CellPress

A unified architecture of transcriptional regulatory elements

Robin Andersson[1*], Albin Sandelin[1,2], and Charles G. Danko[3,4,†]

[1] The Bioinformatics Centre, Department of Biology, University of Copenhagen, Copenhagen, Denmark
[2] Biotech Research and Innovation Centre (BRIC), University of Copenhagen, Copenhagen, Denmark
[3] Baker Institute for Animal Health, College of Veterinary Medicine, Cornell University, Ithaca, NY, USA
[4] Department of Biomedical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY, USA

Gene expression is precisely controlled in time and space through the integration of signals that act at gene promoters and gene-distal enhancers. Classically, promoters and enhancers are considered separate classes of regulatory elements, often distinguished by histone modifications. However, recent studies have revealed broad similarities between enhancers and promoters, blurring the distinction: active enhancers often initiate transcription, and some gene promoters have the potential to enhance transcriptional output of other promoters. Here, we propose a model in which promoters and enhancers are considered a single class of functional element, with a unified architecture for transcription initiation. The context of interacting regulatory elements and the surrounding sequences determine local transcriptional output as well as the enhancer and promoter activities of individual elements.

The classical view: enhancers and promoters as distinct regulatory elements

Transcriptional regulatory DNA sequences encode intricate programs of gene expression that control almost all cellular functions. These regulatory sequences recruit transcription factors (TFs) in a DNA sequence-dependent fashion, allowing cells to precisely control the rates of chromatin decompaction, transcription initiation, and the release of RNA polymerase II (RNAPII) into productive elongation [1,2]. Classically, a sharp distinction has been drawn between two types of regulatory sequence: promoters and enhancers.

Promoters are DNA sequences that regulate and initiate RNAPII transcription at proximal transcription start sites (TSS). RNAPII core promoters can encompass as little as 100 base pairs (bp) of DNA surrounding the TSS, which often contains one or more degenerate copies of the core DNA sequence elements, including the TATA box and the initiator (INR). These core promoter elements are recognized by general transcription factors (GTFs), such as TFIID and TFIIB, which are responsible for recruiting and assembling the RNAPII pre-initiation complex (PIC) [3]. RNAPII PIC assembly and transcription initiation are further facilitated by TFs bound proximally to core promoters. Although promoters were originally found at TSSs of known genes, more-direct experimental methods, such as sequencing 5′ ends of RNAs, have identified promoters genome wide and showed that a much larger proportion of mammalian genomes is associated with transcription initiation than can be accounted for by annotated gene models [4,5]. Promoters, which we use here to refer to units of proximal and core promoter regions collectively, work together with other regulatory regions, such as enhancers and silencers, to regulate all stages of RNAPII transcription from RNAPII recruitment to transcriptional elongation.

Enhancers, in contrast to promoters, are generally considered TF-binding regulatory regions distal to gene TSSs. The first enhancer discovered was a 72-bp tandem repeat upstream of early genes in the simian virus 40 (SV40) genome [6,7]. This sequence was reported to increase transcription of the β-globin gene more than 200-fold when inserted into the same recombinant expression vector, irrespective of its position, distance, and orientation relative to the target gene promoter [8]. Importantly, transcription of the β-globin gene invariably initiated at the β-globin promoter, indicating that the enhancer worked to stimulate transcription from the target promoter at a distance. Enhancers were shortly thereafter discovered in the mouse genome [9] and are today considered key players of transcriptional regulation across Eukaryota. Although different models have been proposed to explain how enhancers regulate gene expression over long genomic distances (Box 1), several recent studies suggest that chromatin architecture places enhancers in close 3D proximity with target gene promoters [10]. Based on these observations, enhancers are classically defined by their ability to increase transcriptional output from target genes.

However, recent studies are challenging the conventional wisdom that enhancers and promoters are distinct entities. These results are changing the way that regulatory elements are defined and identified. Here, we review recent progress in the area, focusing on new genome-wide studies reporting broad similarities in the biochemical and DNA sequence properties of mammalian enhancers and

Corresponding authors: Andersson, R. (robin@binf.ku.dk); Danko, C.G. (dankoc@gmail.com).
Keywords: transcriptional regulation; enhancer; promoter; core promoter; eRNA.
*Twitter: @robin_andersson
†Twitter: @charlesdanko

0168-9525/
 http://dx.doi.org/10.1016/j.tig.2015.05.007

Box 1. Models for enhancer function

Several models have been proposed for how enhancers regulate expression at gene promoters over long genomic distances, in some cases up to a megabase in units of linear DNA. A popular hypothesis is that enhancers regulate transcription by looping into close 3D proximity with target gene promoters [66]. This model is supported by several independent experimental observations, such as the frequencies of distal DNA sequence ligation using chromatin conformation capture (3C), and by florescence *in situ* hybridization (FISH) [67]. However, although 3C and FISH both indicate physical interactions between enhancers and promoters, specific interactions are not always reproducible between experimental techniques [68], suggesting that technical work still needs to be done before we are able to reproducibly capture interactions between enhancers and promoters. Other models have also been proposed that explain how enhancers regulate gene expression, including promoter tracking [69] and enhancer–promoter linking via protein bridges [70], and these are generally not mutually exclusive with DNA looping. In addition to the constraints on transcriptional regulation posed by larger chromatin architectures, context-dependent properties influence whether physical proximity of regulatory elements lead to increased transcriptional output, including enhancer specificity for certain core promoter elements [44,71], promoter competition [72], and insulation [73].

Little is known about the molecular mechanisms by which enhancers regulate the transcription level at target gene promoters. Several mechanisms have been proposed, including: (i) facilitating or enabling transcription initiation by supplying needed factors such as the Mediator [74], GTFs [27], or RNAPII [23]; (ii) affecting the release rate of RNAPII into productive elongation by activation of positive transcription elongation factor b (P-TEFb) [75]; or (iii) by recruitment of the super elongation complex [76] (for a recent review of enhancer function, see [2]).

gene promoters. Based on these new findings, we argue here that the discriminatory view of enhancers and promoters as distinct regulatory elements is problematic because it suggests that they have distinct functions. Rather, their broad similarities and overlapping functional properties suggest a unified view of enhancers and promoters as a single class of 'regulatory element', each with an intrinsic ability to drive local transcription (i.e., act as a 'promoter') or enhance distal transcription (i.e., act as an 'enhancer') with varying strengths, and that their primary function is likely context dependent.

The classical approach: classification of regulatory elements by histone modifications

Several studies have revealed post-translational modifications of chromatin that are characteristic of regulatory sequences across the genome (Box 2) [11]. Certain marks have genomic distributions that correlate broadly with the expected location of either enhancers or gene promoters. Most notably, a seminal study found that histone H3 lysine 4 trimethylation (H3K4me3) is highly enriched in regions that are proximal to the 5′ end of gene annotations, where active gene promoters are expected to reside, whereas H3K4me1 is more frequently found distal to gene promoters [12]. This led to the suggestion that H3K4me3 and H3K4me1 could discriminate between promoters and enhancers. Given that these modifications are not mutually exclusive, their signal ratio (H3K4me1:H3K4me3), which was found to be low at promoters and high at enhancers, was suggested to aid in discrimination [13]. In addition, histone H3 lysine 27 acetylation (H3K27ac) has been proposed to distinguish active enhancers and promoters from inactive ones [14,15]. Over the past few years, classification of regulatory sequences based on histone modifications has become widely adopted by the genomics and gene regulation communities.

Active enhancers can independently work as promoters

Recent observations point to numerous functional similarities between enhancers and gene promoters. RNAPII binds and initiates transcription at active enhancers [16,17], demonstrating that at least some enhancers direct the biological process that is a defining feature of gene promoters. More recently, it has become clear that tens of thousands of enhancers initiate RNAPII in opposing orientations from local TSSs and transcribe enhancer-templated noncoding RNAs (eRNAs), albeit in many cases with a lower rate of transcription initiation than occurs at gene promoters [18–22] (Figure 1A). Although RNAPII could be partially supplied to enhancers by gene promoter regions that are physically proximal, at least some enhancers are able to initiate transcription independently of gene promoters. For example, GTF and RNAPII recruitment to enhancers upstream of the α-globin gene precedes promoter recruitment and is unaffected in MEL cell line hybrids lacking the α-globin gene promoter [23]. Another example of the dual role of enhancers with local promoter activities are a group of α-globin enhancers contained within nitrogen permease regulator-like 3 (*Nprl3*) gene, which act as both enhancers to the α-globin gene promoter and alternative promoters to noncoding RNAs sharing exons with *Nprl3* [24]. Similar to active gene promoters, active enhancers are demarcated by a well-positioned array of nucleosomes surrounding a nucleosome-depleted region (NDR) [21,25,26], and the distance between divergent TSSs and NDR edges is constrained [19,21,25]. Consistent with their ability to recruit RNAPII, enhancer NDRs are enriched in typical core promoter sites, including TATA and INR motifs, and bind GTFs [19,22,27,28] (Figure 1A). In general, enhancers are depleted of CpG islands and seem to recruit similar repertoires of lineage-specific master regulator TFs as CpG-poor gene promoters [21]. Taken together, these observations suggest that enhancers work independently as promoters and are similar to gene promoters in terms of DNA sequence, nucleosome positioning, and TF binding.

Although many enhancers are transcribed, they represent a subset of enhancers predicted from histone modifications. However, untranscribed enhancers in one cell type are often transcribed in another [21,25], suggesting that many enhancers have promoter activities in an appropriate context. More generally, this observation raises the question of what makes transcribed enhancers different from untranscribed ones. Although this question is unresolved, a few key observations have been made. Mammalian enhancers producing eRNAs are more likely to interact with gene promoters [29] and be validated by *in vitro* reporter gene assays [21,30] compared with untranscribed enhancers identified using histone marks. Furthermore, using a transgenic mouse enhancer assay, tissue specificity of eRNA expression was shown to correctly predict tissue-specific *in vivo* enhancer activity of