

# Intranuclear Targeting of DNA Replication Factors

Heinrich Leonhardt, Hans-Peter Rahn, and M. Cristina Cardoso\*

Max Delbrueck Center for Molecular Medicine, 13122 Berlin, Germany

**Abstract** Mammalian nuclei are highly organized into functional compartments. Major nuclear processes like DNA replication and RNA processing take place in distinct foci. These microscopically visible foci are formed by the assembly of, for example, DNA replication factors and associated proteins into megadalton complexes often referred to as protein machines or factories. Thus far, two proteins, DNA ligase I and DNA methyltransferase (DNA MTase), have been analyzed in greater detail. In both cases, the assembly process appears to be controlled by distinct targeting sequences that were attached to the catalytic protein core in the course of evolution and mediate the association with replication factories in mammalian cells. The dynamics of these nuclear structures throughout the cell cycle are analyzed using green fluorescent protein (GFP). Further studies are needed to elucidate the architecture, regulation, and role of these subnuclear structures. *J. Cell. Biochem. Suppl.* 30/31:243–249, 1998. © 1998 Wiley-Liss, Inc.

**Key words:** functional organization; nucleus; targeting sequence; DNA replication; nuclear matrix; cell cycle; DNA methyltransferase; DNA ligase I; PCNA; DNA replication factors; GFP

The most prominent feature of mammalian cells is the nucleus, which can even be spotted with a simple light microscope. Over the last decades, considerable work has been geared toward a better understanding of the controlled transport in and out of the nucleus and its biological consequences. According to traditional views and wisdom, that can still be found in many textbooks, nuclear proteins are subjected to random diffusion and chance encounters with their respective substrates (Fig. 1, left). These views were challenged by several lines of evidence discussed earlier in detail [Leonhardt and Cardoso, 1995]. In this paper, we focus on the organization of DNA replication in mammalian cells. The first example of a replication protein that did not always show a disperse distribution in the nucleus, and therefore did not fit into the random and free diffusion model, was proliferating cell nuclear antigen (PCNA). During S phase, PCNA redistributes to distinct nuclear foci that actively incorporate BrdU, referred to as replication foci. Subsequently, a number of other proteins were found to be con-

centrated in these replication foci, including DNA MTase [Leonhardt et al., 1992], DNA polymerase  $\alpha$  [Hozák et al., 1993], DNA ligase I [Cardoso et al., 1997], replication protein A, cyclin A, and cdk2 [Cardoso et al., 1993]. This is often referred to as functional organization of the nucleus, as proteins are localized at subnuclear sites, where their respective process, e.g., DNA replication, takes place (Fig. 1, center). In this model, the various proteins are drawn as a loose grouping, which immediately raises the question as to what keeps them together. From biochemical analyses it becomes clear that replication proteins assemble into replication complexes, which themselves have auxiliary and regulatory proteins attached and form altogether megadalton complexes [Nogushi et al., 1983; Malkas et al., 1990] termed replication machines or even factories (Fig. 1, right). These replication factories are obviously bound to the DNA that they are replicating, but they are also tethered to the nuclear matrix, which renders them insoluble and resistant to extraction [Hozák et al., 1993]. There is good evidence that this higher level of organization into protein machines and factories exists in mammalian cells. It is easy to imagine that these higher-order structures can greatly enhance the efficiency of DNA replication and methylation *in vivo*, but the actual role and

Contract grant sponsor: Council for Tobacco Research; Contract grant sponsor: Deutsche Forschungsgemeinschaft.

\*Correspondence to: M. Cristina Cardoso, Franz Volhard Clinic, Wiltbergstr. 50, 13122 Berlin, Germany.

E-mail cardoso@fvk-berlin.de

Received 9 September 1998; Accepted 10 September 1998

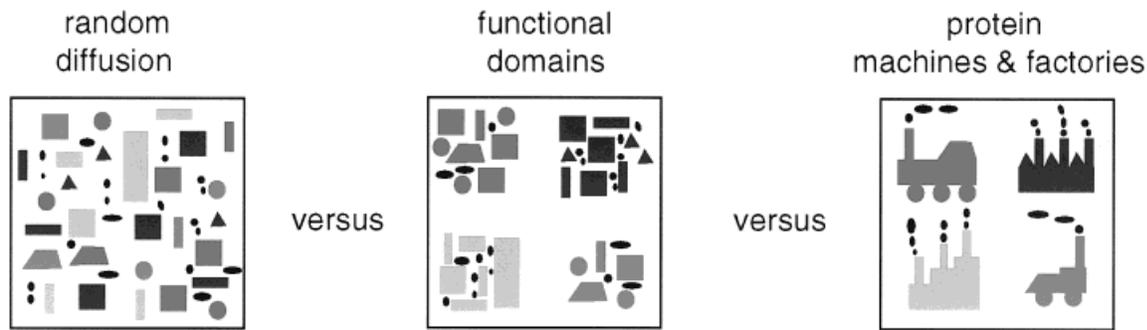


Fig. 1. Functional organization of the nucleus. Alternative, but not mutually exclusive, models of nuclear organization are outlined. The nucleus is schematically represented as an open square and nuclear proteins are represented by geometric forms.

Same colors, proteins, indicate participation in the same biochemical process like, e.g., DNA replication or RNA processing. The models are discussed in the text. **Color plate on page 327.**

regulation of these structures remain to be elucidated.

#### PROTEIN TARGETING TO NUCLEAR REPLICATION FOCI

The view of the nucleus as a highly organized subcellular compartment in which distinct biochemical processes are temporally and spatially arranged in discrete subnuclear domains (Fig. 1) raised the question of how that organization is achieved. Since there are no intranuclear membranes separating the different subnuclear compartments, the factors involved in the different processes must bind to some underlying framework. The existence of a nuclear skeleton or matrix that could provide the structural basis for the organization of the chromatin and its associated metabolic processes, although for some time controversial, is now increasingly accepted [see Cardoso and Leonhardt, 1998, and references therein]. Two different mechanisms could, in principle, explain the nonrandom distribution of some nuclear factors. Enzymes that participate in a particular process could be concentrated in discrete subnuclear domains due to a high local substrate concentration; alternatively, protein domains other than the catalytic part of these enzymes could target these factors to their respective functional sites within the nucleus.

In the case of the subnuclear DNA replication foci, this hypothesis has been investigated for two essential mammalian enzymes, DNA methyltransferase (DNA MTase) and DNA ligase I. Both enzymes were shown to be redistributed to sites of active DNA replication during S phase. DNA MTase acts postreplicatively and is responsible for the exact duplication of the DNA meth-

ylation patterns in every cell cycle [reviewed in Leonhardt and Bestor, 1993]. DNA ligase I is involved in the ligation of Okazaki fragments generated during lagging strand DNA synthesis [reviewed in Lindahl, 1992]. Both enzymes have a catalytic and a regulatory domain. If their subnuclear localization is controlled by substrate concentration one would expect the catalytic domain to be required for association with replication foci. Otherwise, one would expect a distinct region in the regulatory domain controlling their localization. To test these two possibilities, various parts of these enzymes were fused to unrelated proteins and assayed for their subnuclear localization. With these experiments, targeting sequences were identified and mapped in the regulatory domain of DNA MTase and DNA ligase I that fulfill the following criteria: (1) by itself, this domain is necessary, and (2) sufficient for subcellular localization; (3) it works position independently; (4) it is separated from the catalytic domain; and (5) it targets heterologous proteins to the respective subcellular compartment [Leonhardt et al., 1992; Cardoso et al., 1997].

Recent studies have identified additional parts of the DNA MTase as additional replication factory targeting sequences [Chuang et al., 1997; Liu et al., 1998]. Figure 2 summarizes the location of these targeting sequences in different colors. Alignment of these enzymes with bacterial or yeast homologues shows that these targeting sequences are evolutionary "add-ons" that are not conserved in lower eukaryotes and prokaryotes (Fig. 2). A short amino acid motif present in both DNA MTase and DNA ligase I and several other replication and repair enzymes was recently proposed to function as a

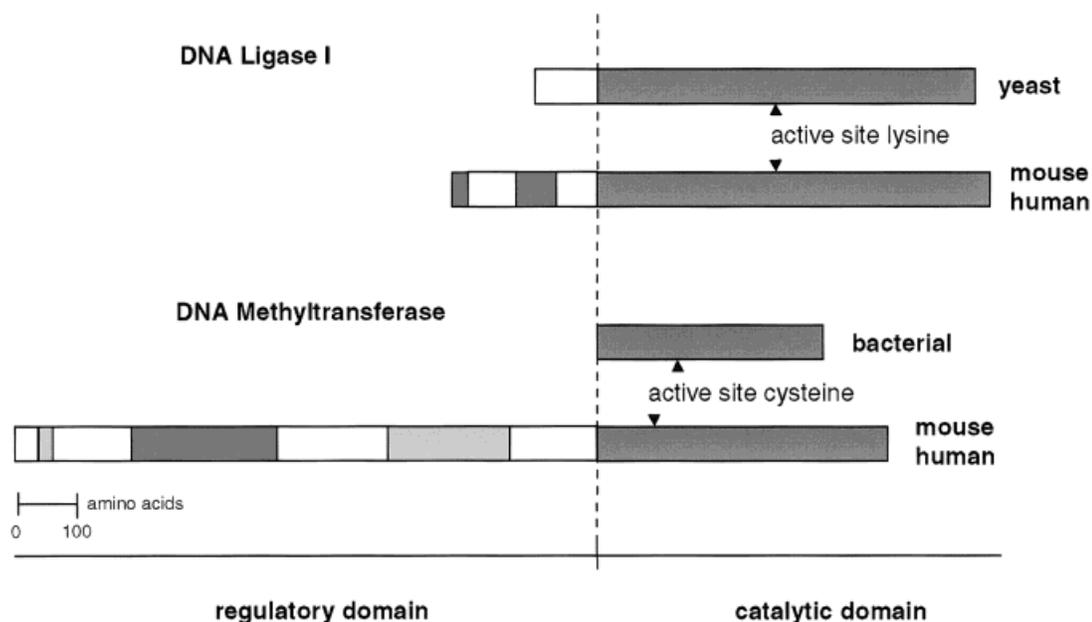
general replication foci targeting sequence [Montecucco et al., 1998]. This general targeting sequence motif has been shown to bind to the DNA polymerase processivity factor PCNA (see the following section). Two alternative but not mutually exclusive hypotheses can be postulated for the existence of multiple targeting sequences to the same subnuclear compartment and a general targeting motif. Separate domains of one enzyme could interact with different factors at the replication foci; thus, different sequences could independently target to these subnuclear structures. Alternatively, or in addition, targeting could involve sequential binding to a first site in the replication factory functioning as a “waiting room” and, at the precise time of action, to a second binding platform (perhaps, the PCNA clamp, see below) bringing the respective enzyme in closer proximity to its substrate.

In all these circumstances, targeting sequences constitute regulatory protein domains which ensure that the numerous factors are present at the right time at the right place in the nucleus. They represent the molecular code for the assembly of the replication/methylation machinery, as well as other protein machines.

This higher-order structure can explain how complex processes like replicating the genome with about 50,000 origins occurs in a timely coordinated fashion once, and only once, per cell cycle and is coupled with precise duplication of epigenetic features such as DNA methylation.

#### IS PCNA AN ASSEMBLY PLATFORM FOR THE REPLICATION MACHINERY?

The classical and first example of a protein associated with replication foci during S phase is PCNA [Bravo and Macdonald-Bravo, 1987]. PCNA has been shown to be essential *in vivo* since expression of antisense oligonucleotides to PCNA inhibited proliferation in mouse fibroblasts [Jaskulski et al., 1988] and deletion of the PCNA gene in budding and fission yeast was lethal [Bauer and Burgers, 1990; Waseem et al., 1992]. PCNA is a small ubiquitous protein that is highly conserved from yeast to humans with no known enzymatic activity but nevertheless required for DNA repair, recombination, and replication [reviewed by Kelman, 1997]. PCNA was first identified as the processivity factor of DNA polymerase  $\delta$  in 1987 [Bravo et al., 1987; Prelich et al., 1987] which indi-



**Fig. 2.** Replication targeting sequences in mammalian enzymes. The structure of two different components of the replication and methylation factory is schematically shown and compared with their respective counterparts from yeast and bacteria. Both DNA ligase I and DNA MTase are highly conserved within their respective catalytic domain across species. They differ, however, in their regulatory domain. The mouse and human

enzymes have an extended regulatory domain containing targeting sequences (*in color*) that are both necessary and sufficient for targeting to replication foci and are not present in the respective yeast or bacterial enzyme. The relative position of the catalytic center (lysine and cysteine residues) is indicated with a *black arrow tip*. **Color plate on page 327.**

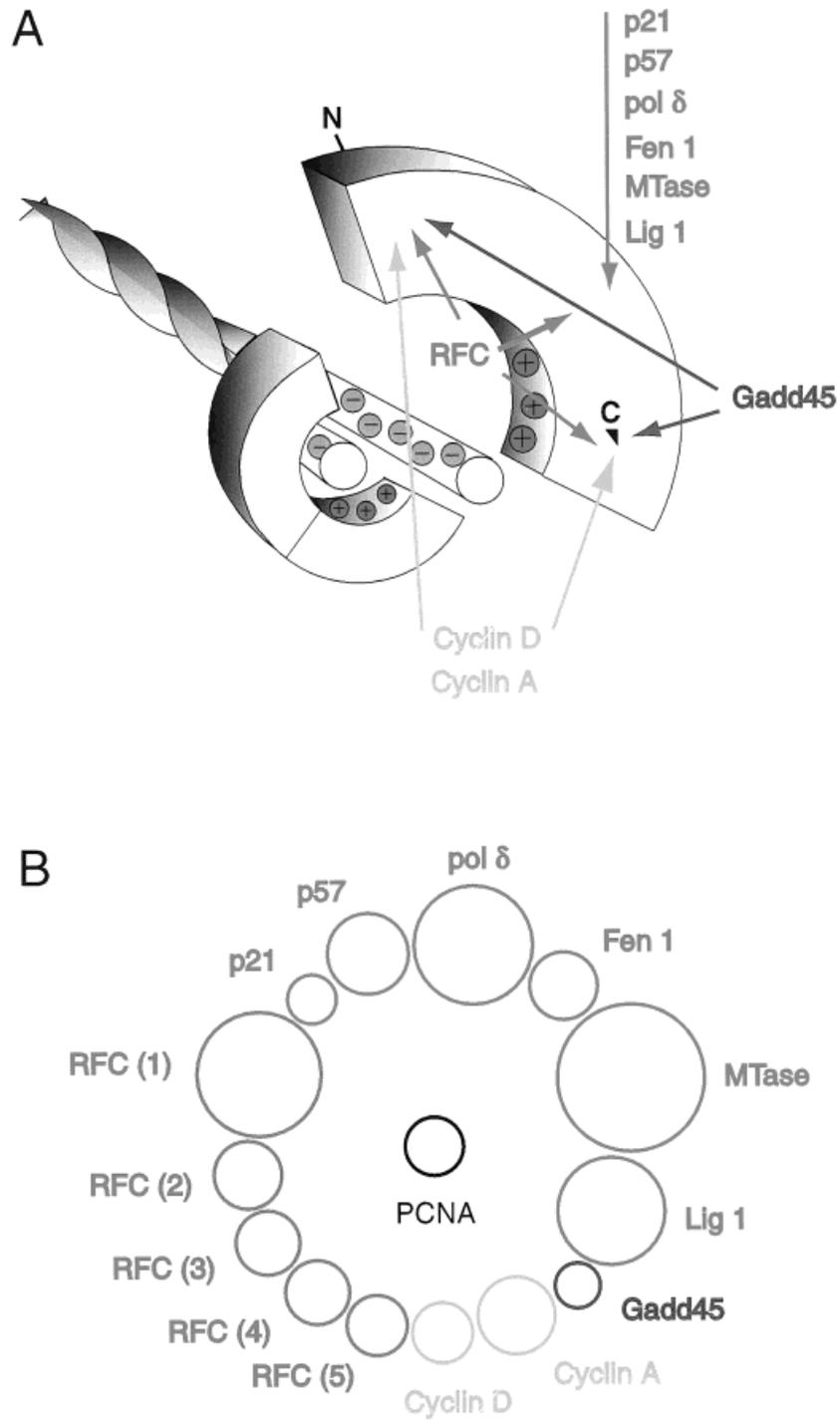
cated its functional homology to the prokaryotic processivity factors  $\beta$ -subunit of DNA polymerase III and gene 45 product of bacteriophage T4. A series of elegant biochemical studies in the prokaryotic and in the eukaryotic DNA replication systems demonstrated that PCNA forms a homotrimeric ring loaded onto duplex DNA with the assistance of the ATPase activity of replication factor C (RFC). The hole in the PCNA ring is large enough to allow the DNA to be threaded through and thus serves as a clamping platform for DNA polymerases, as well as other enzymes, ensuring high processivity. The crystal structure of yeast [Krishna et al., 1994] and human PCNA [Gulbis et al., 1996] have confirmed the topological predictions showing the characteristic sixfold symmetry of the PCNA ring with a highly negative charge at the surface and a central channel with a positive potential facing the negatively charged DNA phosphate backbone. The most asymmetric element of this structure is the long interdomain connector loop running along the outer surface and linking the N- and C-terminal domains of each monomer. A large number of proteins have been shown to interact with PCNA. The list includes DNA metabolic enzymes and cell cycle regulators [reviewed in Kelman, 1997; Kelman and Hurwitz, 1998]. Figure 3A illustrates the PCNA trimer encircling the DNA molecule and, in a schematic manner, the different proteins binding to PCNA and their respective binding sites. These mapping experiments have relied mostly on biochemical approaches as well as mutation analyses and di-hybrid screens in yeast. From the diagram in Figure 3A, it becomes obvious that most of these factors are supposed to bind to the same domains and their binding is, necessarily, mutually exclusive. An important note at this point is that due to its highly charged surface PCNA is frequently isolated in di-hybrid screens as an interacting protein. This "problem" is so common that PCNA entered the list of di-hybrid screen artifacts ([http://www.fccc.edu/research/labs/golemis/main\\_false.html](http://www.fccc.edu/research/labs/golemis/main_false.html))! The same electrostatic properties can equally affect the results of biochemical binding assays, which can select unspecific electric attractions between a few often partially purified components. In particular, DNA ligase I and DNA MTase were recently shown in biochemical assays to bind PCNA [Levin et al., 1997; Chuang et al., 1997]. This binding depends on the sequences that target these proteins to sub-

nuclear replication foci. It should be noted that the additional targeting sequence of DNA MTase (Fig. 2, orange box) that binds PCNA [Chuang et al., 1997] has a net positive charge. A similarly positively charged motif has recently been identified in DNA ligase I as binding to PCNA [Montecucco et al., 1998]. Electrostatic complementarity, though in some cases representing true interactions, can also easily lead to erroneous conclusions and should be controlled for.

Focusing again on the PCNA side, many of these targeting sequences seem to compete with p21 for binding to PCNA. In addition to DNA MTase and DNA ligase I, also the endonuclease Fen 1 and DNA polymerase  $\delta$ , both involved in DNA replication, have been shown to bind to the same region of PCNA (Fig. 3A). Other regions of PCNA are also preferred binding sites for several different factors. One could envisage a situation where different partners bind to different monomers in the same PCNA clamp simultaneously, for example one PCNA molecule binding to one polymerase, another to DNA MTase and another to DNA ligase I. Even if each of the three PCNA monomers binds different factors, that would still not accommodate all reported binding proteins. Altogether, there would be a molecular "overcrowding" at the PCNA ring. This becomes more evident when one compares the sizes of these different PCNA interacting proteins to the PCNA molecule itself, as it is shown in Figure 3B. Clearly, steric hinderance should be a major consideration in the discussion of these interactions.

At this point there are several open questions. How then, if all these interactions occur *in vivo*, do all these molecules assemble via PCNA at the replication machinery? Can PCNA be the only platform responsible for the specific localization of these proteins at the replication sites? Is there an intermediate platform from which they are recruited at the exact time of their action during the replication process to PCNA, and therefore brought into closer proximity to their target? Can the different targeting sequences mapped in the same protein be a reflexion of this two platform recruitment or are these differences caused by unspecific artifacts?

Further experiments are needed to study the architecture of the replication machinery and at the next higher organization level the replication and methylation factories using classical biochemical methods such as cross-linking studies, that were successfully employed to eluci-



**Fig. 3.** PCNA is a central component of the DNA replication machinery. **A:** Schematic representation of the trimeric PCNA ring with a positively charged central channel encircling the negatively charged DNA phosphate backbone. Proteins reported to interact with PCNA are listed and grouped according to their published binding sites, indicated by *arrows* in the

PCNA molecule. **B:** Illustration of possible steric hindrance problems in the replication machinery. The size of proteins reported to interact with PCNA is shown in comparison with PCNA itself. Circles are drawn to scale according to the relative size of the individual proteins and are arranged around the PCNA monomer. Colors as in **A**. **Color plate on page 328.**

date the structure of ribosomes, which are the first and best-studied example of a cellular machinery. These biochemical analyses should be combined with live cell studies on the temporal order of assembly of these proteins at subnuclear compartments and with fluorescence resonance energy transfer (FRET) studies of protein-protein interactions *in vivo*, as described below.

#### GFP AS A TOOL TO MONITOR PROTEIN TARGETING IN LIVING CELLS

The structure and composition of replication factories are known to change during the cell cycle. They are gradually assembled probably starting as early as in mitosis and disassembled after finishing DNA replication. For lack of alternative, this process has long been studied with populations of synchronized cells trying to piece together a dynamic cycle out of isolated "snapshots" and biochemical time points. Apart from the fact that snapshots do not make good movies and might also get the order of events wrong, the synchronization procedure also is far from being perfect and represents in itself a severe interference with cellular processes. Such problems are somewhat reminiscent of Heisenberg's Uncertainty Principle from quantum physics; that, if applied to cell biology, would mean that the closer you examine a process, the less it is what it used to be, meaning that the process of close examination changes the examined process. Almost 4 years ago, a green fluorescent protein (GFP) was described [Prasher et al., 1992; Chalfie et al., 1994] that appears to be an ingenious bypass of what we shall call the Uncertainty Principle of Cell Biology. GFP can be expressed in cells just like any other protein, and its fluorophore is formed in an autocatalytic process in the presence of oxygen [Heim et al., 1994]. Thus, GFP has been used in transcriptional fusions to monitor gene expression in living cells or entire organisms and in translational fusions to monitor protein localization.

The latter approach has already been used to visualize the localization and targeting of DNA ligase I in living cells [Cardoso et al., 1997]. Fusion of either the entire DNA ligase I gene or only its targeting sequence with GFP showed a cell cycle-dependent redistribution as observed also in snapshots of fixed and immunostained cells. DNA ligase I-GFP fusions showed a dispersed distribution during G1, G2, and M phase and were tar-

geted to replication foci during S phase. Therefore, these results directly confirm the existence of subnuclear structures in living cells.

This approach can now be used to study protein targeting to replication foci, meaning the assembly and disassembly of replication factories, during the cell cycle as well as in response to external stimuli causing, e.g., DNA damage. This strategy has also been used to study the dynamics of a pre-mRNA splicing factor in living cells [Misteli et al., 1997].

As powerful as these new GFP-based approaches may be, they also have several limitations. To begin with, GFP has to be expressed in the cell to be visualized, which is the first interference with the examined biological system. Another problem is the fact that GFP has to be fused to the protein under investigation, which not only changes the properties of the protein potentially causing artefactual localization but may also generate a fusion protein with dominant negative mutant properties. The most common artefact is the formation of inclusion body-like structures upon overexpression of insoluble GFP fusion protein. To avoid these problems one has to know beforehand what to expect, which means doing traditional immunofluorescence stainings.

Many GFP fusions are detrimental or even toxic to cells in transient transfection experiments and it may thus become the study of a dying cell. A good control is therefore to establish stable cell lines. If the GFP fusion is stably expressed over many cell division cycles, one can assume that it is sort of "biocompatible." Finally, even if all these problems are successfully solved, GFP still has to be excited to emit some detectable signal. Prolonged exposure to light causes first cell cycle arrest and ultimately cell death. Therefore, illumination has to be minimized and ultraviolet (UV) components of the light should be avoided.

Over the last years, much progress has been made through the development of more sensitive cameras, better adapted filters and media with reduced autofluorescence. Most importantly, enhanced GFP mutants with improved quantum yield are available now, which can be detected at much lower protein expression levels and/or shorter exposure times, which together help minimize the disturbance of cellular processes. Last but not least, there are now GFP mutants with distinguishable spectral properties, which can be imaged side by side in

one and the same cell. It is therefore possible to follow, for example, two components of the replication factories and thus establish the temporal order of their assembly during the cell cycle as well as possible interactions with cell cycle regulators. Moreover, also the interaction of two proteins labeled with two different GFP mutants can be monitored in living cells by FRET [Miyawaki et al., 1997]. If the two GFP mutants are in close vicinity, as is the case with two interacting proteins, the emitted light of one mutant can excite the second mutant.

In summary, GFP is clearly not the sole and ultimate answer to the Uncertainty Principle of Cell Biology but can, in combination with traditional cell biology and biochemical approaches, help elucidate biological processes such as DNA replication and will thus reduce the Uncertainty.

#### REFERENCES

- Bauer GA, Burgers PM (1990): Molecular cloning, structure and expression of the yeast proliferating cell nuclear antigen gene. *Nucleic Acids Res* 18:261–265.
- 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.
- Bravo R, Frank R, Blundell PA, Macdonald-Bravo H (1987): Cyclin/PCNA is the auxiliary protein of DNA polymerase- $\delta$ . *Nature* 326:515–517.
- 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.
- 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, 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.
- Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC (1994): Green fluorescent protein as a marker for gene expression. *Science* 263:802–805.
- 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.
- Gulbis JM, Kelman Z, Hurwitz J, O'Donnell M, Kuriyan J (1997): Structure of the C-terminal region of p21(WAF1/CIP1) complexed with human PCNA. *Cell* 87:297–306.
- Heim R, Prasher DC, Tsien RY (1994): Wavelength mutations and posttranslational autooxidation of green fluorescent protein. *Proc Natl Acad Sci USA* 91:12501–12504.
- Hozak P, Hassan AB, Jackson DA, Cook PR (1993): Visualization of replication factories attached to nucleoskeleton. *Cell* 73:361–373.
- Jaskulski D, deRiel JK, Mercer WE, Calabretta B, Baserga R (1988): Inhibition of cellular proliferation by antisense oligodeoxynucleotides to PCNA cyclin. *Science* 240:1544–1546.
- Kelman Z (1997): PCNA: structure, functions and interactions. *Oncogene* 14:629–640.
- Kelman Z, Hurwitz J (1998): Protein-PCNA interactions: A DNA-scanning mechanism? *Trends Biochem Sci* 23:236–238.
- Krishna TS, Kong XP, Gary S, Burgers PM, Kuriyan J (1994): Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell* 79:1233–1243.
- Leonhardt H, Bestor TH (1993): Structure, function and regulation of mammalian DNA methyltransferase. In Jost JP, Saluz HP (eds): "DNA Methylation: Molecular Biology and Biological Significance." Basel: Switzerland: Birkäuser Verlag, pp 109–119.
- 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.
- Levin DS, Bai W, Yao N, O'Donnell M, Tomkinson AE (1997): An interaction between DNA ligase I and proliferating cell nuclear antigen: implications for Okazaki fragment synthesis and joining. *Proc Natl Acad Sci USA* 94:12863–12868.
- Lindahl T, Barnes DE (1992): Mammalian DNA ligases. *Annu Rev Biochem* 61:251–281.
- 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.
- Malkas LH, Hickey RJ, Li C, Pedersen N, Baril EF (1990): A 21S enzyme complex from HeLa cells that functions in simian virus 40 DNA replication in vitro. *Biochemistry* 29:6362–6374.
- Misteli T, Caceres JF, Spector DL (1997): The dynamics of a pre-mRNA splicing factor in living cells. *Nature* 387:523–527.
- Miyawaki A, Llopis J, Heim R, McCaffery JM, Adams JA, Ikura M, Tsien RY (1997): Fluorescent indicators for  $Ca^{2+}$  based on green fluorescent proteins and calmodulin. *Nature* 388:882–887.
- 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.
- Noguchi H, Prem veer Reddy G, Pardee AB (1983): Rapid incorporation of label from ribonucleoside diphosphates into DNA by a cell-free high molecular weight fraction from animal cell nuclei. *Cell* 32:443–451.
- Prasher DC, Eckenrode VK, Ward WW, Prendergast FG, Cormier MJ (1992): Primary structure of the Aequorea victoria green-fluorescent protein. *Gene* 111:229–233.
- Prelich G, Tan CK, Kostura M, Mathews MB, So AG, Downey KM, Stillman B (1987): Functional identity of proliferating cell nuclear antigen and a DNA polymerase- $\delta$  auxiliary protein. *Nature* 326:517–520.
- Waseem NH, Labib K, Nurse P, Lane DP (1992): Isolation and analysis of the fission yeast gene encoding polymerase delta accessory protein PCNA. *EMBO J* 11:5111–5120.