Molecular Cell
Review

DNA Replication—A Matter of Fidelity

Rais A. Ganai[1,2] and Erik Johansson[1,*]
[1]Department of Medical Biochemistry and Biophysics, Umeå University, SE 901 87 Umeå, Sweden
[2]Howard Hughes Medical Institute, Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, NY 10016, USA
*Correspondence: erik.tm.johansson@umu.se
http://dx.doi.org/10.1016/j.molcel.2016.05.003

The fidelity of DNA replication is determined by many factors, here simplified as the contribution of the DNA polymerase (nucleotide selectivity and proofreading), mismatch repair, a balanced supply of nucleotides, and the condition of the DNA template (both in terms of sequence context and the presence of DNA lesions). This review discusses the contribution and interplay between these factors to the overall fidelity of DNA replication.

Introduction

An appropriate level of fidelity during DNA replication ensures the ability of organisms to transfer genetic information from one generation to the next and contributes to the diversity of life. It is essential that the semiconservative duplication of DNA gives a nearly perfect end product, otherwise important genes might carry mutations that lead to disease or cell death. In this review, we will discuss the contributions of the DNA polymerases, mismatch repair proteins, DNA template, and nucleotide pool to the fidelity of DNA replication.

DNA replication is a tightly regulated process that begins with the activation of origins of DNA replication (reviewed by Rivera-Mulia and Gilbert, 2016). Replication forks are established at these origins, and these consist of numerous proteins that contribute to the synthesis of a new chromosome. Numerous proteins make up the Pol III holoenzyme complex in *Escherichia coli*, and the core of this complex consists of two identical DNA polymerases synthesizing both the leading and lagging DNA strands (McHenry, 2011). The situation is more complex in eukaryotes, and the initial purification of a replication progression complex by Gambus et al. (2006) contained numerous factors, but it was lacking the three eukaryotic replicative DNA polymerases: DNA polymerase α (Pol α), DNA polymerase δ (Pol δ), and DNA polymerase ε (Pol ε). Recently, a milestone was reached when the minimal set of proteins required for activation of a eukaryotic origin and the establishment of DNA synthesis was identified (Yeeles et al., 2015). Both Pol α and Pol ε were shown to be required, as was expected from a large body of studies over the past 40 years. Pol α consists of four subunits, of which two form the primase and two form the DNA polymerase (Pellegrini, 2012). The primase activity of Pol α is responsible for synthesizing the RNA primer that is required for DNA replication to start, and once this primer is laid down, the enzyme switches to DNA synthesis activity. The roles of Pol δ and Pol ε were less clear when the eukaryotic replication machinery was first being investigated, but over the last 10 years, an overwhelming body of evidence has shown that Pol ε copies the leading strand and that Pol δ copies the lagging strand during normal DNA replication. Genetic experiments with a Pol ε M644G mutator showed that the Pol ε mutant introduces errors specifically on the leading strand (Pursell et al., 2007), and using the same approach with a Pol δ L612M mutant, it was shown 1 year later that the Pol δ mutant introduces errors specifically on the lagging strand (Nick McElhinny et al., 2008). Initially this was only shown with reporter genes in the proximity of specific origins, but a whole-genome study confirmed that the strand bias was true for the Pol δ L612M mutant across the entire yeast genome (Larrea et al., 2010). Recently, these studies were challenged by a paper from Prakash and coworkers in which they claimed that Pol δ replicates both the leading and the lagging strands and that Pol ε is limited to proofreading errors made by Pol δ (Johnson et al., 2015). However, concerns were raised regarding several technical issues in that work, including very low mutation rates, suggesting that the reported results were influenced by suppressor mutations (Burgers et al., 2016).

The initial in vivo results suggesting that Pol ε copies the leading strand and Pol δ copies the lagging strand during normal DNA replication were also in agreement with biochemical and genetic experiments showing that Pol δ facilitates the removal of the RNA primer from each Okazaki fragment when providing Fen-1 (an endonuclease) with an optimal substrate (Garg et al., 2004). The biochemical evidence for this strand bias was further strengthened when a purified CMG complex (the replicative helicase) was shown to selectively position Pol ε on the leading strand (Georgescu et al., 2014). Finally, several independent groups have recently shown that Pol δ replicates the lagging strand and Pol ε replicates the leading strand across the genome by mapping the ribonucleotides (rNTPs) that are introduced by Pol δ and Pol ε on each strand (Clausen et al., 2015; Daigaku et al., 2015; Koh et al., 2015; Reijns et al., 2015). These examples, and other contributions by many research labs, have shown that there is a strong bias for Pol ε on the leading strand and for Pol δ on the lagging strand under normal DNA replication conditions. Thus, Pol δ and Pol ε have the greatest impact on the fidelity of DNA replication because together they synthesize at least 90% of the eukaryotic genome (Nick McElhinny et al., 2008). Pol α also has an impact on fidelity because it synthesizes up to about 10% of the genome each time the genome is replicated. Prokaryotes, archaea, and eukaryotes all have mismatch repair systems that recognize errors made by the replicative DNA polymerases. This lowers the mutation rate to a level that is acceptable for the propagation of the species but still allows for slow changes to the genetic code.