



HHS Public Access

Author manuscript

Trends Biochem Sci. Author manuscript; available in PMC 2016 March 01.

Published in final edited form as:

Trends Biochem Sci. 2015 March ; 40(3): 127–129. doi:10.1016/j.tibs.2015.01.005.

Evolution and Diversification of the Basal Transcription Machinery

Sascha H.C. Duttke¹

¹Section of Molecular Biology, University of California, San Diego, La Jolla, CA 92093, USA

Abstract

Transcription initiation was once thought to be regulated primarily by sequence-specific transcription factors with the basal transcription machinery being largely invariant. Gradually it became apparent that the basal transcription machinery greatly diversified during evolution and new studies now demonstrate that diversification of the TATA-binding protein (TBP) family yielded specialized and largely independent transcription systems.

Keywords

Basal transcription machinery; diversification; evolution; gene regulation; system factors; TATA-binding protein (TBP); TBP related factor (TRF); transcription

Main Text

Neither gene number nor genome size correlates with organismic complexity. Rather, it is the ability to express genes at particular times in distinct tissues that enables differentiation and adaptation. The diversity in gene regulatory processes is reflected in the manifold life forms we marvel at today.

Transcription is the initial step in gene expression and to date, most well-known examples for the evolution of transcriptional regulation are provided by sequence-specific transcription factors such as Hox proteins. The basal transcription machinery itself has long been regarded as universally conserved with a passive role in gene regulation. However, gradually it became apparent that the basal transcription machinery diversified during evolution.

The size and subunit composition of the basal transcription machinery drastically increased during the course of evolution, containing about 6 subunits in Bacteria, up to 15 in Archaea [1], and substantially more in Eukarya, which encode at least three different RNA polymerases (RNAPs, reviewed in [1]). All eukaryotic polymerases emerged from a

© 2015 Elsevier Ltd. All rights reserved.

Correspondence: sduttke@ucsd.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

common ancestor, a notion reflected in their structural and functional conservation (reviewed in [2], [3]). The RNAP II transcription apparatus is the likely most closely related to the one found in Bacteria and Archaea. It contains approximately 45 subunits in yeast and humans (Figure 1A), not including transcriptional co-activators like the mediator. As all signals regulating the initiation of gene expression are ultimately integrated by the basal transcription machinery, it might be imagined that its alteration could provide an important means for regulatory diversification.

The bacterial basal transcription machinery consists of the five RNAP core subunits and one of several promoter specificity σ -factors. *Escherichia coli* has six alternative σ -factors, *Bacillus subtilis* 18, and *Streptomyces coelicolor* an astonishing 63. The expansion of σ -factors in *S.coelicolor*, an unusual soil bacterium with a broad range of metabolic processes and ‘multicellular development with distinct tissues’ [4], provides a bacterial example for how more intricate gene regulation enabled adaptation and a relatively complex life style.

In Archaea and Eukarya, promoter specificity is mediated by the basal transcription factors. The basal eukaryotic RNAP II factors include TF-IIA, -IIB, -IID, -IIE, -IIF and IIH, with TFIID serving as the anchor [5]. The canonical TFIID complex is composed of the TATA-binding protein (TBP) as well as 8-14 TBP-associated factors (TAFs) [6]. TFIID binds specific core promoter elements and chromatin modifications. Most notably, TAF1/2 interact with the Initiator motif, TBP with the TATA box and TAF6/9 with the downstream promoter element (DPE) [5]. In Archaea, recruitment of RNAP is facilitated by TFB, TBP and often TFE [1]. Systematic analyses of archaeal genomes indicate the absence or orthologs to TFIIA, TFIIF, and TFIH. Comparison of the transcription machineries among the three domains of life - Bacteria, Archaea and Eukarya, also reveals a diversification in basal transcription factors (Figure 1A). Within the Eukarya, the transcription complexes may appear to be similar. Nevertheless, as exemplified by the machineries of the unicellular eukaryote yeast and mammals, there are at least two notable differences. First, the number and subunit composition of transcriptional co-activators increased during metazoan evolution. Second, TBP-related factors (TRFs) and specialized TBP-associated factors (TAFs) emerged, and with them, non-canonical TFIID complexes.

The TATA-binding protein (TBP) is an integral part of all eukaryotic and archaeal transcription machineries [1,5,7]. Duplications of the TBP gene have occurred several times during evolution yielding paralogs, so far only known in Eukarya [6,7]. Three recent studies now report distinct functions for the TBP family member TRF2 in mediating transcription initiation [7-9], and therefore the existence of at least two largely independent RNAP II transcription systems. The observation that TBP family proteins support distinct and largely independent transcription programs led to the proposal that they be referred to as ‘system factors’ [7]. Intriguingly, the concept can further be expanded to other RNAPs and TBP family members.

The first TBP-related factor discovered, TRF1, is found in some insects and mediates RNAP III-dependent transcription of the tRNA genes. Transcription from most other RNAP III genes, such as the U6 spliceosomal RNA, is TBP-dependent [10]. In addition, TRF1 has been shown to mediate RNAP II transcription of the *Drosophila tudor* gene, but its general

role in RNAP II transcription remains largely speculative. TRF2 emerged immediately prior to the evolution of the bilateria and is crucial for development, cell cycle progression, and cell differentiation in worms, flies, frogs and fish. Mice lacking TRF2 display normal development but are infertile due to defects in spermatogenesis (summarized in [6,7]). TRF2 does not bind the TATA box, likely due to the loss of three highly conserved phenylalanines that are essential for kinking the DNA and ultimately anchoring TBP [7]. Indeed, TRF2 appears to be the only TBP-paralog that depends on its associated factors for localization. Such factors likely include coactivators and basal factors like TFIIA and TFIIB or TFIID (particularly TAF6/9, TAF7L). TRF2 is required for the initiation of transcription from TCT core promoter motif-dependent genes [8] and is a preferential regulator of DPE-containing core promoters in *Drosophila* [9]. It is further required for the initiation of transcription from certain core promoters lacking currently known core promoter motifs like the *Drosophila* histone H1 gene or the terminal deoxynucleotidyl transferase (TdT) gene in chicken, suggesting the existence of other, currently unknown TRF2-dependent transcription systems. The third TRF, TRF3, is conserved among vertebrates but does not appear to occur in jawless fish (Gnathostomata, Duttke SHC, Doolittle RF and Kadonaga JT *unpublished data*), which may suggest that it fulfils a special role in jawed vertebrates. TRF3 binds the TATA box and is highly homologous to TBP but its functional separation from TBP is yet incompletely understood. However, it appears that TRF3 plays an important role in mediating the expression of vertebrate-specific embryonic genes, particularly in ovaries [6]. It remains to be shown whether TRF2/3 are also involved in mediating RNAP III-dependent transcription.

In this sense, RNAPs with different TBP-family proteins present different transcription systems (Figure 1B), similar to σ -factors in Bacteria. This subdivision of transcription systems increases the regulatory potential and may have allowed proliferating and differentiated cell types to sustain distinct transcriptional programs, thereby reducing gene regulatory constraints during development.

In conclusion, many past studies have been performed based on the assumption of a largely invariant basal transcription machinery, a notion that is challenged by the previous discovery of alternative basal transcription factors [6] and now through the identification of multiple distinct basal transcription complexes [7]. The identification of distinct basal transcription machineries complements previous studies (reviewed in [5]) describing enhancers specifically promoting transcription for one core promoter type (e.g. TRF2-dependent transcription) but not another (e.g. TBP-dependent transcription). The evolution of diverse, and, at times, specialized transcription machineries argues against the basal transcription machinery being a homogeneous on/off switch, but rather a highly diverse relay station for different regulatory programs, enabling an additional level in gene regulation.

Acknowledgments

I thank Russell Doolittle, Tamar Juven-Gershon, George Kassavetis and Minsung Kim for discussion and James Kadonaga for inspiration and incredible mentorship. I apologize for the limited citation of references due to format restrictions. S.H.C.D. is the recipient of the University of California at San Diego Molecular Biology/Cancer Center Fellowship. This work was supported by National Institutes of Health grant R01 GM041249 to James T. Kadonaga.

References

1. Werner F, Grohmann D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nature Reviews Microbiology*. 2011; 9:85–98.
2. Vannini A, Cramer P. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Molecular Cell*. 2012; 45:439–46. [PubMed: 22365827]
3. Duttke S. RNA polymerase III accurately initiates transcription from RNA polymerase II promoters in vitro. *Journal of Biological Chemistry*. 2014; 289:20396–404. [PubMed: 24917680]
4. Bentley S, et al. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature*. 2002; 417:141–147. [PubMed: 12000953]
5. Smale S, Kadonaga J. The RNA polymerase II core promoter. *Annu Rev Biochem*. 2003; 72:449–479. [PubMed: 12651739]
6. Goodrich J, Tjian R. Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. *Nature Reviews Genetics*. 2010; 11:549–558.
7. Duttke S, et al. TRF2 and the evolution of the bilateria. *Genes & Development*. 2014; 28:2071–2076. [PubMed: 25274724]
8. Wang YL, et al. TRF2, but not TBP, mediates the transcription of ribosomal protein genes. *Genes & Development*. 2014; 28:1550–1555. [PubMed: 24958592]
9. Kedmi A, et al. Drosophila TRF2 is a preferential core promoter regulator. *Genes & Development*. 2014; 28:2163–2174. [PubMed: 25223897]
10. Verma N, et al. Differential Utilization of TATA Box-binding Protein (TBP) and TBP-related Factor 1 (TRF1) at Different Classes of RNA Polymerase III Promoters. *Journal of Biological Chemistry*. 2013; 288:27564–27570. [PubMed: 23955442]

Abbreviations

RNAP	RNA polymerase
TBP	TATA-binding protein
TF	transcription factor
TRF	TBP-related factor

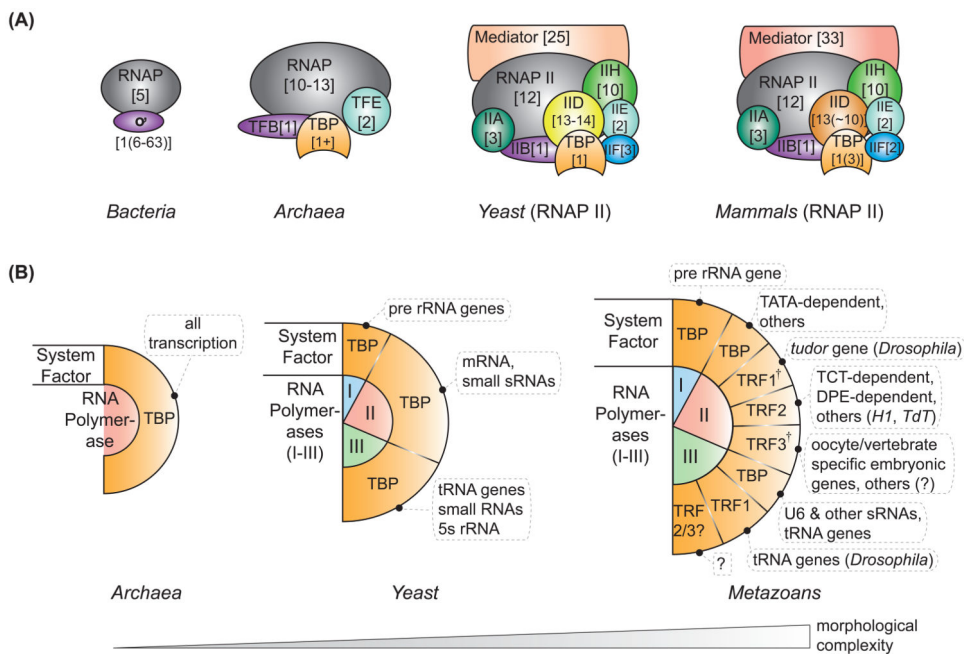


Figure 1. Diversification of the basal transcription machinery and the evolution of ‘System Factors’. **(A)** Subunit composition of the basal transcription machineries in the three domains of life. The number and composition of basal transcription factors increased during evolution. Subunit numbers are given in “[]” and the number of genomically encoded alternative factors is in “()”. **(B)** TATA-binding protein (TBP) family proteins guide different RNA polymerase transcription programs (dashed boxes) and thereby further subgroup the polymerases in distinct and largely (but not entirely) independent transcription systems [7]. † Note: TRF1 is specific to some insects while TRF3 emerged in jawed animals.