DNA Repair xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

DNA Repair

journal homepage: www.elsevier.com/locate/dnarepair

The complex choreography of transcription-coupled repair

Graciela Spivak*, Ann K. Ganesan

Biology Department, Stanford University, 385 Serra Mall, Stanford, CA 94305-5020, USA

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Nucleotide excision repair
Transcription-coupled repair
DNA damage
DNA repair

ABSTRACT

A quarter of a century has elapsed since the discovery of transcription-coupled repair (TCR), and yet our fascination with this process has not diminished. Nucleotide excision repair (NER) is a versatile pathway that removes helix-distorting DNA lesions from the genomes of organisms across the evolutionary scale, from bacteria to humans. TCR, defined as a subpathway of NER, is dedicated to the repair of lesions that, by virtue of their location on the transcribed strands of active genes, encumber elongation by RNA polymerases. In this review, we will report on newly identified proteins, protein modifications, and protein complexes that participate in TCR in *Escherichia coli* and in human cells. We will discuss general models for the biochemical pathways and how and when cells might choose to utilize TCR or other pathways for repair or bypass of transcription-blocking DNA alterations.

1. Introduction

The genetic material of all living organisms must be protected against the constant threats posed by environmental agents and byproducts of cellular metabolic processes. Even the simplest unicellular beings possess mechanisms for the prevention and repair of damage to their DNA. Several dedicated pathways that have evolved to deal with structurally different types of lesions or non-canonical forms of DNA will be addressed in other sections in this volume.

Here, we will focus on transcription-coupled repair (TCR), which has been defined as a subpathway of nucleotide excision repair (NER). NER is a ubiquitous mechanism that detects a variety of bulky lesions, including those that significantly distort the structure of DNA. The process begins with the recognition of a DNA lesion. Then, dual incisions of the damaged DNA strand, one on either side of the lesion, are produced. The lesion-bearing oligonucleotide is removed, a patch is synthesized using the undamaged strand as a template, and the patch is ligated to the contiguous strand. TCR was first described in rodent and human cells, then in *Escherichia coli*, in yeast and in other organisms (reviewed in [1,2]), and was defined as the recognition and repair of DNA lesions or structures occurring in the transcribed strand of active genes. Serendipity intervened: the observation of TCR was possible because the model lesions utilized in those initial experiments, cyclobutane pyrimidine dimers (CPDs), are repaired rather slowly in the genome as a whole, permitting detection of the faster repair of transcribed strands by TCR.

TCR occurs when a translocating RNA polymerase (RNAP) encounters an encumbrance to its forward progress. Arrested transcription complexes may severely affect cellular functions and survival, inhibiting the production of essential transcripts, blocking DNA replication, and signaling cell death pathways. Moreover, an RNAP able to bypass a lesion can generate mutant, perhaps deleterious, transcripts. RNAP pausing or arrest, with the mediation of TCR factors, may recruit NER proteins. In this perspective, we will examine current data and ideas and propose models of how these processes occur in prokaryotes and in highly evolved eukaryotes.

2. Global and transcription-coupled repair in bacteria

It has been widely accepted that in global genomic NER (GGR) in *E. coli*, a complex of the UvrA and UvrB proteins recognizes a DNA lesion and binds to it. UvrC then associates with UvrB and produces two nicks in the damaged DNA strand, one on either side of the lesion. UvrA dissociates from UvrB either before or after UvrC binds, and UvrB dissociates from the DNA after the dual incisions. DNA helicase I (the UvrD protein) and DNA polymerase I cooperate to remove the oligonucleotide containing the lesion and replace the missing bases using the undamaged complementary strand as a template. Finally, ligase I joins the newly synthesized sequence to the contiguous DNA strand. In TCR, RNAP, rather than the UvrAB complex, recognizes a lesion as it transcribes a gene. Typically, the lesion is located in the transcribed strand, causing the RNAP to stall or arrest. Initially, the arrested polymerase complex may occlude the lesion and prevent its repair; however, the Mfd (TCRF)

* Corresponding author. Tel.: +1 650 723 2425; fax: +1 650 725 1848.
E-mail addresses: gspivak@stanford.edu (G. Spivak), anngan@stanford.edu (A.K. Ganesan).

http://dx.doi.org/10.1016/j.dnarep.2014.03.025
1568-7864/

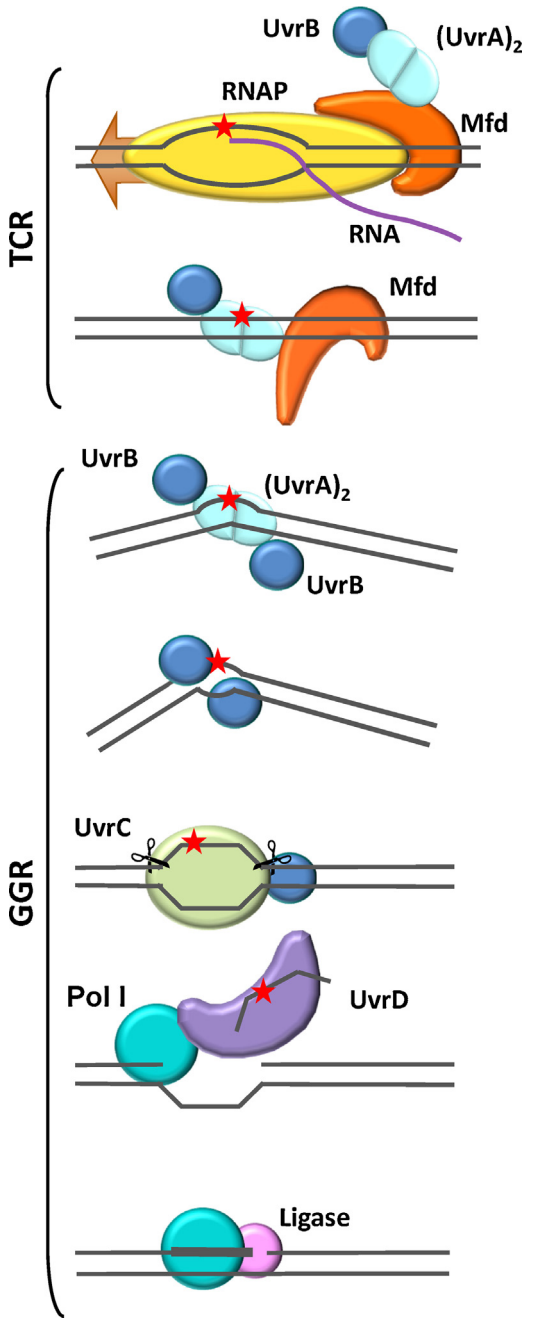

Fig. 1. *A conventional schematic of NER in E. coli.* In TCR, Mfd is activated by binding to a stalled RNAP. It dissociates the RNAP and nascent transcript from the DNA and recruits UvrA complexed with UvrB. The UvrAB complex binds to the DNA and recognizes and verifies the damage to be repaired. Repair then proceeds through the same reaction sequence as GGR. UvrA dissociates from the preincision complex leaving one or two molecules of UvrB bound to the DNA. UvrC interacts with UvrB and catalyzes two nicks in the DNA, one on either side of the lesion. The combined action of UvrD (helicase I) and DNA polymerase I removes the oligonucleotide containing the lesion, as well as UvrB and UvrC, from the DNA and results in the synthesis of a patch using the undamaged complementary strand as a template. DNA ligase I seals the patch to the contiguous DNA strand.

protein can interact with the polymerase, dissociate it from the lesion, and attract the UvrAB complex. Repair then proceeds as for GGR (Fig. 1). The unambiguous criterion of TCR is a strand bias in repair; the transcribed strand (TS) is repaired more rapidly than the nontranscribed strand (NTS).

Although the general scheme of NER is known in some detail, many questions remain. For example, the sequence of association and dissociation of the various proteins is not entirely clear (reviewed in [3]). Similarly, it is not certain that all the relevant proteins or all their functions in repair have been identified.

In the case of GGR, the exact composition of the UvrAB complex has been controversial. Currently, a variety of structural and kinetic studies support the proposal that the recognition–verification complex contains two molecules of UvrA and two molecules of UvrB ($UvrA_2UvrB_2$), but the exact configurations of the subunits in the protein complex and in the DNA–protein complex are still uncertain (reviewed in [3]). Similarly, the sequence of events leading to the formation of the incision complex is unclear. In part, this is due to the recent observation that UvrC can facilitate the binding of UvrB to DNA in the absence of UvrA [4]. In fact, it has been proposed that most or all of the UvrC in a cell is complexed with UvrB. This scenario is more plausible if there are only 10–20 molecules of UvrC in a cell [4,5] than if there are 400–500 [6,7]. If most of the UvrC in a cell is normally complexed with UvrB, previous ideas about the sequence of assembly of the subunits in the incision complex will require revision. In the case of TCR, the situation becomes even more complicated because of the presence of the RNAP, the other proteins in the elongation complex (EC), and the Mfd protein.

The typical core bacterial RNAP contains five subunits, $\alpha_2\beta\beta'\omega$. The holoenzyme also contains a σ factor that determines the promoter sequence recognized by the complex. For most genes expressed during the normal growth of *E. coli*, σ^{70} is used. The holoenzyme can transcribe DNA in vitro, and the process consists of several stages, including initiation, elongation, and termination. The σ factor (which binds strongly to the β′ subunit and weakly to the β subunit of RNAP [8–10]) is required for initiation but not for elongation, and it usually dissociates as the elongation complex (EC) forms and the RNAP moves away from the promoter region [11,12]. When σ^{70} is bound to the RNAP in vitro, it interferes with the interaction between the polymerase and Mfd [13,14], consistent with the observation that wild type *E. coli* Mfd does not release the polymerase from an initiation complex in vitro [15] and with observed patterns of repair [16,17]. These results indicate that Mfd will interact with the polymerase only after the EC forms. Furthermore, the interaction appears to require that the EC be stalled, possibly because, at least in vitro, Mfd makes several attempts before it interacts productively with the EC [18]. In principle, this should have important implications for the rate of repair.

One focus of study has been the factors determining the rate of repair in TCR compared to GGR. It has been proposed that the rate-limiting step in NER may be the small amount of UvrC available [4], but this idea does not account for the different rates of repair observed for TCR and GGR or for the TS compared to the NTS because UvrC is required in each case. One might imagine that the search for DNA damage by the pre-incision complex is a rate determinant, considering the vast amount of undamaged DNA that must be scanned for relatively rare lesions (reviewed in [19]) and that 1D scanning by the EC might lead to more rapid encounters with lesions in transcribed strands than 3D scanning by the UvrAB pre-incision complex; however, after the EC stalls, the pre-incision complex must then find it. The Mfd protein not only dissociates the arrested EC from a lesion but it also associates with UvrA, and the strand bias in TCR depends on the interaction between Mfd and UvrA. Mutants in which the UvrB homology domain of Mfd has been modified so that interaction with UvrA is compromised show the same rate of repair of UV induced damage in the TS as in the NTS even when the mutant proteins are capable of dissociating the EC from DNA [17]. Questions remain about the exact structure of the Mfd, UvrAB, DNA complex. For example, does the complex contain 1 molecule of Mfd, 2 of UvrA and 1 of UvrB or can it accommodate the $UvrA_2UvrB_2$ complex? Is UvrC associated with UvrB in this complex when the complex binds to DNA or does UvrC associate with UvrB after the complex binds to DNA? Recently, it has

Please cite this article in press as: G. Spivak, A.K. Ganesan, The complex choreography of transcription-coupled repair, DNA Repair (2014), http://dx.doi.org/10.1016/j.dnarep.2014.03.025