Structure and Function in the Nucleus: Subnuclear Trafficking of DNA Replication Factors

M. Cristina Cardoso,* Anje Sporbert, and Heinrich Leonhardt
Max Delbrück Center for Molecular Medicine, 13125 Berlin, Germany

Abstract

The traditional view of the eukaryotic cell nucleus as a more or less amorphous milieu in which proteins and nucleic acids are freely floating has been challenged by an ever increasing number of reports uncovering highly organized structures where biological processes are concentrated together with their corresponding factors. The identification and utilization of protein domains that are necessary and sufficient for targeting to different subnuclear compartments have begun to elucidate the molecular principles underlying this structural organization and its dynamic behavior. The combination of biochemical, cell biology, and biophysical approaches to study nuclear structure and function should help to elucidate how these higher-order structures organize and coordinate countless enzymatic activities in time and space within the mammalian nucleus. J. Cell Biochem. Suppls. 32/33:15–23, 1999.

r 1999 Wiley-Liss, Inc.

Key words: functional organization; nucleus; targeting sequence; DNA replication; PCNA; GFP; FRET; FRAP; FLIP

Although the nucleus is a hallmark of eukaryotic cells, we still know remarkably little about its structure and function. A good example to illustrate this point is the nucleolus, easily seen with a light microscope and described by cytologists more than a century ago. Despite having been long exploited as a diagnostic tool, its function in ribosome biogenesis was only established during the mid-1960s [Granboulan and Granboulan, 1965]. Thereafter, several lines of evidence point to additional functions of the nucleolus in processing and export of many different types of RNAs (e.g., some mRNAs, tRNAs, SRP RNA, telomerase RNA); recently, the concept of a plurifunctional nucleolus has been proposed [reviewed in Pederson, 1998]. Thus, after more than a century of research we are still far from a complete understanding of the structure and function of the nucleolus, not to mention the nucleus itself.

In addition to the nucleolus, studies concentrated on the organization of chromatin, initially on mitotic chromosomes, and later, with the development of in situ hybridization techniques, on interphase chromatin as well. These latter investigations demonstrated a surprising level of organization of chromatin in the so-called “chromosome territories” [Manuelidis, 1985; Schardin et al., 1985]. These findings were totally contrary to the common view of the chromatin as “noodles in a nucleoplasmic soup.” Those earlier results raised the question as to how immobile chromosomes are in the interphase nucleus. Recent work using different methods to tag specific chromosome domains in living yeast and Drosophila showed that although chromatin can move freely within the nucleus, a specific chromatin segment can only do so within a limited subregion of the nucleus, which they termed constrained Brownian motion [Marshall et al., 1997].

During the last few years, the development of new antibodies combined with the ability to tag proteins fluorescently in vivo, as well as recent advances in live-cell microscopy, have provided a wealth of information on how the different enzymatic processes are coordinated both spatially and temporally in the nucleus.

PROTEIN SORTING BEYOND THE NUCLEAR ENVELOPE

The first indications came from the use of human autoantibodies in immunofluorescence, which showed that a group of proteins (the Sm
family) that bind to nuclear RNA were concentrated in nuclear "speckles" [Lerner et al., 1981; Perraud et al., 1979; Spector et al., 1983]. These reports provided the first evidence that the nucleus was neither a "chromatin soup" nor a "bag of free-floating proteins." At the electron microscopic level, the speckled compartment corresponds to two morphologically different structures: the interchromatin granules (IGCs) and the perichromatin fibrils (PFs). These reports were followed by the identification and subnuclear localization of many other RNA and protein components of the splicing machinery at these structures and led to the concept that spliceosomes are assembled or stored in the IGCs. Recruitment of splicing factors from the IGCs to sites of active transcription in vivo is regulated by phosphorylation [Misteli and Spector, 1996] and is dependent on the large subunit of RNA polymerase II [Misteli and Spector, 1999].

An increasing number of reports on the localization of proteins that participate in different nuclear processes have revealed a variety of subnuclear compartments and the concentration of different factors at sites where their function is required is commonly referred to as "functional organization of the nucleus" [reviewed in Leonhardt and Cardoso, 1995]. Since this subnuclear compartmentalization is not achieved by separating membranes, new principles have to be evoked to explain this unforeseen organization. The general picture of subcellular protein sorting usually depicts the newly translated protein remaining in the cytosol or, dependent on the presence of specific peptide sequences, imported into the endoplasmic reticulum, mitochondria, or through the nuclear pores into the nucleus. Figure 1 shows a more updated, but far from complete, diagram of subcellular protein sorting beyond the nuclear envelope. This scheme raises the question as to how this organization is accomplished at the molecular level. The first indications came from studies mapping the domains of human retroviral proteins required for nucleolar accumulation. Deletion analysis led to the identification of a nucleolar targeting sequence in the HIV Rev and HTLV Rex proteins. These short (about 16-amino acid-long) sequences could mediate nucleolar localization of heterologous proteins such as β-galactosidase and were shown to be essential for Rev and Rex protein function [Hatanaka, 1990; Kubota et al., 1989]. Subsequently, analysis of two Drosophila splicing regulators indicated that an arginine/serine-rich (RS) domain in these factors was necessary and sufficient for localization at the speckled compartment [Li and Bingham, 1991]. These RS sequences were not required for activity but greatly enhanced it in vivo.

In addition to RNA metabolism, one of the major nuclear processes is the duplication of the genome at each cell cycle, encompassing not only the genetic information, but also the epigenetic information in the form of DNA methylation and chromatin condensation. All these features have to be accurately duplicated at each cell cycle. Any mistakes need to be corrected before the cell divides. How are all these at first sight independent biochemical processes coordinated within the nucleus during each cell cycle? The first hints came from the cell cycle-dependent redistribution of DNA methyltransferase (Dnmt1), the enzyme responsible for maintaining the DNA methylation pattern in mammalian cells. The protein was specifically localized at subnuclear sites of DNA replication; a specific targeting sequence in the N-terminal regulatory domain was shown to be necessary and sufficient for this redistribution [Leonhardt et al., 1992]. Subsequent studies identified additional regions within the regulatory domain of this enzyme that also target to replication foci during S-phase [Chuang et al., 1997; Liu et al., 1998]. Meantime, several proteins involved in cell cycle regulation (cycin A and cdk2) [Cardoso et al., 1993; Sobczak-Thepot et al., 1993], chromatin assembly (CAF-1) [Krude, 1995], DNA repair (uracil-DNA glycosylase) [Otterlei et al., 1999], and DNA replication enzymes and auxiliary factors [reviewed in Leonhardt et al., 1998] have been shown specifically to redistribute to subnuclear replication foci during S-phase. Several of these proteins bind in vitro to a central component of the replication machinery, the proliferating cell nuclear antigen (PCNA). This binding has been proposed to mediate their subnuclear localization [Warbrick, 1998]. Unfortunately, as we discussed before [Leonhardt et al., 1998], the inflation of factors allegedly binding to PCNA via the same conserved motif poses some steric hindrance problems; that is, it is hard to imagine how so many proteins can all bind to so small a protein as PCNA. Alternatively, we proposed the existence of intermediate binding sites where the different factors would dock at
the replication factories and at the precise time of their action the individual factors would bind transiently and sequentially to PCNA coming into close contact with their substrate [Leonhardt et al., 1998]. The use of other techniques described below to study protein-protein interaction in living cells in real time should help answer at least some of these questions.

Although several problems remain to be solved, the localization of factors involved in DNA methylation, chromatin assembly, cell cycle, and postreplicative DNA repair at replication foci during S-phase provides the first clue as to how these processes are coordinated within the nucleus, to ensure that genetic and epigenetic information is accurately duplicated at each cell cycle. Furthermore, the identification of subnuclear targeting sequences that are separate from the catalytic domain of the enzymes shows that these factors are not concentrated at their respective sites due to high substrate concentration, but rather are actively organized into protein factories that function like assembly lines. The existence and dynamics of these complexes can now be investigated in living cells by making appropriate translational fusions with the green fluorescent protein (GFP). As exemplified by the images depicted in Figure 1, these fusions labeled the different subnuclear compartments indicated near the micrograph, demonstrating the feasibility of this approach.

**FOCI, DOTS, AND FACTORIES**

The renewed interest in the field of nuclear structure and function during the past few years, in addition to its obvious positive aspects, has also brought some confusion concerning the nomenclature. We will therefore attempt to clarify the terminology used. At the light microscopic level, DNA replication can be visualized as dots, more often called foci. They
can be seen from either the chromatin or the replication machinery point of view. The former corresponds to labeling of the newly synthesized strand for a short pulse with modified nucleotides. The latter corresponds to the specific association of different nuclear proteins involved in the duplication of the genome (and its epigenetic features) with discrete sites within the nucleus during S-phase. Therefore, we define "replication foci" as subnuclear sites of ongoing DNA replication. This is not only a morphological definition, but a functional one as well. What do these dots represent at the molecular level? DNA replication is initiated at regions within the genome named "origins" and proceeds bidirectionally to duplicate the flanking DNA. The stretch of DNA duplicated from a single origin is called a "replicon." The different enzymes, auxiliary factors, and other regulatory proteins involved in the duplication of the genome in one replicon unit are organized into megacomplexes called "replisomes" or "replication machines." In each focus, many replication forks are thought to be active, leading to the concept that a focus seen in the light microscope corresponds to a "replication factory" containing many replication machines or replisomes.

Our knowledge of the organization of chromatin within these sites and the assembly and disassembly of the multienzyme complexes that form the factories are still very limited. Most studies on DNA replication have been performed in test tubes with rather artificial conditions and using mostly naked DNA, instead of the more "natural" chromatin substrate. Only since the mid-1980s, with the development of techniques that permit visualization of sites of DNA replication in the cell nucleus and the generation of antibodies and fusion proteins that permit the visualization of replisome clusters in fixed and more recently also live cells, could a more biological approach become possible. On the one hand, direct visualization of chromatin is now possible in living cells by loading of fluorescently labeled nucleotides that become incorporated during DNA replication and, surprisingly enough, are not removed by the DNA repair enzymes [Manders et al., 1999; Zink et al., 1998]. These studies showed an organization of the interphase chromosomes into subcompartments of about 400- to 800-nm diameter, referred to as subchromosomal foci. These foci most likely correspond to clusters of replicons forming a stable unit of chromosome structure, which are coordinately replicated during early S-phase and remained together over many cell cycles [Jackson and Pombo, 1998].

On the other hand, the generation of translational fusions of GFP to components of the replication apparatus permits labeling of the replication machines in living cells [Cardoso et al., 1997], setting the stage for real-time studies of the dynamics of this fundamental biological process. This turns out not to be a trivial undertaking, as the expression of fusion proteins targeted to this subnuclear compartment is often deleterious for cell viability. Indeed most studies so far using GFP to monitor protein localization have been performed in transient overexpression assays with the possible caveat of studying a dying cell [for further discussion, see Leonhardt et al., 1998]. Thus, a good control for the physiological relevance of the dynamics observed is to perform the experiments using stable cell lines expressing the fusion protein, which means that the expression level is compatible with cell cycle progression. The generation of stable cell lines expressing low levels of chimeric fusion proteins of GFP and DNA replication-associated proteins will allow, for the first time, direct study of replication foci dynamics in vivo. As shown in Figure 2 replication foci patterns have been proposed to change in a characteristic manner throughout S-phase. It is now possible to test whether these pattern changes occur in vivo and in synchronous waves at specific times in S-phase. In one of the few subnuclear compartments analyzed in four dimensions, the coiled body, large-scale movements, as well as coalescence of coiled bodies, were observed during the cell cycle in plant cells [Boudonck et al., 1999]. Likewise, DNA replication factories might move to unreplicated chromatin and smaller foci might fuse, forming the larger subnuclear foci typical of late S-phase. Alternatively, replication foci might not substantially change their nuclear position but rather disassemble (white circles in Fig. 2) and reassemble (filled circles in Fig. 2) at new sites in an asynchronous manner, as opposed to synchronous waves.

The next open question is how individual factories form. Are individual replication foci gradually assembled and disassembled, suggesting that the many replication machines that compose them are not all activated synchro-
nously (see upper row of focus magnification in Fig. 2), or are the individual replication machines of one factory simultaneously assembled (see lower row of focus magnification in Fig. 2). The answers to these and other questions are now possible. The stage is set to investigate, in real time in living cells or organisms, the inter-relationships between the multiprotein complexes that carry out the different metabolic processes in the nucleus and their substrate, the chromatin.

**SHEDDING SOME LIGHT ON NUCLEAR STRUCTURE AND FUNCTION**

Passive utilization of GFP to follow protein localization in real time in living cells is only one application of this amazing new tool. During the past few years, several biophysical techniques have been adapted and optimized to probe the dynamics of molecules in living cells as well as protein-protein interactions and GFP fusion proteins have proved well suited for these investigations.

Over the years, many studies have relied on coprecipitation of proteins from soluble extracts, to determine molecular interactions. Clearly, many of these interactions do exist in living cells, but others do not. Indeed, subcellular localization has shown in several cases that these allegedly interacting factors were not even present in the same subcellular compartment. In addition, in situ extraction protocols have indicated that many macromolecular complexes are present in insoluble structures [reviewed in Leonhardt and Cardoso, 1995], which are usually discarded in biochemical analyses. The ability to express GFP variants with different spectral properties simultaneously in cells enables us to exploit a biophysical effect called fluorescence resonance energy transfer (FRET) to study molecular interactions in real time in living cells. This effect occurs if the emission spectrum of the donor fluorophore overlap with the excitation spectrum of the acceptor fluorophore and if the distance between the two fluorophores is less than 10 nm (Fig. 3A). Thus,

Fig. 2. Dynamics of DNA replication factories in the mammalian cell nucleus. The subnuclear patterns of replication foci (filled circles) were proposed to change throughout S-phase. In early S-phase, most foci are distributed all over the nuclear interior. By mid-S-phase, they are visible mostly at the nuclear and nucleolar (gray color) periphery. Later larger clusters emerge that often coincide with areas of heterochromatin. The change of patterns may happen asynchronously by disassembly and reassembly at new sites or by movement of the previously assembled factories. All sites used during S-phase are shown by circles and the assembly of factories at a specific site is shown in filled circles. The assembly (filled circles within a replication focus) and disassembly of replication machines at one such replication factory in mid to late S-phase is shown in the blowup of one focus below, and it may happen sequentially (shown above) or synchronously (shown below) at each focus.
FRET can be used to obtain quantitative temporal and spatial information about the interaction of molecules in intact living cells.

To date, two pairs of GFP mutants have been used in FRET-based investigations of biological systems (blue FP/GFP and cyan FP/yellow FP) [for a review, see Pollok and Heim, 1999]. One study used FRET to test the homodimerization of the transcription factor Pit-1 in the nucleus by cotransfection of BFP- and GFP-fusions of this protein in HeLa cells [Day, 1998]. However, this technique requires a microscope with a very accurate optical setup. The donor fluorophore must be stimulated selectively and a good separation of the emitted light from the donor and the acceptor is required to permit independent measurement of the fluorescence of each fluorophore. Furthermore, the absence of FRET effect does not necessarily mean that there is no interaction between the two fluorescently labeled proteins. Due to sterical reasons, the interference between the two fluorophores might just be hindered, preventing any energy transfer. The applications of GFP in FRET-based applications are not limited to the possible GFP variant pairs, as combinations with other fluorescently labeled molecules can be used, greatly increasing the possible variations.

In addition to protein-protein interactions, GFP fusion proteins can be used to measure the mobility of proteins in living cells. This is based on the ability to photobleach irreversibly, that is, render a population of fluorophores nonfluorescent with very high intensity light within a small area and monitor the exchange between the bleached and unbleached population of fluorophores. There are two variations of photobleaching experiments [reviewed in White and Stelzer, 1999]: (1) after bleaching a small region once, the recovery of fluorescence is monitored in the bleached area (fluorescence recovery after photobleaching [FRAP]) (Fig. 3B); and (2) a region of the cell is bleached repeatedly and, following the loss of fluorescence in the rest of the cell is monitored (fluorescence loss in photobleaching [FLIP]).

These experiments provide information about the fraction of mobile and immobile fluorescent protein and permit determination of diffusion coefficients of the fluorescently tagged proteins in living cells, as well as the exchange of proteins between different subcellular compartments. As these experiments are performed in living cells, one can directly determine the (im)mobilization of proteins throughout the cell cycle or upon specific stimuli. For example, FRAP studies recently showed that lamin B receptor-GFP fusion (LBR-GFP) exists in two pools in interphase cells: an immobile fraction in the nuclear envelope and a highly mobile fraction in the endoplasmic reticulum (ER). During mitosis, however, the immobile fraction relocates to the ER, resulting in only highly mobile LBR-GFP [Ellenberg et al., 1997]. In another study, factors involved in DNA repair, which are in a freely diffusible state in the nucleoplasm, were shown upon induction of

Fig. 3. A: Principle of fluorescence resonance energy transfer (FRET). FRET is a phenomenon in which an excited donor fluorophore (brought into the excited state S1 from the ground state S0 by absorption of light with a certain wavelength) transfers part of its energy to a second acceptor fluorophore (see Jablonski energy-level diagram). This longer-wavelength fluorophore then emits fluorescent light by returning from its excited state S1 to the ground state S0. The efficient transfer of energy between the fluorophores requires (1) an overlap between the emission spectrum of the donor and the excitation spectrum of the acceptor molecule; (2) that the two fluorophores are in close proximity (less than 10 nm); and (3) an appropriate relative orientation to each other. These resonance conditions make FRET ideal to study protein-protein interactions. The development of GFP mutants with different spectral properties and the expression of GFP fusions in cells allow the application of FRET in living cells. Thus, if the two fluorophore-tagged proteins are in very close proximity, by excitation (hv1), of the donor fluorophore (blue fluorescent protein in this scheme), one can observe emission of fluorescent light (hv2) from the acceptor fluorophore (green fluorescent protein in this scheme), combined with a decrease of emission from the donor fluorophore. B: Principle of fluorescence recovery after photobleaching (FRAP). In FRAP, a short intense pulse of light is used to render a fluorophore irreversibly non-fluorescent (dark barrels in this scheme) within a small micrometer-size region. The fluorescence intensity is monitored before photobleaching (I1), and immediately after photobleaching (I0), as well as the recovery of fluorescence in the bleached area as a function of time (see fluorescence intensity plot). Recovery of fluorescence occurs as a result of the exchange between bleached and unbleached population of the fluorophore, or the fluorescent protein and the fraction of fluorescent protein participating in this exchange is considered mobile. Immobile fluorescent proteins not moving during the duration of the experiment in the cell nucleus, are either attached to the nuclear envelope or to insoluble subnuclear structures. Quantitative determination of the final fluorescence intensity I1 relative to I0 provides information about the amount of fluorescent protein that is mobile. The time point after photobleaching where one-half of the final fluorescence intensity (I1/2) is regained can be used to calculate effective diffusion coefficients of the fluorescent proteins and thus reveal the dynamics underlying the steady-state distribution of this protein inside living cells.
Figure 3.

A

Fluorescence resonance energy transfer (FRET)

Jablonski energy-level diagram

B

Fluorescence recovery after photobleaching (FRAP)
DNA damage to be transiently immobilized probably, while repairing the damaged DNA [Houtsmuller et al., 1999]. These methods are not restricted to the use of GFP fusions. Other recent studies have used photobleaching of chromatin labeled with a membrane-permeant derivative of ethidium bromide, to investigate chromatin movement within the nucleus [Abney et al., 1997]. The lack of fluorescence recovery observed in these experiments is consistent with the confinement of chromosomes to defined areas within the interphase nucleus (chromosome territories), in agreement with other reports of chromatin dynamics in living cells mentioned above.

These new biophysical approaches to cell biology questions should help in the near future to elucidate such questions as (1) the mechanism of how factors within the nucleus are “transported” to their site of action, (2) how they are retained at these sites, and (3) what the docking components are, that is to say, how this organization into higher-order structures is achieved.

One of the major impediments to rapid progress in our knowledge of how different processes are spatially and temporally coordinated within the mammalian nucleus has been the lack of adequate in vitro systems that reproduce with fidelity the in vivo architecture. The best example of a cell-free replication system uses Xenopus egg extracts and DNA that reconstitute in vitro nuclei that look morphologically normal [Newport, 1987]. Although this system has provided much valuable information on the control of DNA replication, a major drawback has been the lack of temporal regulation of DNA replication and a total lack of specificity for the DNA being replicated. Recently, some success has been attained in the field of DNA replication in mammalian cells with the establishment of a cell-free in vitro nuclei system initiating DNA replication efficiently and under cell cycle control [Krude et al., 1997]. Thus far, this in vitro nuclei system has been used to test the requirement of different factors for initiation of DNA replication [Engel et al., 1999; Stoeber et al., 1998] but also shows preservation of subnuclear DNA replication foci and might therefore become a useful tool for more rapid dissection of these structures in mammalian cells.

Finally, to elucidate all the different functions associated with subnuclear structures, and to understand how these enzymatic activities are structurally coordinated within the nucleus, it would also be advantageous to isolate these structures and elucidate their composition and architecture. For sometime, only the nucleoli and the nuclear pores had been successfully isolated. Recently, however, the isolation of interchromatin granule clusters was described and a complete list of its components is now within reach [Mintz et al., 1999]. During the next few years, we should see other structures isolated and their composition determined.

In the next millenium, the combination of the newly emerging techniques and optimized protocols outlined are expected to yield a plethora of information about the components of subnuclear compartments, their dynamics, and interactions and will most likely uncover new functions for some compartments. But, as often occurs in nature, the whole is more than the sum of all its parts. To explain biological systems will take more than the enumeration of their parts. We will have to understand their ordered temporal and spatial grouping.

REFERENCES


