Transcription pausing: biological significance of thermal fluctuations biased by repetitive genomic sequences

Masahiko Imashimizu and David B. Lukatsky

Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo, Japan; Department of Chemistry, Ben-Gurion University of the Negev, Beer Sheva, Israel

ABSTRACT
Transcription of DNA by RNA polymerase (RNAP) takes place in a cell environment dominated by thermal fluctuations. How are transcription reactions including initiation, elongation, and termination on genomic DNA so well-controlled during such fluctuations? A recent statistical mechanical approach using high-throughput sequencing data reveals that repetitive DNA sequence elements embedded into a genomic sequence provide the key mechanism to functionally bias the fluctuations of transcription elongation complexes. In particular, during elongation pausing, such repetitive sequence elements can increase the magnitude of one-dimensional diffusion of the RNAP enzyme on the DNA upstream of the pausing site, generating a large variation in the dwell times of RNAP pausing under the control of these genomic signals.

KEYWORDS
RNA polymerase; transcription pausing; repetitive genomic sequence; thermal fluctuations; Brownian ratchet

Introduction of statistical approach: why is it necessary?

When one characterizes the function of a biomolecular complex, determining a single particular energy state or a single particular structure of the complex does not always tell us about its function. This is because in the context of different cellular compartments with sizes ranging from tens to hundreds of nanometers, fluctuations of energy (or a structure) are not negligible but are essential in determining regulatory mechanisms [1].

The term fluctuations that we are using throughout this review means stochastic (i.e. random) variations of microscopic parameters of a biomolecular system due to the presence of thermal noise. Thermal noise is usually defined as random or nearly random, diffusive motion of molecules (e.g. water molecules, small organic molecules, lipids, peptides, proteins, nucleic acids, and their complexes etc.) constituting a living cell [1]. For example, due to the presence of thermal noise, a thermally fluctuating biomolecular complex maintains conformational heterogeneity, i.e. adopts multiple conformational states or structures rather than a single structure. In this review, we use the term microscopic state in order to describe a particular conformational state of a biomolecular complex. Such conformational diversity represents a biomolecular complex’s dynamic nature, which is necessary for performing its functions, which might include catalysis of chemical reactions, or providing a net directional motion along DNA. Our review stresses the key point that within a cell environment, thermal fluctuations are not entirely random but rather biased. In order to define the term biased thermal fluctuations, let us consider an example of a transcription factor searching for its specific binding site along the genomic DNA molecule [2]. In this case, diffusion of a transcription factor on DNA is not entirely random, but it is rather affected by the DNA sequence in a highly non-random way [3]. Such a non-random influence on random fluctuations of biomolecular complexes is termed here biased fluctuations. This review discusses the effect of biased thermal fluctuations on the process of DNA transcription by RNA polymerase (RNAP), especially on the pausing of the RNAP elongation complex.
The use of high-throughput sequencing (HT-seq) enables us to analyze tens of thousands of nucleic-acid-RNAP complexes in a single experiment [4]. This approach provides a comprehensive snapshot on the entire ensemble of different microscopic states (i.e. microscopic heterogeneity) of the active transcription elongation complex, consisting of DNA, RNA-DNA hybrid, and RNAP. Such data enable us to evaluate the effect of microscopic heterogeneity and DNA sequence features on the enzymatic function of RNAP. In particular, we have recently investigated the transcription elongation complex genome-wide in *E. coli* [4,5]. Our HT-seq data identify nascent RNA transcripts whose 3’ ends are associated with RNAP during elongation and provide the probability distribution of multiple states of RNAP elongation complexes. This data enable us to identify paused RNAPs at single base-pair resolution along the entire genomic DNA sequence [4,6].

Concepts of statistical mechanics can provide a theoretical framework for handling the HT-seq data containing the information on microscopic heterogeneity of a biomolecular complex. In particular, statistical mechanics enables us to connect the entire ensemble of different microscopic states to macroscopic quantities, such as the average RNAP density in a particular genomic coordinate, assuming thermodynamic equilibrium [5]. Our statistical mechanics approach stresses the key fact that the measured average RNAP density at a given genomic location depends on the DNA context surrounding this location [5]. On the other hand, concepts of chemical kinetics and thermodynamics, which have been historically applied for modeling biomolecular reactions, need to assume homogeneity in the system [7]. Therefore, such approaches would be able to interpret biological mechanisms involving microscopic heterogeneity only when connected with statistical mechanical concepts. Although the importance of thermal fluctuation and microscopic heterogeneity on molecular mechanisms of many biological reactions have been recognized [1], to the best of our knowledge, statistical mechanical concepts have not been applied for understanding the mechanism of RNAP pausing during transcription elongation.

In order to tackle this issue, we have developed the concept of non-consensus protein-DNA binding [3]. This concept allows us to describe microscopic heterogeneity stemming from thermal fluctuations as entropy-dominated free energy, which strongly depends on repetitive DNA sequence elements associated with a protein. Such a statistical fluctuation approach allows us to achieve a quantitative, probabilistic, differentiation between statistical ensembles of the two different groups of DNA sequences with high and low affinities to a specific target protein. In particular, one group of DNA sequences with high affinity to a specific protein has extensive repetitive DNA sequence elements (characterized as a relatively low average free energy of non-consensus protein-DNA binding); another group of DNA sequences with low affinity to the same protein has no repetitive sequence elements (characterized as a relatively high average free energy) [5]. We have shown the significance of multiple functional conformational states of a biomolecular system, which contrasts to the commonly adopted view that thermal fluctuations destabilize specific, structural states of a biomolecular system [5].

**Biology of transcription pausing**

One may be comfortable when riding on a train, even though the train moves very fast. This is because the movement of a train is a homogeneous rail tracking, where every step of the movement is exactly the same (relative to the size of a train). On the other hand, one would be extremely uncomfortable when riding on an RNAP transcribing RNA along DNA, because during elongation, the net speed of RNAP’s forward movement depends on DNA sequence heterogeneity, i.e. the speed of the RNAP movement per base is not constant but it is rather characterized by some probabilistic distribution. We have defined such RNAP movement as heterogeneous DNA tracking [8]. At certain positions, the RNAP moves slowly with a dwell time per base of ~1 sec or longer, and this is defined as transcription pausing [9–11].

Transcription pausing is a widespread phenomenon [8,12,13]. In promoter-proximal regions, RNAP pausing is observed in 10–40% of *Drosophila* genes [14–16] and 10–20% of *E. coli* genes [17]. There is a long research history in studying the phenomena of transcription pausing in both eukaryotic and prokaryotic cells. Those studies are also closely linked with the development of high-throughput techniques analyzing the nascent 3’ RNA ends of transcripts associated with transcribing RNAP [12,18]. Recent HT-seq studies from *E. coli* to human cells reveal an increased density of pausing not only in promoter-proximal regions but
also at several sites through the length of a gene, both in sense and anti-sense orientations [4,6,19,20].

Accumulating lines of evidence suggest that one of the major roles for RNAP pausing in promoter-proximal regions is to maintain active promoters in highly expressed genes. In eukaryotes, when RNAPII stably pauses just downstream of a promoter, it keeps chromatin open [13]. Thus a dynamic equilibrium between open and closed, bistable chromatin states in the vicinity of the promoter is biased toward an open state, which can be essential for providing sufficient time to associate regulatory transcription factors with the promoter and with the paused RNAPII during transcription initiation [21]. Another important role of RNAP pausing, which is observed outside of promoter-proximal regions, is a kinetic coupling of transcription to other chromosome machines, such as the spliceosome in eukaryotic cells [12,22]. For instance, pausing of RNAPII due to a roadblock by an eleven zinc finger protein affects co-transcriptional alternative splicing in the CD45 gene, and increases exon inclusion frequency [23].

Transcription pausing does not always play positive regulatory roles. One negative consequence of pausing is genome rearrangements, which are highly toxic to cells [24–26]. Pausing can be long-lived when it is stabilized by backtracking of RNAP. Backtracking is enhanced when RNAP misincorporates a nucleotide that is noncomplementary to the template DNA, or when encountering protein roadblocks on DNA [8,27,28]. During backtracking, the 3’ end of the nascent RNA is disengaged from the active site of RNAP. Such an event profoundly stabilizes the paused state [29,30]. This stably paused complex will collide with other proteins that are moving along the genomic DNA such as the DNA replication complex [31,32]. Collisions of paused transcription complexes with replisomes are frequent occurrences that cause DNA damage usually via the formation of R-loops, molecular structures composed of an RNA-DNA hybrid and a looped-out single DNA strand [24,31,32]. The R-loop is sensitive to DNA breaks and mutations in the single-stranded DNA region. Surprisingly, recent sequencing studies estimate that R-loops occupy 5% or more of the length of the genome in some cell types [33–35]. Accumulation of R-loops leads to human diseases including cancer development and neurodegeneration [36]. Maintenance of genome integrity is a crucial issue for both prokaryotic and eukaryotic cells, and hence, both types of cells are equipped with several different co-transcriptional mechanisms that minimize transcription-replication conflicts [37].

In prokaryotes, a translating ribosome on the nascent transcript behind RNAP may prevent many of the pervasive RNAP pauses that are stabilized by backtracking [38]. In particular, the leading ribosome on the nascent RNA behind RNAP may always prevent RNAP [39] from moving backwards on the RNA. In the absence of ribosome and translation coupling with transcription, RNAP can backtrack for a long distance. Despite this, recent genome-wide studies identified considerable backtracked pauses [4] and likely resultant R-loops formed within the coding regions of genes of prokaryotes [33] suggest that the association of translating ribosomes with the 5’ end of nascent RNA may not be always sufficient to prevent spontaneous RNAP backtracking.

**What factors discriminate between physiologically-relevant pausing and physiologically-irrelevant pausing?**

Transcription pausing stabilized by RNAP backtracking appears to have physiological roles for specific regulatory purposes [40,41]. However, pausing, if it ubiquitously occurs throughout transcribed regions in the genome, could be harmful since it provides a source of transcription-replication conflicts [37]. What factors affect pausing in a cell? What factors help to suppress excessive pausing events from occurring in a cell?

Recently, we and other research groups independently developed a method allowing the genome-wide analysis of transcription pausing, using *E. coli* cells as a model system. Using their methods, we and others identified a highly conserved pause-inducing element (PIE) [4,10,42]. The conserved motifs of PIE represented as GNNNNNTGCG are spread out along the entire length of the RNA-DNA hybrid (Fig. 1) [4]. It has been shown that the PIE sequence can impede forward translocation of RNAP, as well as the following NTP addition step to cause pausing with ~1 sec or longer duration. These long pauses are detectable by HT-seq experiments [4]. However, deeper analysis demonstrated that the presence of the PIE sequence alone is not responsible for RNAP pausing on the *E. coli* chromosome; an additional genome-wide mechanism is necessary to cause pausing [5]. Finally, we then showed that thermal fluctuations biased by
repetitive DNA sequences constitute the global mechanism determining the fate of RNAP pausing [5].

In order to investigate the PIE-independent mechanism for pausing in *E. coli*, we defined “false” PIES in such a way that the information content [43] (i.e. the index of the base conservation) of DNA sequence is as high as “true” PIES, however, no reliable pausing was experimentally detected for false PIES [5]. Therefore, the specific, consensus motif characterizing the true PIE group is the same as the corresponding consensus motif for the false PIE group. This is despite the fact that pausing events are detected in the former group but not in the latter group. Surprisingly, we showed that true PIES constitute only 3%, and the rest 97% of DNA sequences constitute false PIES. Most importantly, we identified that DNA sequences flanking true PIES contain more repetitive sequence elements than the sequences flanking the false PIES even though there is no significant difference in the average GC content of those flanking sequences between true and false PIES [5].

Control of transcription pausing by repetitive genomic sequences

How do repetitive sequences enhance pausing? An intuitive explanation follows. Similar to our past work modeling protein-DNA interactions, here we assume that the binding energy of RNAP to DNA per base is dominated by thermal fluctuations [3,44–49]. Thus, even though RNAP may pause at PIE, thermal, diffusive motion (i.e. sliding) of the paused complex on DNA determines the overall stability of the paused state. In other words, the stability of the paused state is determined by the overall free energy of the elongation complex (dominated by entropy) rather than by the binding energy of a unique “static” state. When DNA sequences located upstream of consensus PIES are enriched with certain repetitive elements, the number of possible conformational states of a paused complex will increase [5]. As a result of this effect, statistically on average, upon pausing, RNAP has a higher probability to diffuse backwards and upstream of such PIES towards DNA sequences enriched in repetitive elements (Fig. 2). We use the term diffusional backtracking stabilized by biased thermal fluctuations in order to describe this effect. This intuitive explanation is indeed validated by our quantitative statistical mechanical modeling [44]. In particular, in order to estimate the statistical magnitude of non-consensus RNAP binding to DNA along the entire *E. coli* genome, we computed the free energies of non-consensus RNAP-DNA binding for three groups of DNA sequences, those surrounding true PIES, false PIES and a control non-PIE group (Fig. 1). Both the experimentally determined true PIES and the predicted false PIES consist of very similar 11-nt-long consensus motifs represented by GNNNNNNTGCG, located in the center of the sequences (Fig. 1). The control group consisting of 20,000 sequences was selected from the *E. coli* genome at randomly chosen locations, by using a position weight matrix of the 11-nt-long true PIES (Fig. 1) [5]. Thus, DNA sequences belonging to this control group lack the 11-nt-long consensus motifs. Strikingly, our analysis reveals that the DNA context upstream of true PIES is special and characterized by a significantly lower average free energy of non-consensus binding as compared to false PIE regions (Fig. 1). We interpreted that such an asymmetric profile and lower magnitude of the non-consensus free energy leads to enhanced RNAP sliding on DNA upstream of...
PIE when RNAP temporarily pauses during elongation (Figs. 1 and 2), We also interpreted that the higher magnitude of the non-consensus free energy surrounding false PIE prevents such RNAP sliding (Figs. 1 and 2). These notions were verified by in vitro experiments [5].

**Biological significance of Brownian-ratchet type mechanism controlling pausing**

It is notable that the statistical control of transcription pausing by biased thermal fluctuations on repetitive genomic sequences belongs to a type of Brownian ratchet mechanism of transcription elongation [7,8,50,51]. According to the original view, during elongation, RNAP can diffuse forward and backward on DNA but the net forward motion on DNA is generated and biased by cognate NTP binding to the RNAP active site of the elongation complex [52]. This mechanism prevents spontaneous backward motion by allowing elongation to the next cycle through condensation of the cognate NTP to the 3' end of the transcript [7,8,50]. Based on a molecular structure of the elongation complex [53] combined with biochemical
and single-molecule measurements [52,54], this mechanism assumes two rapid thermal fluctuations during elongation: (i) translocation fluctuation of the RNA-DNA hybrid and (ii) conformational fluctuation of the RNAP active site that includes the two protein motifs termed bridge helix and the trigger loop [53,55].

It is important to note that this proposed ratchet mechanism describes a pause-free elongation (homo- geneous tracking). However, it has not sufficiently described sequence-specific pausing due to heterogeneous DNA tracking [8]. Our findings thus extend the Brownian-ratchet mechanism of transcription elongation by incorporating the entropy-dominated mechanism controlling RNAP pausing (Fig. 2). In particular, upon pausing at any PIE sequence, the cognate NTP binding to the RNAP active site is inhibited, allowing time for RNAP to diffuse backward on DNA. Repetitive sequence elements upstream of PIE can increase the number of the paused complex conformations induced by thermal fluctuations, and therefore enhance diffusional backtracking of RNAP. On the other hand, non-repetitive sequences do not significantly increase the number of conformations of the complex during pausing, which leaves RNAP in a non-backtracked conformation, allowing RNAP to resume elongation rapidly (Fig. 2). This mechanism determines (statistically) the fate of pausing at PIE using Brownian motion biased by repetitive DNA sequences (Fig. 2).

We suggest therefore that the positive (i.e. enrichment of repetitive DNA sequence elements upstream of consensus PIE) and negative (i.e. enrichment of non-repetitive DNA sequence elements upstream of consensus PIE) genomic-design strategies [47] for controlling RNAP pausing represent a physiologically crucial mechanism for E. coli cells (Fig. 2). This strategy prevents pervasive pausing at all such PIE sequences. If it occurred everywhere it would lead to extensive DNA damage via transcription-replication collisions. By enhancing pausing at defined places where repetitive sequences are kept, the cell may maintain mechanisms for regulating pausing and pausing in turn can regulate different stages of gene transcription (i.e. initiation, elongation and termination) and the related biological processes such as transcript degradation. Indeed, our HT-seq measurement of pausing throughout the E. coli genome showed that true PIEs are enriched in regulatory regions including 5’ untranslated regions, anti-sense transcripts-coding regions, and intrinsic transcription terminators (Fig. 2) [4,5]. Future characterization of false PIEs in the context of transcription-replication conflict and the resultant R-loop formations will help to further elucidate biological significance of the negative design strategy for transcription pausing.

**Abbreviations**

- bp base pair
- nt nucleotide
- NTP ribonucleoside triphosphate
- HT-seq high-throughput sequencing
- PIE pause-inducing element
- RNAP RNA polymerase

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

**Acknowledgments**

We thank Donald L. Court for critical reading of the manuscript.

**Funding**

This work was supported by the Japan Society for the Promotion of Science (JSPS) (number 16H06692).

**References**

[1] Nelson PC, Radosavljevic M, Bromberg S. Biological physics: energy, information, life. New York: W.H. Freeman and Co.; 2008.

[2] Shvets AA, Kolomeisky AB. Sequence heterogeneity accelerates protein search for targets on DNA. J Chem Phys. 2015;143:245101.doi:10.1063/1.4937938.

[3] Afek A, Schipper JL, Horton J, et al. Protein-DNA binding in the absence of specific base-pair recognition. Proc Natl Acad Sci USA. 2014;111:17140–17145. doi:10.1073/pnas.1410569111.

[4] Imashimizu M, Takahashi H, Oshima T, et al. Visualizing translocation dynamics and nascent transcript errors in paused RNA polymerases in vivo. Genome Biol. 2015;16:98. doi:10.1186/s13059-015-0666-5.

[5] Imashimizu M, Afek A, Takahashi H, et al. Control of transcriptional pausing by biased thermal fluctuations on repetitive genomic sequences. Proc Natl Acad Sci USA. 2016;113:E7409–E7417. doi:10.1073/pnas.1607760113.

[6] Churchman LS, Weissman JS. Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature. 2011;469:368–373. doi:10.1038/nature09652.
[7] Shimamoto N. Nanobiology of RNA polymerase: biological consequence of inhomogeneity in reactant. Chem Rev. 2013;113:8400–8422. doi:10.1021/cr400066b.

[8] Imashimizu M, Shimamoto N, Oshima T, et al. Transcription elongation: Heterogeneous tracking of RNA polymerase and its biological implications. Transcription. 2014;5:e28285-1–10. doi:10.4161/trns.28285.

[9] Herbert KM, La Porta A, Wong BJ, et al. Sequence-resolved detection of pausing by single RNA polymerase molecules. Cell. 2006;125:1083–1094. doi:10.1016/j.cell.2006.04.032.

[10] Larson MH, Mooney RA, Peters JM, et al. A pause sequence enriched at translation start sites drives transcription dynamics in vivo. Science. 2014;344:1042–1047. doi:10.1126/science.1251871.

[11] Larson MH, Landick R, Block SM. Single-molecule studies of RNA polymerase: one singular sensation, every little step it takes. Mol Cell. 2011;41:249–262. doi:10.1016/j.molcel.2011.01.008.

[12] Mayer A, Landry HM, Churchman LS. Pause & go: from the discovery of RNA polymerase pausing to its functional implications. Curr Opin Cell Biol. 2017;46:72–80. doi:10.1016/j.ceb.2017.03.002.

[13] Adelman K, Lis JT. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat Rev Genet. 2012;13:720–731. doi:10.1038/nrg3293.

[14] Zeitlinger J, Stark A, Kellis M, et al. RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nat Genet. 2007;39:1512–1516. doi:10.1038/ng.2007.26.

[15] Guenthner MG, Levine SS, Boyer LA, et al. A chromatin landark and transcription initiation at most promoters in human cells. Cell. 2007;130:77–88. doi:10.1016/j.cell.2007.05.042.

[16] Muse GW, Gilchrist DA, Nechaev S, et al. RNA polymerase is poised for activation across the genome. Nat Genet. 2007;39:1507–1511. doi:10.1038/ng.2007.21.

[17] Marr MT, Roberts JW. Function of transcription cleavage factors GreA and GreB at a regulatory pause site. Mol Cell. 2000;6:1275–1285. doi:10.1016/S1097-2765(00)00126-X.

[18] Jonkers I, Lis JT. Getting up to speed with transcription elongation by RNA polymerase II. Nat Rev Mol Cell Biol. 2015;16:167–177. doi:10.1038/nrm3953.

[19] Mayer A, di Iulio J, Maleri S, et al. Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. Cell. 2015;161:541–554. doi:10.1016/j.cell.2015.03.010.

[20] Nojima T, Gomes T, Grosso AR, et al. Mammalian NETSeq reveals genome-wide nascent transcription coupled to RNA processing. Cell. 2015;161:526–540. doi:10.1016/j.cell.2015.03.027.

[21] Rogatsky I, Adelman K. Preparing the first responders: building the inflammatory transcriptome from the ground up. Mol Cell. 2014;54:245–254. doi:10.1016/j.molcel.2014.03.038.

[22] Carrillo Oesterreich F, Preibisch S, Neugebauer KM. Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. Mol Cell. 2010;40:571–581. doi:10.1016/j.molcel.2010.11.004.

[23] Shukla S, Kavak E, Gregory M, et al. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature. 2011;479(7371):74–79. doi:10.1038/nature10442.

[24] Nudler E. RNA polymerase backtracking in gene regulation and genome instability. Cell. 2012;149:1438–1445. doi:10.1016/j.cell.2012.06.003.

[25] Aguilera A, Garcia-Muse T. R loops: from transcription byproducts to threats to genome stability. Mol Cell. 2012;46:115–124. doi:10.1016/j.molcel.2012.04.009.

[26] Helmrich A, Ballarino M, Nudler E, et al. Transcription-replication encounters, consequences and genomic instability. Nat Struct Mol Biol. 2013;20:412–418. doi:10.1038/nsmb.2543.

[27] Imashimizu M, Oshima T, Lubkowska L, et al. Direct assessment of transcription fidelity by high-resolution RNA sequencing. Nucleic Acids Res. 2013;41:9090–9104. doi:10.1093/nar/gkt698.

[28] James K, Gamba P, Cockell SJ, et al. Misincorporation by RNA polymerase is a major source of transcription pausing in vivo. Nucleic Acids Res. 2017;45(5):1105–1113. doi:10.1093/nar/gkw969.

[29] Nudler E, Mustaev A, Lukhtanov E, et al. The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. Cell. 1997;89:33–41. doi:10.1016/S0092-8674(00)80180-4.

[30] Komissarova N, Kashlev M. RNA polymerase switches between inactivated and activated states By translocating back and forth along the DNA and the RNA. J Biol Chem. 1997;272:15329–15338. doi:10.1074/jbc.272.24.15329.

[31] Dutta D, Shatalin K, Epshtein V, et al. Linking RNA polymerase backtracking to genome instability in E. coli. Cell. 2011;146:533–543. doi:10.1016/j.cell.2011.07.034.

[32] McGary K, Nudler E. RNA polymerase and the ribosome: the close relationship. Curr Opin Microbiol. 2013;16:112–117. doi:10.1016/j.mib.2013.01.010.

[33] Leela JK, Syeda AH, Anupama K, et al. Rho-dependent transcription termination is essential to prevent excessive genome-wide R-loops in Escherichia coli. Proc Natl Acad Sci USA. 2013;110:258–263. doi:10.1073/pnas.1213123110.

[34] Wahba L, Costantino L, Tan FJ, et al. S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. Genes Dev. 2016;30:1327–1338. doi:10.1101/gad.280834.116.

[35] Sanz LA, Hartono SR, Lim YW, et al. Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals. Mol Cell. 2016;63:167–178. doi:10.1016/j.molcel.2016.05.032.

[36] Santos-Pereira JM, Aguilera A. R loops: new modulators of genome dynamics and function. Nat Rev Genet. 2015;16:583–97. doi:10.1038/nrg3961.

[37] Hamperl S, Cimprich KA. Conflict resolution in the genome: How transcription and replication make it
[38] Proshkin S, Rahmouni AR, Mironov A, et al. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. Science. 2010;328:504–508. doi:10.1126/science.1184939.

[39] Miller OL Jr, Hamkalo BA, Thomas CA Jr. Visualization of bacterial genes in action. Science. 1970;169:392–395. doi:10.1126/science.169.3943.392.

[40] Nechaev S, Fargo DC, dos Santos G, et al. Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. Science. 2010;327:335–338. doi:10.1126/science.1181421.

[41] Lemay JF, Larochelle M, Marguerat S, et al. The RNA exosome promotes transcription termination of backtracked RNA polymerase II. Nat Struct Mol Biol. 2014;21:919–926. doi:10.1038/nsmb.2893.

[42] Vvedenskaya IO, Vahedian-Movahed H, Bird JG, et al. Transcription. Interactions between RNA polymerase and the “core recognition element” counteract pausing. Science. 2014;344:1285–1289. doi:10.1126/science.1253458.

[43] Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 1990;18:6097–6100. doi:10.1093/nar/18.20.6097.

[44] Sela I, Lukatsky DB. DNA sequence correlations shape nonspecific transcription factor-DNA binding affinity. Biophys J. 2011;101:160–166. doi:10.1016/j.bpj.2011.04.037.

[45] Afek A, Cohen H, Barber-Zucker S, et al. Nonconsensus protein binding to repetitive DNA sequence elements significantly affects eukaryotic genomes. PLoS Comput Biol. 2015;11:e1004429. doi:10.1371/journal.pcbi.1004429.

[46] Afek A, Lukatsky DB. Genome-wide organization of eukaryotic preinitiation complex is influenced by nonspecific protein-DNA binding. Biophys J. 2013;104:1107–1115. doi:10.1016/j.bpj.2013.01.038.

[47] Afek A, Lukatsky DB. Positive and negative design for nonspecific protein-DNA binding affinity in the vicinity of functional binding sites. Biophys J. 2013;105:1653–1660. doi:10.1016/j.bpj.2013.08.033.

[48] Afek A, Lukatsky DB. Nonspecific protein-DNA binding is widespread in the yeast genome. Biophys J. 2012;102:1881–1888. doi:10.1016/j.bpj.2012.03.044.

[49] Afek A, Sela I, Musa-Lempel N, et al. Nonspecific transcription-factor-DNA binding influences nucleosome occupancy in yeast. Biophys J. 2011;101:2465–2475. doi:10.1016/j.bpj.2011.10.012.

[50] Guajardo R, Sousa R. A model for the mechanism of polymerase translocation. J Mol Biol. 1997;265:8–19. doi:10.1006/jmbi.1996.0707.

[51] Svetlov V, Nudler E. Macromolecular micromovements: how RNA polymerase translocates. Curr Opin Struct Biol. 2009;19:701–707. doi:10.1016/j.sbi.2009.10.002.

[52] Bar-Nahum G, Epshtein V, Ruckenstein AE, et al. A ratchet mechanism of transcription elongation and its control. Cell. 2005;120:183–193. doi:10.1016/j.cell.2004.11.045.

[53] Gnatt AL, Cramer P, Fu J, et al. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. Science. 2001;292:1876–1882. doi:10.1126/science.1059495.

[54] Abbondanzieri EA, Greenleaf WJ, Shaevitz JW, et al. Direct observation of base-pair stepping by RNA polymerase. Nature. 2005;438:460–465. doi:10.1038/nature04268.

[55] Wang D, Bushnell DA, Westover KD, et al. Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. Cell. 2006;127:941–954. doi:10.1016/j.cell.2006.11.023.