DNA Repair xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

DNA Repair

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

Molecular basis of transcriptional fidelity and DNA lesion-induced transcriptional mutagenesis

Liang Xu [a], Linati Da [a], Steven W. Plouffe [a], Jenny Chong [a], Eric Kool [b,*], Dong Wang [a,*]

[a] Skaggs School of Pharmacy and Pharmaceutical Science, University of California San Diego, La Jolla, CA 92093-0625, United States
[b] Department of Chemistry, Stanford University, Stanford, CA 94305-5080, United States

ARTICLE INFO

Article history:
Available online xxx

Keywords:
RNA polymerase II
Transcription elongation
Fidelity control
DNA damage recognition
Transcriptional lesion bypass
Synthetic nucleotide analogs
Unlocked nucleic acid
Nonpolar isostere
Chemical biology
Structural biology
Computational biology

ABSTRACT

Maintaining high transcriptional fidelity is essential for life. Some DNA lesions lead to significant changes in transcriptional fidelity. In this review, we will summarize recent progress towards understanding the molecular basis of RNA polymerase II (Pol II) transcriptional fidelity and DNA lesion-induced transcriptional mutagenesis. In particular, we will focus on the three key checkpoint steps of controlling Pol II transcriptional fidelity: insertion (specific nucleotide selection and incorporation), extension (differentiation of RNA transcript extension of a matched over mismatched 3′-RNA terminus), and proofreading (preferential removal of misincorporated nucleotides from the 3′-RNA end). We will also discuss some novel insights into the molecular basis and chemical perspectives of controlling Pol II transcriptional fidelity through structural, computational, and chemical biology approaches.

1. Introduction

Eukaryotic RNA polymerase II (Pol II) is one of the central enzymes for the first key step of gene expression [1]. During transcription, Pol II reads the DNA template and synthesizes a complementary RNA strand. The functional RNA molecules include precursors of protein-coding messenger RNAs as well as non-coding RNAs which may have important and diverse biological roles. Therefore, maintaining a highly accurate transfer of genetic information from DNA to RNA (high transcriptional fidelity) is essential for the process of life [2].

Pol II, as a highly specific "DNA reader" and "RNA writer", is able to maintain a low error rate during transcription (less than 10^{-5}) [3,4]. The Pol II active site forms a network of interactions that specifically recognizes cognate nucleoside triphosphates and excludes non-cognate ones through three fidelity checkpoint steps [3,5]. These three checkpoint steps include insertion (specific nucleotide selection and incorporation), extension (differentiation of RNA transcript extension of a matched over mismatched 3′-RNA terminus), and proofreading (preferential removal of misincorporated nucleotides from the 3′-RNA end) (Fig. 1) [3,5,6].

RNA Pol II has also been proposed to be a highly selective DNA damage sensor, since it constantly scans the transcribed genome during transcription [7,8]. In fact, a potentially lethal challenge that all cells and organisms must constantly face is the generation of harmful genomic DNA lesions caused by endogenous and environmental agents [9,10]. There can be as many as one million DNA lesions generated in a cell per day [11]. Many of these lesions cause significant DNA structural and chemical alterations. The presence of DNA lesions within highly transcribed genomic regions significantly alters Pol II transcription with deleterious consequences [12–15].

Biochemical and genetic studies have shown that the action Pol II takes when encountering DNA damage is lesion specific [7,13–17]. Pol II can either bypass, backtrack, or stall at DNA lesions. Pol II transcriptional bypass may cause misincorporation within the RNA transcript, termed transcriptional mutagenesis [15,18]. Pol II backtracking leads to intrinsic or transcription factor IIS (TFIIS)-mediated RNA transcript cleavage. Pol II stalling initiates a specialized DNA repair pathway, termed transcription-coupled repair (TCR) [13,15–17]. The TCR pathway, first discovered by the Hanawalt Lab, specifically repairs DNA lesions in the transcribed genome [13,16,19–23]. In this pathway, the Cockayne

* Corresponding authors at: University of California San Diego, Skaggs School of Pharmacy and Pharmaceutical Science, La Jolla, CA 92093-0625, United States; Department of Chemistry, Stanford University, Stanford, CA 94305-5080, United States.
E-mail addresses: kool@stanford.edu (E. Kool), dongwang@ucsd.edu (D. Wang).

http://dx.doi.org/10.1016/j.dnarep.2014.03.024
1568-7864/

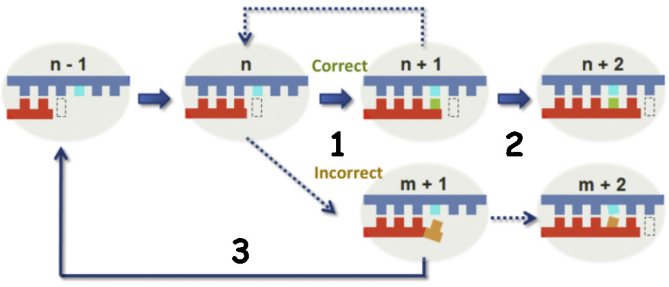

Fig. 1. Schematic model showing the three fidelity checkpoint steps of Pol II transcription. The matched nucleotide (green) is selected over an incorrect substrate (orange) during the incorporation step (step 1) and the RNA transcript can be efficiently elongated from a matched pair (step 2). In addition, proofreading is required to remove misincorporated nucleotides, backtrack, and reactivate the transcription process (step 3). The DNA template and RNA product are shown in blue and red, respectively. The active site of Pol II is shown in a dash box.

Syndrome B protein (CSB) is one of the first proteins recruited to the arrested Pol II site and is involved in the early stages of TCR, although the detailed recruiting mechanism remains to be elucidated [22,24–28]. Finally, as the last resort to remove persistently arrested Pol II, the arrested Pol II undergoes ubiquitylation and degradation [7,14,29]. Defects in genes involved in TCR cause premature aging and several human diseases, such as UV-sensitive syndrome, Cockayne Syndrome, Xeroderma Pigmentosum, Trichothiodystrophy, and Cerebro-Oculo-Facio-Skeletal Syndrome [13,30–35]. Several excellent reviews on transcription-coupled DNA damage processing can be found elsewhere [13–17]. In this review, we will focus on discussing some novel insights into the molecular basis and chemical perspectives in controlling Pol II transcriptional fidelity and DNA lesion-induced transcriptional mutagenesis through structural, computational and chemical biology approaches.

2. Recent studies on elucidating the molecular basis of transcriptional fidelity

2.1. Understanding the molecular basis of transcriptional fidelity through structural biology approaches

Pol II is an RNA polymerase containing 12 subunits with a total molecular weight of 550 kDa. A combination of genetic, biochemical, and structural studies have shed new light on the Pol II transcriptional mechanism at an atomic level over the last decade [5,36–56]. There are several excellent reviews focusing on Pol II enzymatic catalysis and transcriptional regulation published elsewhere [6,57–61]. Here we will focus on understanding the molecular basis of the three key checkpoint steps in controlling Pol II transcriptional fidelity during transcription elongation phase.

The first atomic three-dimensional crystal structures of Pol II (apo form) and the Pol II transcribing elongation complex were solved in 2001 by the Kornberg group [36,62]. The Pol II elongation complex is in a "pre-translocation" state in which the Pol II active site is still occupied by the newly added nucleotide at the 3′-RNA end. The structure of this Pol II elongation complex reveals two key structural features: first, the DNA template bends about 90 degrees as it crosses over the bridge helix and enters the Pol II active site; second, the RNA forms a short upstream RNA–DNA hybrid with the template DNA strand extending from the Pol II active site. In order to make the active site available for the next nucleotide addition, Pol II must translocate along the DNA template one base pair downstream, referred to as the "post-translocation" state (Fig. 2a) [40,41]. The Pol II elongation complex in a post-translocation state was subsequently solved using a synthetic RNA/DNA scaffold to mimic the transcription bubble during transcription [40,41]. In this structure, the active site of the Pol II elongation complex is empty and available for substrate binding, while the Pol II protein has almost an identical conformation with that of the pre-translocation state.

Having the Pol II elongation complex structure in a post-translocation state makes it possible to address how the nucleotide substrate is selected and bound to the Pol II active site using structural biology approaches. A major challenge to overcome during protein crystallization is to separate two kinetically coupled events: nucleotide binding and nucleotide addition. Two approaches were employed to allow nucleotide binding in the Pol II active site while preventing nucleotide addition at the 3′-RNA end [37,39–41]. In one strategy, the 3′-RNA terminus of the Pol II elongation complex at a "post-translocation" state was modified with 3′-deoxyribose. NTPs can bind to the active site but subsequent addition is prevented due to the lack of a 3′-OH on the RNA chain. An alternative strategy is to employ non-hydrolyzable NTP analogues to prevent nucleotide addition. Using these two approaches, a series of structures of Pol II elongation complexes with matched or mismatched substrates binding in the active site were determined.

These structural studies provided several important insights into understanding the molecular basis of the first checkpoint step of Pol II transcriptional fidelity (nucleotide selection and addition). The matched NTP binding to the Pol II active site in which its nucleobase forms a Watson–Crick base pair with the DNA template base is referred to as the "addition site" (Figs. 2b and 3a). A key structural feature is that, upon binding of the matched NTP, a conserved motif of Pol II, termed the trigger loop, switches from an inactive open state to an active closed state (Fig. 3a–c [37]). It is important to note that the closure of the active site is a common strategy for many other enzymes to greatly facilitate enzymatic efficiency and selectivity [63–65]. In addition, Pol II shares a universal two-metal ion catalytic mechanism with many other nucleic acids enzymes [66–70]. Two magnesium ions are required to bind the active site for Pol II catalytic activity (Fig. 3a–c). All key functional moieties of the substrate, including the nucleobase, sugar and triphosphate, are recognized by the Pol II active site through a substrate recognition network [37]. The nucleobase of NTP is base paired with the DNA template and further sandwiched by the 3′-RNA primer and Leu1081(Rpb1) from the Pol II trigger loop to ensure correct positioning in both lateral and vertical directions. Both hydroxyl groups of the sugar moiety are recognized by Pol II residues (hydrogen bonding networks formed by Arg446(Rpb1), Asn479(Rpb1), Gln1078(Rpb1), and Asn1082(Rpb1)). The triphosphate moiety is coordinated by two catalytic magnesium ions and His1085(Rpb1), Lys1020(Rpb2) and Arg766(Rpb2) (Fig. 3a). Altogether, these interactions form a network to ensure the NTP is correctly positioned for nucleotide addition and to also ensure correct nucleotide selection. In sharp contrast, the mismatched nucleotide is bound in an inverted conformation in which the nucleobase is facing away from the DNA template, referred to as the "entry site", and the trigger loop is left in the inactive open conformation. Therefore, the rate of misincorporation is several orders of magnitude slower than the rate of correct substrate incorporation [3,5]. Consistently, several mutations in the trigger loop and the nearby active site that significantly affect Pol II transcription fidelity and catalytic efficiency have been identified [52,53,56,71–78].

In the second fidelity checkpoint step, Pol II transcript extension from a mismatched 3′-RNA terminus is much less efficient than extension from a fully matched end. A series of Pol II structures containing a mismatched RNA terminus provided the structural basis for our understanding of this second checkpoint step. The mismatched 3′-RNA terminus causes a distortion in the RNA/DNA

Please cite this article in press as: L. Xu, et al., Molecular basis of transcriptional fidelity and DNA lesion-induced transcriptional mutagenesis, DNA Repair (2014), http://dx.doi.org/10.1016/j.dnarep.2014.03.024