SUBMITTED VERSION

Nan Hao and Keith E. Shearwin
Hidden secrets of sigma54 promoters revealed

http://hdl.handle.net/2440/124218

Hidden Secrets of Sigma54 Promoters Revealed

Nan Hao[1] and Keith E. Shearwin[1,*]

[1]Department of Molecular and Cellular Biology, School of Biological Sciences, University of Adelaide, Adelaide, SA 5005, Australia.

*Correspondence: keith.shearwin@adelaide.edu.au (K.E. Shearwin)

Keywords

Abstract

Bacterial σ^{54} promoters are the DNA binding motif for σ^{54} containing RNA polymerase holoenzymes. A recent study using a combination of synthetic oligonucleotide library screening, biochemical characterization and bioinformatics has uncovered a new and unexpected role for σ^{54} promoters, encoding a form of bacterial "insulator sequence" to dampen unwanted translation.

To constantly monitor and swiftly respond to even subtle fluctuations in their environment, bacteria require orchestrated expression of a large number of genes. One strategy that bacteria utilize to coordinate such rapid and finely tuned responses is to use alternative sigma factors. For example, DNA bound AAA+ transcriptional activators, known as bacterial enhancer binding proteins (bEBPs), bind to an upstream activating sequence (UAS) to activate σ^{54} containing RNA polymerase (RNAP) and switch on σ^{54} controlled genes [1, 2]. With the exception of the σ^{54} protein, all other *E. coli* sigma factors, including the house-keeping σ^{70} and the specialized σ^{19}, σ^{24}, σ^{28}, σ^{32} and σ^{38} factors, also have their own dedicated set of anti-sigma factors [3]. These anti-sigma factor proteins operate by physically sequestering the sigma factors, providing a layer of post-transcriptional regulation. Until now though, expression of σ^{54} regulated genes was thought to be controlled only at the transcriptional level. Is such a single control mechanism likely to be sufficient to precisely regulate the expression of these important genes? Amit *et al.* [4] provide some clues to this question, finding within σ^{54} promoters an unexpected sequence element which helps prevent accidental expression of σ^{54} controlled genes arising from transcriptional read-through.

Prior to the Amit study [4], σ^{54} promoters were known to act as master switches to turn on genes involved in nitrogen assimilation, motility, host colonization, and biofilm formation [1]. Amit *et al.* initially set out to study the interplay between different bacterial UASs and σ^{54} promoters but noticed that of 66 enhancer-promoter combinations tested, one of the promoters, namely the *E. coli glnK* promoter (*glnKp*), exerted a strong inhibitory effect against downstream gene expression. While it may have been tempting to write off the result for this single promoter as an outlier, Amit *et al.* pursued this curious finding and uncovered a hidden aspect of σ^{54} promoters that was quite unexpected.

The approach they took was to combine massively parallel oligonucleotide library synthesis [5-9] with fluorescence activated cell sorting and deep sequencing (*aka* Flow-seq). (Figure 1). Such oligonucleotide library based screening methods, based on the rapid and inexpensive synthesis of a very large number of variant sequences, permit function-based separation of the library into variant sub-populations and their

analysis by next generation DNA sequencing. Flow-seq uses flow cytometric sorting, based on expression levels of a fluorescent reporter gene, to select subsets of library variants. Subsequent sequencing of these pools and bioinformatic analysis of the results allows dissection of the DNA sequence determinants critical for the selected activity. Even with such a high throughput method, a complete systematic survey of the sequence space of a 50bp promoter (4^{50}) is still beyond the reach of current technologies. As a result, Amit et al [4] custom designed an oligonucleotide library that consisted of over 12 000 variants including i) annotated σ^{54} promoters; ii) σ^{54} core-like sequences from *E. coli* and *V. cholera* (but distinct from the annotated σ^{54} promoters); iii) non-σ^{54} core sequences; and iv) a set of annotated *E. coli* σ^{70} promoters. Subsets of variants exhibiting silencing, as measured by reduced expression of a downstream mCherry reporter (Figure 1), were selected by flow cytometry and sequenced.

Intriguingly, it was found that the silencing effect is manifested not by a single recognizable position weight matrix, but by the presence of loosely defined, short CT-rich motifs of 3-5 bp (CTmers), where the strength of silencing was dependent on the number and the position of the CTmers. Mechanistically, their findings suggest that a subset of σ^{54} controlled genes are indeed controlled at the post-transcriptional level, by utilizing part of the σ^{54} promoter sequence as an anti-Shine-Dalgarno (aSD) element (Figure 1B). In the bacterial genomic context, such messenger RNAs might arise by transcriptional read-through from an upstream promoter. The read-through transcript would have σ^{54} derived CU enriched sequences, positioned ideally to basepair with the downstream AG-rich Shine-Dalgarno (SD) sequence. This aSD:SD interaction inhibits translation, by blocking access of ribosomes to the ribosome binding site (RBS) in the mRNA [10]. It will be interesting to see whether similar aSD elements are identified within other stress response promoters, such as the σ^{19}, σ^{24}, σ^{28}, σ^{32} and σ^{38} dependent promoters.

It is sobering to note that this previously unrecognised aspect of σ^{54} promoter biology is unlikely to have been revealed by traditional low throughput promoter analysis experiments. As we enter the synthetic biology era of custom designed genetic circuits, with the goal of plug-and-play DNA elements, such hidden functions could

be considered to be 'bugs' in the DNA software. High throughput oligonucleotide based screening approaches represent one way to identify, and perhaps even take advantage of, such features. One suspects that many more hidden functions remain to be uncovered, particularly in the larger, less explored eukaryotic genomes.

References

1. Francke, C. et al. (2011) Comparative analyses imply that the enigmatic Sigma factor 54 is a central controller of the bacterial exterior. BMC Genomics 12, 385.
2. Zhang, N. et al. (2016) The bacterial enhancer-dependent RNA polymerase. Biochem J 473 (21), 3741-3753.
3. Trevino-Quintanilla, L.G. et al. (2013) Anti-Sigma Factors in E. coli: Common Regulatory Mechanisms Controlling Sigma Factors Availability. Curr Genomics 14 (6), 378-87.
4. Amit, R. et al. (2017) A synthetic oligo library and sequencing approach reveals an insulation mechanism encoded within bacterial σ54 promoters. Cell Rep XX (X), XX-XX.
5. Goodman, D.B. et al. (2013) Causes and effects of N-terminal codon bias in bacterial genes. Science 342 (6157), 475-9.
6. Mutalik, V.K. et al. (2013) Precise and reliable gene expression via standard transcription and translation initiation elements. Nat Methods 10 (4), 354-60.
7. Kosuri, S. and Church, G.M. (2014) Large-scale de novo DNA synthesis: technologies and applications. Nat Methods 11 (5), 499-507.
8. Sharon, E. et al. (2012) Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. Nat Biotechnol 30 (6), 521-30.
9. Rocklin, G.J. et al. (2017) Global analysis of protein folding using massively parallel design, synthesis, and testing. Science 357 (6347), 168-175.
10. Espah Borujeni, A. et al. (2017) Precise quantification of translation inhibition by mRNA structures that overlap with the ribosomal footprint in N-terminal coding sequences. Nucleic Acids Res 45 (9), 5437-5448.

Figure legend

Figure 1. (A) An oligonucleotide library-based high throughput screen was developed to decipher the underlying mechanism of insulation encoded within a subset of σ^{54} promoters. Insulation was measured as reduced expression of a downstream mCherry protein, expressed by transcriptional read-through from an upstream promoter. Note that the σ^{54} promoter itself is not active in the absence of an upstream activating sequence (UAS) or enhancer binding protein (bEBP). The Flow-seq workflow to study the sequence determinants responsible for insulation is indicated. The insulating effect is conferred by the presence of 3-5bp CT rich motifs (CTmers) within the σ^{54} promoter, while the strength of silencing is dependent on the

density and the position of the CTmers. (B) CTmers are proposed to reduce translation by acting as anti-Shine-Dalgarno sequences (aSD), through a base-pairing interaction with the Shine-Dalgarno (SD) sequence, thus blocking the access of ribosomes to the ribosome binding site.