Biochimica et Biophysica Acta 1819 (2012) 482–493

Contents lists available at SciVerse ScienceDirect

Biochimica et Biophysica Acta

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

Review

Intra-nuclear mobility and target search mechanisms of transcription factors: A single-molecule perspective on gene expression☆

Davide Normanno [a,b,c,d,*], Maxime Dahan [b,c], Xavier Darzacq [a,c,**]

[a] Institut de Biologie de l'Ecole normale supérieure (IBENS), CNRS UMR 8197, Ecole normale supérieure, 46, Rue d'Ulm, 75005 Paris, France
[b] Laboratoire Kastler Brossel (LKB), CNRS UMR 8552, Ecole normale supérieure, Université Pierre et Marie Curie — Paris 6, 46, Rue d'Ulm, 75005 Paris, France
[c] Transcription Imaging Consortium, HHMI Janelia Farm Research Campus, 19200, Helix drive, Ashburn VA, USA
[d] Fondation Pierre-Gilles de Gennes, 29, Rue d'Ulm, 75005 Paris, France

ARTICLE INFO

Article history:
Received 28 November 2011
Received in revised form 26 January 2012
Accepted 3 February 2012
Available online 10 February 2012

Keywords:
Target search
Transcription factor
Gene expression regulation
Facilitated diffusion
Single-molecule analysis

ABSTRACT

Precise expression of specific genes in time and space is at the basis of cellular viability as well as correct development of organisms. Understanding the mechanisms of gene regulation is fundamental and still one of the great challenges for biology. Gene expression is regulated also by specific transcription factors that recognize and bind to specific DNA sequences. Transcription factors dynamics, and especially the way they sample the nucleoplasmic space during the search for their specific target in the genome, are a key aspect for regulation and it has been puzzling researchers for forty years. The scope of this review is to give a state-of-the-art perspective over the intra-nuclear mobility and the target search mechanisms of specific transcription factors at the molecular level. Going through the seminal biochemical experiments that have raised the first questions about target localization and the theoretical grounds concerning target search processes, we describe the most recent experimental achievements and current challenges in understanding transcription factors dynamics and interactions with DNA using in vitro assays as well as in live prokaryotic and eukaryotic cells. This article is part of a Special Issue entitled: Nuclear Transport and RNA Processing.

1. Introduction

Gene expression is the process by which the information encoded into genomic DNA is used by cells to synthesize a functional gene product through a cascade of steps consisting of transcription, post-transcriptional control, mRNA splicing, translation, and post-translational modifications. Regulation of gene expression is crucial for cells to achieve the control over timing, location and amount of gene products. This, in turn, potentiates cell functions and adaptation to internal and external stimuli. All steps in gene expression are tightly regulated and, in the case of transcription, regulation is orchestrated mainly through a class of nuclear proteins termed as transcription factors (TFs). TFs can either directly regulate single genes or initiate cascade events that trigger, in a time-controlled manner, the successive activation of several genes. This is essential for the appropriate control of cell cycle [1,2], adaptability, metabolism and differentiation [3] as well as for organ and tissue development [4]. Many aberrant events culminating with the development of tumors and cancer arise from breakdowns in transcription regulation. For instance, TFs are often overrepresented in tumors [5] and a third of human developmental disorders have been attributed to dysfunctional TFs [6]. In addition, alteration in the activity and regulatory specificity of TFs is likely to be a major source for phenotypic diversity and evolutionary adaptation [7,8]. The final mapping of the human genome has revealed that human chromosomes contain about 23,000 protein-coding genes [9]. About 2/3 of these genes are expressed to detectable levels [10] and around 6000 to levels sufficient to observe nascent, Bromo-Uridine labeled, mRNA spots in the nucleus of cultured cells [11]. Furthermore, transcription of genes is neither random nor uniform: genes are transcribed with widely different bursting kinetics [12,13], and small variations in the expression level of certain genes may lead to drastic phenotypes or severe illnesses [14].

Most of our basic knowledge about transcription and transcription regulation derives from molecular biology and genetics. Only recently, growing efforts have been devoted to understand transcription quantitatively [15] and dynamically [16] also by directly observing transcription in living systems [17]. At the biochemical level, transcription implies "copying" the information encoded into genomic DNA to RNA transcripts (mRNA, tRNA, or rRNA). In metazoans, RNA Polymerase II (RNA Pol II) machinery transcribes protein-coding

☆ This article is part of a Special Issue entitled: Nuclear Transport and RNA Processing.

* Correspondence to: D. Normanno, Institut de Biologe de l'Ecole normale supérieure (IBENS). CNRS UMR 8197. Ecole normale supérieure. 46, Rue d'Ulm –75005 Paris, France.

** Correspondence to: X. Darzacq, Functional Imaging of Transcription Institut de Biologie de l'Ecole normale supérieure (IBENS), CNRS UMR 8197, 46, Rue d'Ulm 75230, Paris, Cedex 05, France. Tel.: +33 1 44 32 34 62 (Office); fax: +33 1 44 32 39 41.

E-mail addresses: normanno@biologie.ens.fr (D. Normanno), darzacq@biologie.ens.fr (X. Darzacq).

1874-9399/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbagrm.2012.02.001

genes and is responsible for the synthesis of all mRNAs [18]. RNA Pol II machinery is composed of 12 subunits named RPB1 to 12 (or RPB A to L), also termed as the core 12. This unique enzyme transcribes nearly 23,000 genes and therefore the choice of the gene to be transcribed is not due to its intrinsic specificity but instead to hundreds of co-regulators that coordinate its selectivity. Gene activation and transcription by RNA Pol II machinery follow a set of common steps including the assembly of the Pre-Initiation Complex (PIC). PIC is a large macromolecular complex that engages on promoters prior to transcription, its composition is highly regulated and modulates its global specificity [19] during differentiation [20]. Importantly, PIC and elongation components interact with specific TFs, which recognize and bind to specific DNA sequences via a DNA binding domain (DBD), and direct transcription initiation to specific promoters [21] in order to specify a particular response on a particular gene.

The availability of specific TFs at specific sites and, consequently, their mobility in the nucleoplasm are crucial elements for gene expression regulation. In fact, drastically different scenarios can be envisioned depending whether TFs are limiting or are in large excess and depending if they are organized in static pre-assembled structures, usually denominated as transcription factories [22], or if they are capable of effective displacements between different nuclear loci. For instance, gene expression has been shown to become stochastic when TFs are limiting [23]. Many of these aspects are still elusive and the aim of this review is to point out what is known about TFs dynamics and target search mechanisms and their role in gene expression regulation with a single-molecule perspective. In fact, in order to quantitatively understand gene expression regulation at the transcriptional level that may, in turn, pave the way to design specific synthetic regulatory networks [24] for disease treatments, it is necessary to understand the mechanisms underlying TFs target localization at the molecular level.

2. The origins of target (re)search

Barbara McClintock, while studying the mechanisms of mosaic color patterns in maize, discovered interactions between two genetic loci and challenged the concept of the genome as a static set of instructions passed between generations already in 1950 [25]. Nevertheless, the very first discovery of a molecular network underlying and controlling gene expression through a feedback mechanism is widely considered to be the identification of the *lac* operon by Jacob and Monod in 1961 [26]. In particular, Jacob and Monod identified the regulation mechanism that controls lactose metabolism in *E. coli*, in response to external and internal conditions, via the activation and repression of a cluster of three genes, the *lac*ZYA genes. In 1965, Bourgeois et al. showed by suppression mutation experiments that a single specific transcription factor, the *lac* repressor or LacI, was regulating the *lac*ZYA genes [27] while, in 1967, Gilbert and Müller-Hill proved that the LacI had the capability to recognize and directly bind to a specific DNA sequence (*lac* primary operator) placed downstream the promoter region [28].

Basically, the *lac* repressor behaves as a transcriptional switch for the *lac* operon: in the absence of lactose, LacI is strongly bound to the *lac* operator and acts as a roadblock along the DNA template preventing RNA polymerase to transcribe the *lac*ZYA genes, when they are unnecessary (see Fig. 1A). On the contrary, in the presence of lactose in the surrounding medium and in the absence of glucose inside the cell, LacI dissociates from the DNA operator and the *lac*ZYA genes can be transcribed. More precisely, in *E. coli* there are two more DNA (pseudo-)operators located within few hundred of bps around the primary operator and the *lac* repressor can simultaneously bind to two out of the three operators and induce a loop in the DNA [29–31] (see [32] for a more detailed description of the *lac* system). The ability of DNA to form loops is affected by the sequence and the distance between binding sites [33,34], by DNA supercoiling [35,36], and by the presence of other proteins that may exert an influence on the conformation of the DNA [37]. DNA looping is a ubiquitous mechanism and has different functions in different cellular contexts [38–42]. In some cases, as in the prokaryotic operons *lac*, *ara*, and *gal*, DNA looping is involved in the regulation of transcriptional initiation. The function of the DNA loop is not, as initially thought, to mechanically prevent RNA Pol II to engage to the promoter, but instead to increase the effective local concentration of repressor in the proximity of the primary operator that, in turn, increases repression efficiency [43]. Still, the *lac* operon regulation relies in the association and dissociation kinetics of LacI to the DNA operator and, in particular, the association rate will depend on the target localization dynamics.

In the case of *lac* repressor in *E. coli*, the DNA target sequence is 24 bp long [31] while the entire genome is about 4.6 million bp long [44]. Furthermore, the *lac* repressor is auto regulated and, in basal conditions, it is present at a very low copy number: 5 to 10 tetramers per cell [45]. The first attempt to measure *lac* repressor-DNA association kinetics was reported at the beginning of the seventies, when Riggs et al. used a sensitive membrane filtration technique [46] to detect complexes between unlabeled LacI and labeled DNA [47]. In the case of 50 kbp long DNA fragments containing the *lac* primary (and now we know also a secondary) operator, they measured a rather fast binding reaction, following second-order kinetics with a rate constant for protein association to DNA (k_a) of the order of $7\times10^9\,M^{-1}\cdot s^{-1}$ [47]. They then sought to compare the measured value with a numerical estimation for k_a calculated, according to a model based on pure 3D thermal diffusion and random collisions, with the Smoluchowski's equation [48,49]. Strikingly, they found that the association rate constant they measured was almost 100 times larger than the value expected for a reactant of the size of *lac* repressor [47]. This limiting value is often referred to as the diffusion limit [50] and is generally considered to be of the order of $10^8\,M^{-1}\cdot s^{-1}$ [51]. Hence, Riggs et al. put forward for the first time the idea that the "[*lac*] repressor is not simply diffusing randomly but rather is oriented by relatively long-range electrostatic forces toward the DNA. It is therefore worth considering an extreme model of [LacI] oriented diffusion […] by binding and "rolling" or "hopping" along it [DNA]" [47]. Despite awareness of the relevance of their visionary idea, they discarded it because, as they noticed, "several arguments can be brought against this interesting idea". Among those arguments, they listed the very weak nonspecific affinity of *lac* repressor for non-operator DNA (characterized by a dissociation constant of the order of 10^{-3} M [52]) and the fact that the acceleration in target localization they observed could be explained as well as by electrostatic attraction. Also, such a model of "rolling" along the DNA would predict that the fragmentation of DNA in small pieces should reduce the association rate, a fact that they did not observe in their experiments [53].

Recently, it has been shown that the k_a measured by Riggs et al. did not exceed the diffusion limit, for the salt conditions under which the initial experiments were conducted, and that "rolling" events enhance target localization only when they are extremely short and not exceed 80 bp [54]. Nevertheless, Riggs and coworkers raised very relevant questions and triggered the interest of many researchers. In fact, since the first evidences of protein binding to specific DNA sequences [28,55] and the report of the fast (or, better said, not as slow as expected) LacI association kinetics [47], many efforts have been devoted to try to quantitatively understand the way specific transcription factors find their DNA target sequences.

3. Facilitated diffusion and beyond

For more than thirty years, most of the research devoted to elucidate TFs target search mechanisms has been essentially based on biochemical and theoretical grounds. As we will discuss in the following