Critical Reviews™ in Eukaryotic Gene Expression, 9(3&4):345–351 (1999)

Functional Links between Nuclear Structure, Gene Expression, DNA Replication, and Methylation

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

Max Delbrück Center for Molecular Medicine, 13125 Berlin, Germany

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

ABSTRACT: Over the last decades it became clear that mammalian nuclei are highly organized. Nuclear processes like DNA replication and RNA metabolism take place in distinct subnuclear foci, which are enriched for enzymes involved in the corresponding biochemical reactions. This colocalization of functions with their respective factors is often referred to as functional organization of the nucleus. This organization is achieved by assembly of different enzymes and regulatory factors into high-molecular-weight complexes that are tethered to insoluble nuclear structures. Recently, several links between nuclear structure, gene expression, DNA replication, and methylation have been described that illustrate the interrelation of higher-order structures and nuclear functions. New insights into the functional organization of the nucleus and how it could explain the high precision and overall coordination of nuclear processes are discussed.

KEY WORDS: functional organization of the nucleus, targeting sequence, DNA methyltransferase, DNA ligase I, PCNA, chromatin structure.

I. ORGANIZATION AND EFFICIENCY

Recent decades have taken our understanding of biological processes to the molecular level. Soon all players will be known, but the sum of all their individual properties clearly does not explain complex biological phenomena like DNA replication, cell cycle, and development, not to mention the life and death of an entire organism. We will try to develop this argument, taking DNA replication as an example.

By now, most of the basic, essential DNA replication factors and their enzymatic activities are known and can be combined in vitro to replicate small viral genomes like SV40 DNA. The known factors and the sum of their properties can explain the basics, but not the efficiency, accuracy, and genome-wide coordination of DNA replication observed in mammalian cells. In concrete terms, it remains to be explained: (1) how about 3.000.000.000 base pairs of DNA are replicated once and only once, meaning no base pair more and no base pair less; (2) how the activity of about 50.000 origins are coordinated in time and space; (3) how origins either duplicated or deleted by chromosomal rearrangements can be accommodated and do not cause affected regions to be replicated either twice or not at all; (4) how the same origin can be active in some cell types early during S-phase and in some late during S-phase; (5) how origins active at a given time during S-phase are not randomly distributed throughout the nucleus but concentrated in subnuclear foci; (6) how DNA damage leads to a halt of DNA replication until the damage is repaired; (7) how three billion base pairs can be replicated, with all Okazaki fragments ligated and methyl groups added on the newly synthesized strand within a few hours and with very high precision; (8) how the complex higher-order chromatin structure can be restored after DNA replication during which the DNA is stripped of attached proteins; (9) how DNA replication and DNA methylation are coordinated; (10) how DNA replication is coordinated

1045-4403/99/$5.00

with cell cycle progression. All these aspects can be illustrated with some bits of data but not yet really explained and even less reproduced in vitro.

Undoubtedly, in vitro replication assays were the key to a basic understanding of DNA replication, but they also have systematic limitations. Their shortcomings can be best demonstrated with cellular extracts or homogenates that contain the same mix of molecules as living cells, but cannot carry out a precise duplication of the genome. The major difference between test tube assays and living cells is the presence of organization and higher-order structures. Interestingly, *Xenopus* egg extracts that are used frequently for eukaryotic in vitro replication assays are known to at least partially rebuild nuclear structures after addition of DNA (reviewed in Laskey, 1986). Also, the only replication assay until now that can at least partially reproduce the complex regulation of mammalian DNA replication uses isolated nuclei, which obviously brings along most of the required structures (Engel et al., 1999; Krude et al., 1997; Stoeber et al., 1998).

These observations raise the question, what structure and organization can do to help DNA replication and other cellular processes. As cellular components are not evenly distributed throughout the cell, they may either be colocalized with potential binding partners and substrates in the same subcellular compartment or not. Thus, their subcellular localization decides whether certain interactions and reactions can occur or not. A good example for this type of regulation is the nuclear uptake of transcription factors, which can be controlled by, for example, phosphorylation or interacting factors sequestering the nuclear localization signal (reviewed in Schmitz et al., 1991). Thus, the regulation of subcellular localization is a very efficient mechanism to control gene expression but also to control the activity of the gene product itself. An example of the latter is DNA methyltransferase (Dnmt1) that is required for the replication of epigenetic information, that is, for the maintenance of DNA methylation pattern. During early development, when the genome of preimplantation embryos undergoes global demethylation, Dnmt1 is separated from its DNA substrate and is retained in the cytoplasm (Cardoso and Leonhardt, 1999; Carlson et al., 1992).

The subcellular organization, however, goes beyond sorting of proteins into different compartments, thereby controlling their access to substrates and binding partner. The enrichment of enzymes in, for example, the nucleus or even in subnuclear foci represents an increase in the effective local enzyme concentration, as outlined in Figure 1. This organization decreases the entropy of the system and increases the efficiency and precision of diffusion-based biochemical reactions, because the probability of successful collisions of enzymes and substrates increases with their concentration. With these successful collisions also the rate of enzyme-substrate complex formation and their turnover to products increases. Because DNA replication proteins as well as their auxiliary factors are concentrated in subnuclear replication foci during S-phase, they would also benefit from this effect. However, the organization of mammalian DNA replication goes beyond this simple concentration effect and involves the formation of higher-order structures, which can further increase the efficiency and fidelity of this process.

II. HIGHER-ORDER STRUCTURES AND THE ORGANIZATION OF DNA REPLICATION

By now, several factors either directly or indirectly involved in DNA replication have been identified that are evenly distributed in the nucleus during interphase with the exception of S phase when they redistribute to replication foci (Figures 2A and 2B). These replication foci are sites of ongoing DNA replication and can be labeled with thymidine analogs such as bromodeoxyuridine. The first protein identified at these foci was PCNA (Bravo and Macdonald-Bravo, 1987), which forms a trimeric ring around the DNA strand clamping other replication factors to the DNA and thus enhances processivity. Subsequently, further replication proteins were found to be concentrated at these replication foci including DNA polymerase alpha (Hozak et al., 1993), replication protein A (Cardoso et al., 1993), and DNA ligase I (Cardoso et al., 1997; Montecucco et al., 1995). However, also several indirectly involved factors were identified at these sites, including the cell cycle pro-

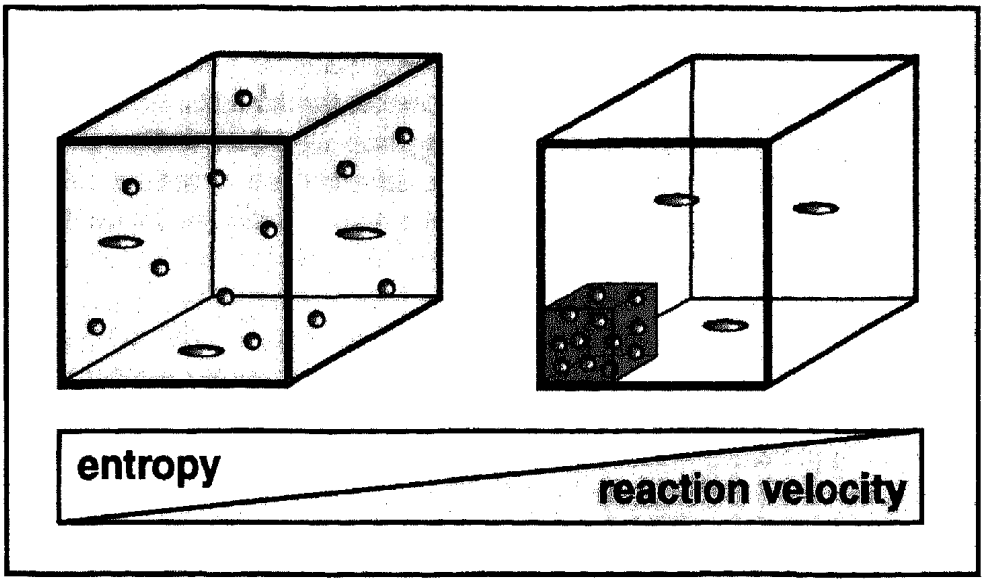

FIGURE 1. The functional organization of living cells accelerates biochemical reactions. Cells and nuclei or subnuclear compartments are schematically outlined as cubes. The enrichment of some enzymes (spheres) and substrates (shades of grey) in subcellular compartments (e.g., nucleus or replication foci) increases their effective local concentration. This reduces the entropy of the cellular system and accelerates the formation of enzyme-substrate complexes and products. The fact that other enzymes (ovals) are excluded further increases the specificity. In principle, this scheme should apply to all diffusion-based reactions in living cells.

teins cyclin A and cdk2 (Cardoso et al., 1993; Sobczak-Thepot et al., 1993), Dnmt1 (Leonhardt et al., 1992), and chromatin assembly factor 1 (CAF-1) (Krude, 1995). Using a fusion of the green fluorescent protein (GFP) with DNA ligase I, recently it was shown that this redistribution to replication foci during S-phase really occurs in living cells (Cardoso et al., 1997).

The localization of all these factors at replication foci raises the question how do they become concentrated at these sites and what prevents their random diffusion in the nucleus. This question was first addressed with Dnmt1, which after DNA replication adds methyl groups to the newly synthesized strand. Various parts of Dnmt1 were fused with an unrelated protein (β-galactosidase from *Escherichia coli*), expressed in mammalian cells and screened for localization at replication foci. With this strategy a targeting sequence could be mapped that is necessary and sufficient for localization at replication foci (Chuang et al., 1997; Leonhardt et al., 1992; Liu et al., 1998). Similar experiments with DNA ligase I led to the mapping of a replication foci targeting sequence also in this enzyme (Cardoso et al., 1997; Montecucco et al., 1998). Both sequences are functionally similar, as they can both direct unrelated proteins to replication foci, but they show no apparent similarities in their amino acid sequence. In both cases the targeting sequence was located in the N-terminal, regulatory domain of the protein and targeting was independent of the catalytic domain. The comparison of DNA ligase genes from different origins showed that the catalytic domain is highly conserved, but the targeting sequence seems to be absent in, for example, the yeast homolog, suggesting that the targeting sequence was added to a conserved enzymatic core some time during evolution probably to cope with the growing complexity in mammalian nuclei (discussed in Cardoso and Leonhardt, 1998).

The mapping of a so-called targeting sequence does not yet explain the localization of these proteins at replication foci. Biochemical fractionation experiments suggest that the various proteins involved in the replication of the genome assemble into large multienzyme complexes (Noguchi et al., 1983; Tom et al., 1996). These replication complexes are obviously bound to the DNA they are replicating and are, in addition,

tethered to some immobile nuclear structures, which then renders them insoluble and resistant to extraction (reviewed in Berezney et al., 1995; Cook, 1999). Therefore, it seems likely that proteins become localized at replication foci simply by binding to these insolubilized multienzyme complexes. In fact, an ever-increasing number of proteins is reported to bind to PCNA, which if true in vivo would raise serious problems of steric hinderance (discussed in Leonhardt et al., 1998). In any case, replication factories seem to be assembled by a network of protein–protein interactions. Several proteins, like, for example, PCNA, do not seem to have additional and dedicated domains for these interactions. In that sense, the targeting sequences of Dnmt1 and DNA ligase I are different as they seem to be recent evolutionary add-ons, which by themselves are necessary and sufficient for localization at replication foci, are independent of the catalytic core of the enzyme and function with unrelated proteins.

It is likely, but not yet shown, that Dnmt1 and DNA ligase I are randomly diffusing in the nucleus and in S phase specific binding or docking sites

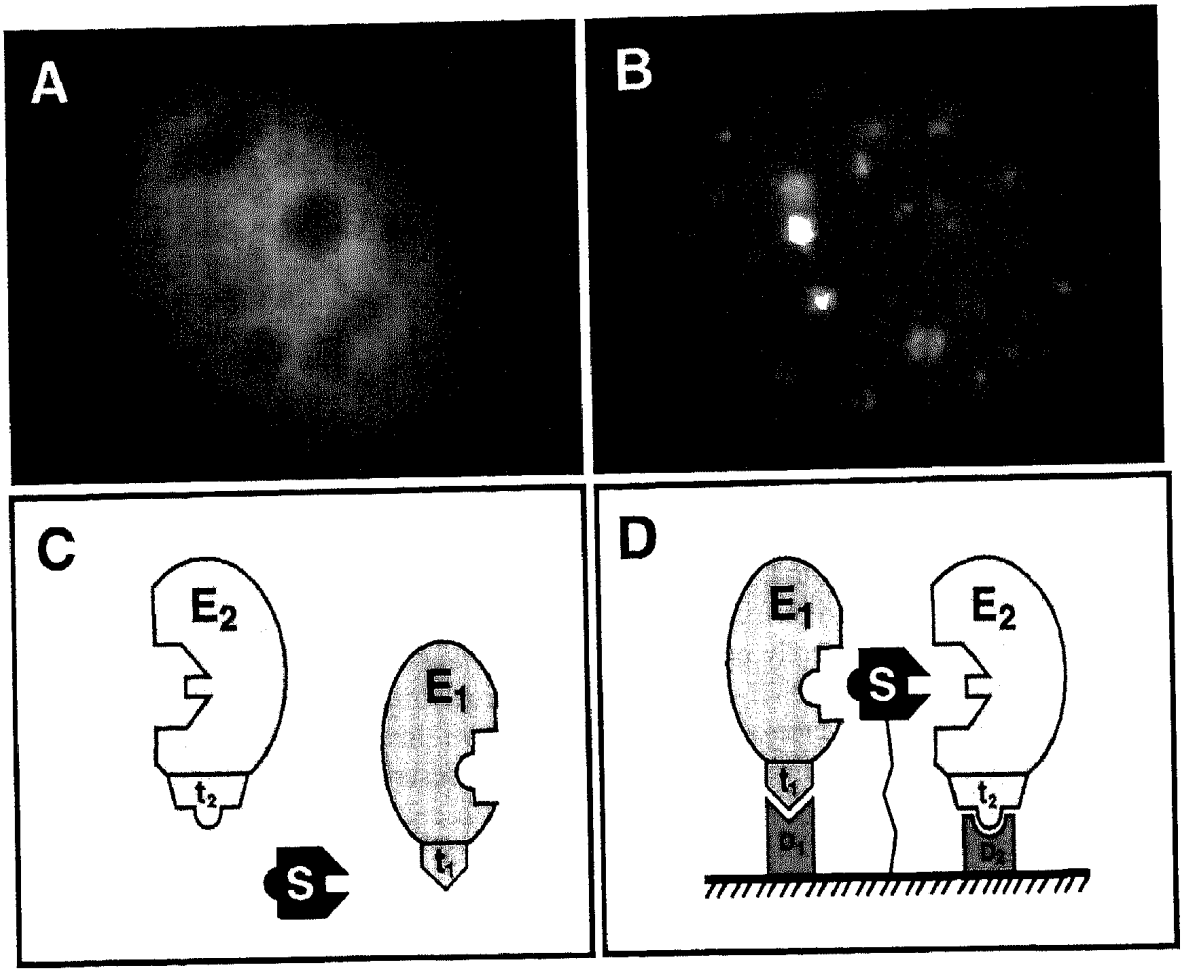

FIGURE 2. Targeting of enzymes to functional domains and molecular assembly lines. The nuclear distribution of a replication factor (PCNA) at two different stages of the cell cycle is shown. During the G1 and G2 phases PCNA is evenly distributed throughout the nucleus and is excluded from nucleoli (A). In S phase PCNA redistributes to subnuclear replication foci (B). A number of other replication factors have been found to be associated with these subnuclear foci of active DNA replication. A model for the functional organization of the nucleus is outlined below (C, D). Enzymes (E1, E2) and substrates (S) are subjected to random diffusion (C). However, at specific times they are targeted to a multienzyme complex (D). This process is mediated by specific targeting sequences (t1, t2) that bind to specific docking sites (D1, D2) at these complexes or assembly platforms. This targeting process ensures that the right enzyme is at the right time at the right place and thus enhances the efficiency and precision of complex, multistep reactions.

(see Figures 2C and 2D) are exposed that bind with high affinity the corresponding targeting sequences. In other words, the targeting sequences seem to dock their attached enzymes to multienzyme complexes like the replication factories, and thereby ensure that the right enzyme is at the right time at the right place. Moreover, enzymes may be fixed in a steric position that is optimal for direct substrate binding (see Figure 2D). This organization into "molecular assembly lines" could explain the efficiency and precision of processive, multistep reactions like DNA replication in mammalian cells.

III. FROM DNA REPLICATION AND METHYLATION TO CHROMATIN STRUCTURE AND GENE EXPRESSION

Correlations between chromatin structure, gene expression, DNA replication and methylation (see Figure 3) are known for quite some time. On one hand, DNA sequences that replicate early in S phase tend to be hypomethylated, folded in an accessible chromatin structure and actively transcribed. On the other hand, DNA sequences that replicate late in S phase tend to be hypermethylated, tightly folded in an inaccessible chromatin structure and not transcribed. These complex correlations raise the question how these processes are connected and coordinated. Part of the answer has already been outlined in the previous section. The organization of DNA replication into complex factories represents a platform for the association of other proteins and thus provides the possibility to integrate different cellular processes in the framework of these higher-order structures.

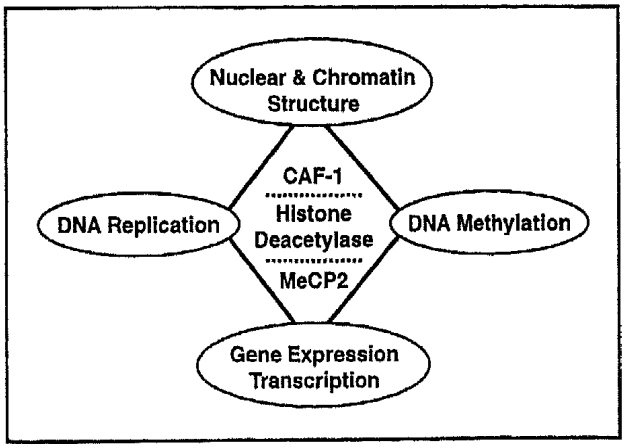

FIGURE 3. Links between structure, gene expression, DNA replication, and methylation. Recently, several functional links between nuclear structure, gene expression, DNA replication, and methylation have been identified. Among the key factors linking these nuclear processes are CAF-1, histone deacetylase, and MeCP2. See text for a detailed discussion.

One of the first links discovered was the targeting of Dnmt1 to replication foci. This association of Dnmt1 with replication factories can explain the efficient coupling of DNA replication and methylation, which seems to ensure the precise maintenance of epigenetic information after DNA replication (Leonhardt et al., 1992). Likewise, targeting of DNA ligase I couples lagging strand synthesis and the ligation of Okazaki fragments and thus seems to ensure the integrity of the genome after DNA replication (Cardoso et al., 1997). Furthermore, the identification of cyclin A and cdk2 at replication foci may represent a link between cell cycle regulation and the control of DNA replication and might transmit the start signal to the replication foci (Cardoso et al., 1993). Finally, the localization of CAF-1 at replication foci suggests that also a direct link between chromatin assembly and DNA replication exists (Krude, 1995; Shibahara and Stillman, 1999). Direct functional links also exist between DNA methylation and gene expression. In some cases, DNA methylation was shown to directly prevent transcription factor binding (Becker et al., 1987). In most cases, however, the effect of DNA methylation on gene expression is mediated by the methyl cytosine binding protein (MeCP2) that was found to be present in a complex containing histone deacetylase (Jones et al., 1998; Nan et al., 1998). Thus, MeCP2 binds methylated DNA and recruits histone deacetylase, which then removes acetyl groups from histones causing chromatin modification and gene repression. This inactivation may be reversed by either recruiting histone acetylases or by active and passive demethylation. Passive demethylation may be caused by transcription factors binding to promoter sequences directly after DNA replication, which would then prevent DNA methyl-transferases from accessing and methylating this DNA sequence and would finally lead to histone acetylation and gene reactivation (reviewed in Ng and

Bird, 1999). Again, this sequence of events connects DNA replication and methylation with chromatin structure and gene expression.

After the discovery of these functional links many interesting questions remain, but these first results clearly show that the various cellular processes are linked through countless protein-protein interactions that form a framework of higher-order structures, which may explain the high efficiency, precision, and overall coordination observed in living cells.

IV. CHALLENGES AND PERSPECTIVES FOR THE NEXT MILLENIUM

The next millenium will undoubtedly bring an unprecedented flood of data. The challenge will be to manage this flood and to make some sense out of it. Pretty soon we will know the entire sequence of the human genome and may compare it with data from our favourite model system be it mouse or yeast or else. That leaves us—give and take a few—with about 100.000 genes/proteins to play with and to assign a place in life. For each of these genes one should have some basic information: (1) transcription, alternative splicing, translation, modification, and degradation in the different cells of an organism at different stages during development; (2) three-dimensional structure; (3) abundance, subcellular localization, and modifications during the cell cycle; (4) interactions with the other 99.999 proteins during development and cell cycle; (5) assembly into higher-order structures; (6) malignant changes of all of the above during disease.

Depending on the gene, this short and certainly incomplete checklist has already been worked on with different intensity though. A literature search for a famous protein like p53 shows that more than 10.000 papers have been directly or indirectly concerned with the analysis of this one protein, but we are still far from a complete understanding. It is clear that this flood of data will bypass the capacity of even the brightest of human minds and will require sophisticated bioinformatics tools to sieve through and to make some sense out of it. After a vastly successful century of analyses subdividing complex questions into more detailed and manageable ones, dissecting organisms down to organs, tissues, cells, organelles, complexes, proteins, nucleic acids, ions, and functional groups, the challenge for the next century is to piece these bits of information again together. Eventually a better understanding of life should emerge with new and more precise possibilities to interfere with it for either the better or the worse.

The situation in molecular biology at the turn of this century is somewhat reminiscent of a story on the severity of science told by the Argentinean writer Jorge Luis Borges. In some ancient empire the science of cartography was developed to perfection to describe the empire with an ever-increasing accuracy. Eventually, the cartographers generated a perfect map that did match the empire point by point and had the size of the empire itself. Later generations less dedicated to the science of cartography left the map to the reign of sun and winter. Nowadays only a few ruins inhabited by beggars are left of that perfect map. The future will show whether our molecular maps of living organisms will share a better fate.

ACKNOWLEDGMENTS

We thank the Max Delbrück Center and the Deutsche Forschungsgemeinschaft for their support in recent years.

REFERENCES

Becker PB, Ruppert S, Schutz G (1987): Genomic footprinting reveals cell type-specific DNA binding of ubiquitous factors. Cell 51:435–443.

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 regula-

tion and coordination of nuclear processes. J Cell Biochem 70:222–230.

Cardoso MC, Leonhardt H (1999): DNA methyltransferase is actively retained in the cytoplasm during early development. J Cell Biol (in press).

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.

Carlson LL, Page AW, Bestor TH (1992): Properties and localization of DNA methyltransferase in preimplantation mouse embryos: implications for genomic imprinting. Genes Dev 6:2536–2541.

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.

Cook PR (1999): The organization of replication and transcription. Science 284:1790–1795.

Engel FB, Hauck L, Cardoso MC, Leonhardt H, Dietz R, von Harsdorf R (1999): A mammalian myocardial cell-free system to study cell cycle reentry in terminally differentiated cardiomyocytes. Circ Res 85:294–301.

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

Jones PL, Veenstra GJ, Wade PA, Vermaak D, Kass SU, Landsberger N, Strouboulis J, Wolffe AP (1998): Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nat Genet 19:187–191.

Krude T (1995): Chromatin assembly factor 1 (CAF-1) colocalizes with replication foci in HeLa cell nuclei. Exp Cell Res 220:304–311.

Krude T, Jackman M, Pines J, Laskey RA (1997): Cyclin/Cdk-dependent initiation of DNA replication in a human cell-free system. Cell 88:109–119.

Laskey RA (1986): Prospects for reassembling the cell nucleus. J Cell Sci (Suppl) 4:1–9.

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.

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.

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.

Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, Bird A (1998): Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature 393:386–389.

Ng HH, Bird A (1999): DNA methylation and chromatin modification. Curr Opin Genet Dev 9:158–163.

Noguchi H, Prem veer Reddy G, Pardee AB (1983): Rapid incorporation of label from ribonucleoside disphosphates into DNA by a cell-free high molecular weight fraction from animal cell nuclei. Cell 32:443–451.

Schmitz ML, Henkel T, Baeuerle PA (1991): Proteins controlling the nuclear uptake of NF-kappa B, Rel and dorsal. Trends Cell Biol 1:130–137.

Shibahara K, Stillman B (1999): Replication-dependent marking of DNA by PCNA facilitates CAF-1-coupled inheritance of chromatin. Cell 96:575–585.

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. Exp Cell Res 206:43–48.

Stoeber K, Mills AD, Kubota Y, Krude T, Romanowski P, Marheineke K, Laskey RA, Williams GH (1998): Cdc6 protein causes premature entry into S phase in a mammalian cell-free system. EMBO J 17:7219–7229.

Tom TD, Malkas LH, Hickey RJ (1996): Identification of multiprotein complexes containing DNA replication factors by native immunoblotting of HeLa cell protein preparations with T-antigen-dependent SV40 DNA replication activity. J Cell Biochem 63:259–267.