions involving part of the enzymatic core. A key factor in the assembly of replication complexes seems to be PCNA, which by itself does not have a catalytic activity but was reported to interact with most replication proteins (reviewed in Tsurimoto, 1999). The sheer number and size of the reported interacting factors makes it hard to imagine how they can all interact with such a small protein as PCNA (discussed in Leonhardt et al., 1998). At this moment it is not clear whether PCNA acts as a loading platform or as a transporter (Figure 2). It is also not clear whether all the interactions observed *in vitro* or with dihybrid screens actually occur *in vivo*, whether they occur at replication sites or elsewhere, or whether proteins after being recruited via PCNA move to other parts of the replication machinery.

Clearly, most of these questions require new live cell imaging techniques such as fluorescence resonance energy transfer (FRET), fluorescence recovery after photobleaching (FRAP), fluorescence loss in photobleaching (FLIP), fluorescence lifetime imaging microscopy (FLIM), fluorescence correlation spectroscopy (FCS), and other still to be invented imaging techniques.

IV. POTENTIAL BENEFITS OF ORDER

Every cell division cycle requires the precise duplication of the genome, which entails actu-

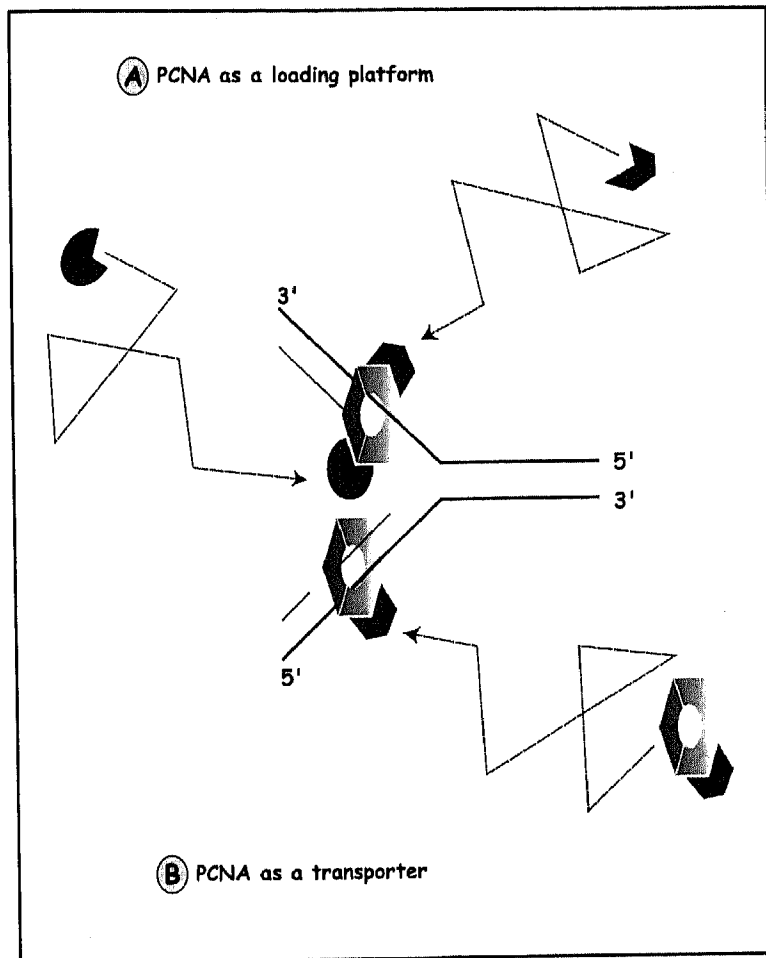


FIGURE 2. Loading platform versus transporter function. (A) PCNA loaded onto replicating DNA works as a loading platform for the several factors involved in DNA synthesis. These factors diffusing through the nucleoplasm collide and stay associated with replication factories. (B) PCNA free in the nucleoplasm associates with several factors involved in DNA synthesis and the resulting complexes are then transported and loaded onto replicating DNA.

ally much more than just DNA replication. To begin with, the decision to proliferate and to enter a new cycle has to be communicated to the replication machinery. Then the DNA (chromatin) has to be stripped of all “packaging material” (e.g., histones) to become accessible for DNA replication. During DNA replication all single-strand breaks (mostly Okazaki fragments) have to be ligated, and the DNA methylation together with other epigenetic information has to be copied onto the newly synthesized strand. Eventually, the higher order structure, the chromatin structure, has to be restored. With this step the DNA

is assembled into an either active or inactive chromatin structure and thus determines in part their future fate, whether it is accessible for regulatory factors, transcriptionally active and early replicating, or rather silenced and late replicating. Finally, the successful duplication of the genome has to be signaled to allow progression into G2 phase. In other words, numerous cellular processes have to be coordinated with DNA replication.

A glance at the list of proteins identified at replication sites (see Table 1), shows that representatives from most of these other cellular processes are actually present at these sites. The ready-

to-go signal could be delivered by the cell cycle proteins, cyclin A and cdk2, which were identified at replication sites. Representatives of the core replication machinery are DNA polymerase alpha, RPA, PCNA, and RFC. A link with the maturation and ligation of Okazaki fragments is represented by FEN 1 and DNA ligase I, assuring the complete duplication of the DNA, meaning the genetic information. Efficient repair of damaged DNA could be achieved by docking DNA repair enzymes to the replication foci as shown for UNG2. In addition, mammalian DNA carries also epigenetic information, which is encoded by chromatin structure and DNA methylation patterns. Both are crucial for the correct development-specific expression of the genome and both have to be restored after each replication round. Also in this case, coordination seems to be accomplished through association of CAF-1 and Dnmt1 with replication sites. Finally, the entire machinery could in part be coupled to an underlying nuclear protein network via lamin B.

Now, it is of course easy to say that test tube assays with soluble components failed to achieve a precise duplication of the genome and that everything can be explained by higher order structures in the nucleus. As convincing as this concept may sound, until now there is no direct evidence supporting this model. Worse, it is hard to imagine experiments that could directly address this question. Of course, it will be possible to disrupt these higher order structures by introducing specific mutations, but it will be hard to distinguish a specific role of these structures in the efficient and coordinated duplication of the genome from unspecific side effects. In the meantime, however, there will be plenty of work to elucidate the architecture of these nuclear structures. The screening for further factors localized at nuclear replication sites is likely to provide new revealing links with other cellular processes and will thus further our understanding of the regulation of DNA replication in mammalian cells. An interesting question is also how these structures are assembled and how they are disassembled again afterward. In summary, the discovery of higher order structures in the mammalian nucleus opened new lines of research with great promises and even greater challenges.

ACKNOWLEDGMENTS

We apologize to the many colleagues whose work was not cited due to space constraints. We thank the Max Delbrück Center and the Deutsche Forschungsgemeinschaft for their support over the past years.

REFERENCES

- Berezney R, Coffey DS (1975): Nuclear protein matrix: association with newly synthesized DNA. *Science* 189: 291-293.
- Berezney R, Mortillaro MJ, Ma H, Wei X, Samarabandu J (1995): The nuclear matrix: a structural milieu for genomic function. *Int Rev Cytol*:1-65.
- Bravo R, Macdonald-Bravo H (1987): Existence of two populations of cyclin/proliferating cell nuclear antigen during the cell cycle: association with DNA replication sites. *J Cell Biol* 105:1549-1554.
- Cardoso MC, Joseph C, Rahn HP, Reusch R, Nadal-Ginard B, Leonhardt H (1997): Mapping and use of a sequence that targets DNA ligase I to sites of DNA replication in vivo. *J Cell Biol* 139:579-587.
- Cardoso MC, Leonhardt H (1998): Protein targeting to subnuclear higher order structures: a new level of regulation and coordination of nuclear processes. *J Cell Biochem* 70:222-230.
- 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.
- Chuang LS, Ian HI, Koh TW, Ng HH, Xu G, Li BF (1997): Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science* 277:1996-2000.
- Dimitrova DS, Gilbert DM (2000): Stability and nuclear distribution of mammalian replication protein A heterotrimeric complex. *Exp Cell Res* 254:321-327.
- Goldman MA, Holmquist GP, Gray MC, Caston LA, Nag A (1984): Replication timing of genes and middle repetitive sequences. *Science* 224:686-692.
- Hozak P, Hassan AB, Jackson DA, Cook PR (1993): Visualization of replication factories attached to nucleoskeleton. *Cell* 73:361-373.
- Jackson DA, Cook PR (1995): The structural basis of nuclear function. *Int Rev Cytol*:125-149.
- Kill IR, Bridger JM, Campbell KH, Maldonado-Codina G, Hutchison CJ (1991): The timing of the formation and usage of replicase clusters in S-phase nuclei of human diploid fibroblasts. *J Cell Sci* 100:869-876.
- Kodama K, Barnes DE, Lindahl T (1991): In vitro mutagenesis and functional expression in *Escherichia coli* of a cDNA encoding the catalytic domain of human DNA ligase I. *Nucleic Acids Res.* 19:6093-6099.
- Krude T (1995): Chromatin assembly factor 1 (CAF-1) co-

- localizes with replication foci in HeLa cell nuclei. *Exp Cell Res* 220:304–311.
- Leonhardt H, Cardoso MC (1995): Targeting and association of proteins with functional domains in the nucleus: the insoluble solution. *Int Rev Cytol* 162B:303–335.
- Leonhardt H, Page AW, Weier HU, Bestor TH (1992): A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* 71:865–873.
- Leonhardt H, Rahn HP, Cardoso MC (1998): Intranuclear targeting of DNA replication factors. *J Cell Biochem Suppl* 31:243–249.
- Leonhardt H, Rahn HP, Weinzierl P, Sporbert A, Cremer T, Zink D, Cardoso MC (2000): Dynamics of DNA replication factories in living cells. *J Cell Biol* 149:271–280.
- Liu Y, Oakeley EJ, Sun L, Jost JP (1998): Multiple domains are involved in the targeting of the mouse DNA methyltransferase to the DNA replication foci. *Nucleic Acids Res* 26:1038–1045.
- Margot JB, Aguirre-Arteta AM, Di Giacco BV, Pradhan S, Roberts RJ, Cardoso MC, Leonhardt H (2000): Structure and function of the mouse DNA methyltransferase gene: Dnmt1 shows a tripartite structure. *J Mol Biol* 297:293–300.
- Misteli T, Caceres JF, Spector DL (1997): The dynamics of a pre-mRNA splicing factor in living cells. *Nature* 387:523–527.
- Moir RD, Montag-Lowy M, Goldman RD (1994): Dynamic properties of nuclear lamins: lamin B is associated with sites of DNA replication. *J Cell Biol* 125:1201–1212.
- Montecucco A, Rossi R, Levin DS, Gary R, Park MS, Motycka TA, Ciarrocchi G, Villa A, Biamonti G, Tomkinson AE (1998): DNA ligase I is recruited to sites of DNA replication by an interaction with proliferating cell nuclear antigen: identification of a common targeting mechanism for the assembly of replication factories. *Embo J* 17:3786–3795.
- Montecucco A, Savini E, Weighardt F, Rossi R, Ciarrocchi G, Villa A, Biamonti G (1995): The N-terminal domain of human DNA ligase I contains the nuclear localization signal and directs the enzyme to sites of DNA replication. *Embo J* 14:5379–5386.
- Nakamura H, Morita T, Sato C (1986): Structural organizations of replicon domains during DNA synthetic phase in the mammalian nucleus. *Exp Cell Res* 165:291–297.
- Otterlei M, Warbrick E, Nagelhus TA, Haug T, Slupphaug G, Akbari M, Aas PA, Steinsbekk K, Bakke O, Krokan HE (1999): Post-replicative base excision repair in replication foci. *Embo J* 18:3834–3844.
- Phair RD, Misteli T (2000): High mobility of proteins in the mammalian cell nucleus. *Nature* 404:604–609.
- Shibahara K, Stillman B (1999): Replication-dependent marking of DNA by PCNA facilitates CAF-1-coupled inheritance of chromatin. *Cell* 96:575–585.
- Simbulan-Rosenthal CM, Rosenthal DS, Hiltz H, Hickey R, Malkas L, Applegren N, Wu Y, Bers G, Smulson ME (1996): The expression of poly(ADP-ribose) polymerase during differentiation-linked DNA replication reveals that it is a component of the multiprotein DNA replication complex. *Biochemistry* 35:11622–11633.
- Sobczak-Thepot J, Harper F, Florentin Y, Zindy F, Brechot C, Puvion E (1993): Localization of cyclin A at the sites of cellular DNA replication. *Experimental Cell Research* 206:43–48.
- Tsurimoto T (1999): PCNA binding proteins. *Front Biosci* 4:D849–858.
- van Driel R, Wansink DG, van Steensel B, Grande MA, Schul W, de Jong L (1995): Nuclear domains and the nuclear matrix. *Int Rev Cytol*:151–189.
- Warbrick E, Coates PJ, Hall PA (1998): Fen1 expression: a novel marker for cell proliferation. *J Pathol* 186:319–324.
- Waseem NH, Lane DP (1990): Monoclonal antibody analysis of the proliferating cell nuclear antigen (PCNA). Structural conservation and the detection of a nucleolar form. *J Cell Sci* 96:121–129.