Foreword: Eukaryotic DNA replication: is time of the essence?

Marie-Noëlle Prioleau • Dean A. Jackson

Published online: 22 January 2010
© Springer Science+Business Media B.V. 2010

The visionary studies of Watson and Crick (1953) first proposed a structure for DNA to suggest a ‘possible copying mechanism for the genetic material’. Ever since, cell biologists have marvelled at the mysteries by which this most fundamental process of cell proliferation is controlled. The importance of mechanisms that regulate DNA synthesis is self-evident, given the biological imperative to maintain genetic integrity from one cell generation to the next. Moreover, in order to preserve the genetic information, it is essential that synthesis is performed with absolute precision, as any errors might generate chromosomal defects, which can ultimately lead to cell transformation or death.

As is often the case in biology, our desire to understand how a synthetic process might fail demands a molecular understanding of that process under normal conditions. In eukaryotes, a molecular

description of the replication machinery and basic mechanisms that regulate synthesis have been developed using biochemical analysis of yeast (Waga and Stillman 1998) and *Xenopus* (Chong et al. 1995) extracts as model systems. These studies have elaborated the key components of the synthetic machinery, the make-up of the origin recognition complex that defines where synthesis can begin and the role of the MCM2-7 complex that acts to ‘license’ DNA for replication and ensure that each potential origin is only able to operate once per cell cycle. At this level, essentially, the same basic machinery is used throughout eukaryotes.

However, while this basic synthetic machinery is now well characterised, many of the mechanisms that control the synthetic process remain enigmatic. This is particularly evident in higher eukaryotes, where increasing organismal complexity and associated increases in genome size place huge demands on the need to perform replication with absolute precision. In higher eukaryotes, epigenetic histone modifications that fix patterns of gene expression during cell specialisation must also be recapitulated during DNA synthesis. Perhaps as part of this process, the mammalian S phase is known to be organised in time so that gene-rich chromosomal R-bands are replicated early in S phase before synthesis of the gene-poor G-bands can begin (Drouin et al. 1994). Put in these simplistic terms, it is easy to see why a temporally structured S phase programme must be a fundamental feature of the eukaryotic S phase. Yet, despite this

M.-N. Prioleau (✉)
Equipe “Domaines Chromatiniens et Réplication”,
Institut Jacques Monod, UMR7592 CNRS,
Université Paris Diderot-Paris 7,
BAT. BUFFON-5ème étage-Labo 542, 15,rue Hélène
Brion, 75205 Paris cedex 13, France
e-mail: prioleau.marie-noelle@ijm.univ-paris-diderot.fr

D. A. Jackson
Faculty of Life Sciences, Manchester Interdisciplinary
Biocentre, University of Manchester,
131 Princess St, Manchester M1 7DN, UK
e-mail: dean.jackson@manchester.ac.uk

obvious logic, the mechanisms by which different regions of the genome might be destined to replicate at particular times are unknown. This special issue of *Chromosome Research* evaluates how analysis of the eukaryotic replication timing programme will allow us to understand this fundamental aspect of chromosome duplication.

The basic concept of replication timing applies similarly in simple and multi-cellular eukaryotes (Raghuraman and Brewer, this volume). However, because of their genome size, key features of the concept are best elaborated in mammalian systems. Complete replication of a mammalian genome requires that synthesis begins at some 40,000 sites. The size of replication units varies widely: ~90% of mammalian replicons are 25–500 kbp in length, and majority of these will be engaged in synthesis for 1–2 h, while a small minority of forks are seen to continue synthesis for longer periods of time. As the mammalian S phase lasts for ~10 h, most forks are only active for intervals that represent ~10% of S phase. This raises obvious questions as to whether specific regions of the genome are destined to engage synthesis at pre-determined times or may be replicated over long periods of S phase, because the sequence is replicated at different times of S phase in different cells. At the two operational extremes, the numerous initiation events that define the replication programme would be activated by deterministic or stochastic mechanisms.

This imperative to understand how the initiation of DNA synthesis is controlled has, over the past 30 years, placed huge emphasis on our analysis of the ‘replication origins’ where synthesis begins. In this regard, an early breakthrough came from the realisation that specific sequences in budding yeast (Hsiao and Carbon 1979) were able to support the autonomous replication of ectopic DNA circles (plasmids). Molecular dissection of these autonomous replicating sequences (ARS) revealed a consensus structure of DNA motifs that was later shown to recruit the origin recognition complex. Inevitably, what followed was a predictable eagerness to find the counterpart of ARS elements in higher eukaryotes. At this point, what everyone in the field thought would be a routine technical challenge gradually revealed that the initiation sites for replication in higher eukaryotes might turn out to have unpredictably enigmatic characteristics (Gilbert 2004). In this volume, Joyce Hamlin and her colleagues give an

excellent personalised account of the struggle to understand the behaviour of mammalian replication origins.

Over 20 years or more, it slowly became apparent that the ARS paradigm for replication origins was not relevant in higher eukaryotes. In fact, early attempts to recapitulate the plasmid-based ARS assay in mammalian systems showed that origin activity correlated broadly with sequence length, so that DNA fragments of ~10 kbp or more had a high probability of supporting DNA synthesis (Heinzel et al. 1991). This and many other lines of evidence supported the view that replication initiation was not defined by genetically controlled sequence elements but based on more subtle features of genome structure, chromatin context and nuclear organisation. Eventually, local transcriptional activity emerged as a major factor that might contribute to origin choice during the initiation of DNA synthesis.

The suggestion that initiation of replication is somehow based on ‘context’ is oddly unsatisfactory for molecular biologists who thrive on mechanistic understanding. Happily, the present post-genomic enthusiasm for genome-wide analysis is beginning to reveal more detail of the general principles of replication control. In this volume, two thought-provoking articles review how genome-wide strategies can be used to define the initiation sites of DNA synthesis. In the first, Schepers and Papior (this volume) analyse how components of the pre-replication complexes—the potential sites of initiation—can be used to explore the distribution of initiation targets on chromatin. In the second, Cadoret and Prioleau (this volume) assess how genome-wide tools can be used to define sites where DNA initiation begins. In this case, analysis relies of the purification of short nascent strands that are enriched in replication initiation zones. However, because of their small size, these fragments are difficult to isolate in large quantities, and as a result, experiments to date have concentrated on functionally informative regions of the human genome—represented in ~1% of the human sequence identified by the ENCODE project. Genome-wide analysis using high resolution microarrays has also been used to map the mammalian replication timing programme (Farkash-Amar and Simon, this volume) and explore how the timing is modulated during cell differentiation (Pope et al., this volume).

Genome-wide analysis of DNA synthesis is beginning to have a significant impact on our understanding

of the synthetic process and basic features of the regulatory mechanisms. However, it is also clear that important mechanistic detail might be lost by using techniques that combine information from huge populations of cells. Notably, biochemical and molecular strategies that are analysed on microarrays will commonly use 10^6 – 10^8 cells, so that any biological variability between cells is not seen; readout from such experiments is inevitably a composite that represents a population average for the sample used. Based on this loss of biological complexity, it is reasonable to argue that analysis of replication within individual cells will also be needed for a complete mechanistic understanding of the replication process (see Shaw et al. in this volume for discussion).

Two general experimental approaches are worthy of consideration as we zoom in on single cells. Historically, the distribution of active replication forks on DNA fibres has been used to explore replication speeds and replicon structure. Traditionally, active replicons were labelled with ^3H -thymidine and visualised by autoradiography. Studying synthesis in mammalian systems, this approach revealed that adjacent replicons were often activated together in small groups (Hand 1978; for an excellent historical perspective, see Gilbert 2007). This general theme of synthetic co-activation of groups of replicons was emphasised when Nakamura et al. (1986) showed that the active sites in nuclei contained synthetic hotspots where many replicon are synthesised together. Eventually, the combined analysis of DNA foci and DNA fibres prepared from the same labelled cell populations suggested that mammalian DNA is packaged as DNA foci that are both structural units of chromosome architecture and functional targets for DNA synthesis (Jackson and Pombo 1998). Importantly, these foci were shown to contain small groups of replicons, which are activated together and appear to be replicated by the synthetic machinery within dedicated replication factories (Hozak et al. 1993). A detailed analysis of the structure and spatial architecture of replication sites is presented here by Toyoaki Natsume and Tomoyuki Tanaka.

The analysis of units of DNA synthesis on isolated DNA fibres provides powerful insight into mechanisms of origin selection and activation (reviewed by Tuduri et al., this volume). As an example, fibre analysis shows that during the normal cycle, potential origins are selected from multiple potential initiation sites, within

licensing groups that typically cover ~10 kbp of DNA. In any cell during a single S phase, origins are selected to engage synthesis in only about one third of these initiation zones. Potential origins that are not used in a particular cell cycle are silenced once replication is activated locally and are eventually replicated passively by extending forks from the neighbouring active replicons. This stochastic feature of origin selection means that the time at which a particular region of the genome is replicated will be dictated by its proximity to the nearest active replication origin.

If stochastic selection of potential replication origins is a general feature of the eukaryotic S phase, it is important to assess how random mechanisms of origin activation can be reconciled with a temporally structured S phase, which demands that part of the regulation is defined by innate properties of the chromatin template and replication machinery. Various features of chromatin structure might accommodate a combination of stochastic and deterministic features. For example, given the general trend to replicate euchromatin and heterochromatin during discrete periods of S phase, it is reasonable to argue that local chromatin environments play a deterministic role in the process of origin selection and activation (discussed in Cayrou et al., this volume). Within euchromatin, early origins are frequently found in the vicinity of the most active genes (Cadoret and Prioleau, this volume), implying that local patterns of gene expression might influence origin selection at the onset of S phase. This could reflect the accessibility of potential origins within the most exposed and plastic chromatin (this chromatin compartment is clearly dynamic in its ability to move and interact locally with different nuclear compartments) or the fact that transcription will locate these domains to the active sites of RNA synthesis within the inter-chromatin nuclear domain, where replication factories also form (Hassan et al. 1994).

This use of chromatin context to define how different classes of chromatin might interact with the synthetic nuclear compartment is now thought to be the dominant feature of nuclear organisation that defines S phase progression. To emphasise this point, we encourage anyone who is interested in understanding how nuclear structure and function are related to read about this fascinating topic here in this special issue of *Chromosome Research*.

In addition to the roles played by chromatin context in regulating S phase, origin selection is also

synchronised with cell cycle progression by coupling origins that are activated preferentially during early or late S phase to the expression of specific cyclin-CDK protein complexes. In budding yeast, different cyclin-CDK complexes—which contain the cyclins Clb5 and Clb6—appear to play a dominant role in defining the origins that are active during early and late S phase (Raghuraman and Brewer, this volume). In this organism, Clb6 plays the major role early in S phase, but this protein degrades during S phase, and origin activation is then supported by Clb5. In mutants with compromised Clb5 activity, the early origins fire as normal, but the activation of late origins is significantly delayed. In human cells, increased expression of specific cyclin-CDK complexes towards mid/late S phase ensures that target origins within heterochromatin are activated at the appropriate time of S phase (discussed in Nakanishi et al., this volume).

The analysis of origin use and replication timing almost inevitably treats individual replication units as independent entities. However, one of the most fascinating features of control in higher eukaryotes is that synthesis appears to be defined holistically, in order to ensure that replication completes on programme within a synthetic period that occupies 10–12 h of the cell cycle. Though the regulatory mechanisms are unknown, it is notable that many conditions that reduce the rate of fork progression—for example, by reducing the concentration of precursor pools—have surprisingly little impact on S phase timing. This compensation for defects in elongation is possible because any decline in the synthetic capacity of engaged forks is recognised by a form of rate sensor, which controls the local density of active forks in order to ensure that optimal levels of synthesis are maintained (Tuduri et al., this volume). This implies that mammalian cells continually assess the absolute level of DNA synthesis and that overall synthesis is somehow regulated by a replication rheostat, which defines the amount of synthesis that is required to complete S phase on time. Intriguingly, this level of regulation appears to define the number of active factories and not the number of active forks that these factories contain, suggesting that factory assembly might be an essential control point.

Studies that have been published to date support a model of S phase progression that is based on a combination of deterministic and stochastic events (discussed in articles throughout this volume). During

S phase, origin selection and activation are defined in large part by the local chromatin environment, with cyclin expression providing an additional selection mechanism (Nakanishi et al., this volume). Based on features of the local chromatin environment, preferred sites are activated at the onset of S phase. As S phase proceeds, synthesis at the preferred sites appears to enhance the probability with which synthesis is activated, and hence spreads, into the neighbouring, genetically coupled, replicon cluster along individual chromosomes (discussed in Shaw et al., this volume). Though this mechanism of spread is not obligatory, the sequential activation of genetically neighbouring clusters will limit the number of isolated forks, which, if allowed to accumulate, might compromise genome integrity. However, while individual DNA foci might be selected for synthesis because of their relationship to foci that were replicated earlier in S phase, the mechanisms by which potential origins are activated within individual foci is stochastic. This is clear from the fact that origin activation is inefficient so that within individual cells, the majority of potential initiation sites are not used to activate synthesis but are replicated passively by forks from neighbouring active origins (Hamlin et al. and Tuduri et al., this volume).

As features that contribute to S phase progression in higher eukaryotes begin to emerge, it becomes exciting to ask how *in silico* modelling can be used to explore how different combinations of these features might influence S phase. Computer-generated simulations of the eukaryotic S phase are presently rather primitive and most explore how the activation of origins with different probabilities can define the efficiency of synthesis. Two papers in this volume (Rhind et al. and Hyrien and Goldar) describe the present state-of-the-art for modelling eukaryotic S phase; the first focuses on unicellular systems and the second addresses how models might be adapted to explore the more complex replication programmes seen in high eukaryotes. Finally, Shaw et al. (this volume) explore how higher-order genome architecture together with variable expression of specific cyclin-CDK complexes might together drive a replication programme that mimics DNA synthesis in mammalian cells.

Recent advances in genome-wide analysis using both high-density microarrays and high-volume DNA sequencing technologies are beginning to revolutionise our ability to explore how eukaryotic DNA synthesis is

controlled to ensure the preservation of genetic integrity. Though many questions remain to be resolved, there is no doubt that the genome-wide data sets combined with data derived from analysis of single cells will continue to reveal fundamental insight into mechanisms that regulate and define the eukaryotic S phase.

It goes without saying that accurate duplication of the genetic material is essential for cell proliferation and hence all life. How such a complex process is achieved with the required accuracy is endlessly fascinating. We are confident that everyone who reads the articles in this volume will immediately recognise the enigmatic nature of the synthetic process and the experimental ingenuity that has been used to explore this fundamental process. We would like to thank all authors for their excellent and uniformly erudite and thought-provoking contributions to this special issue. Our special thanks go to Herbert Macgregor, Editor-in-chief of *Chromosome Research*, for his tireless enthusiasm, support and editorial precision, and to the editorial board for initiating this project on behalf of the Journal.

References

- Chong JPI, Mahbubani HM, Khoo CY, Blow JJ (1995) Purification of an MCM-containing complex a component of the DNA replication licensing system. *Nature* 375:418–421
- Drouin R, Holmquist GP, Richer CL (1994) High-resolution replication bands compared with morphological G-bands and R-bands. *Advances Hum Genet* 22:47–115
- Gilbert GM (2004) In search of the holy replicator. *Nat Rev Mol Cell Biol* 5:848–854
- Gilbert (2007) Replication origin plasticity, Taylor-made: inhibition vs recruitment of origins under conditions of replication stress. *Chromosoma* 116:341–347
- Hand R (1978) Eukaryotic DNA—organization of the genome for replication. *Cell* 15:317–325
- Hassan AB, Errington RJ, White NS, Jackson DA, Cook PR (1994) Replication and transcription sites are colocalized in human cells. *J Cell Sci* 107:425–434
- Heinzel SS, Krysan PJ, Tran CT, Calos MP (1991) Autonomous DNA replication in human cells is affected by the size and source of DNA. *Mol Cell Biol* 11:2263–2272
- Hozak P, Hassan AB, Jackson DA, Cook PR (1993) Visualization of replication factories attached to a nucleoskeleton. *Cell* 73:361–373
- Hsiao C-L, Carbon J (1979) High frequency transformation of yeast by plasmids containing the cloned ARG4 gene. *Proc Nat Acad Sci USA* 76:3829–3833
- Jackson DA, Pombo A (1998) Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells. *J Cell Biol* 140:1285–1295
- Nakamura H, Morita T, Sato C (1986) Structural organizations of replication domains during DNA synthetic phase in the mammalian nucleus. *Exp Cell Res* 165:291–297
- Waga S, Stillman B (1998) The DNA replication fork in eukaryotic cells. *Annu Rev Biochem* 67:721–751
- Watson JD, Crick FHC (1953) Molecular structure of nucleic acids. *Nature* 171:737–738