RegulonDB Glossary

Transcriptional Regulation

Definition

1. Transcriptional regulation refers to the mechanisms by which gene expression is controlled at the level of transcription. It involves the regulation of RNA polymerase's ability to transcribe genes into RNA molecules.

Acronym and synonyms

Description

Overview of transcriptional regulation

The transcription process begins with the formation of the RNA polymerase (RNAP) holoenzyme (Eσ), which is a complex composed of the core RNAP and a σ factor. This holoenzyme is responsible for initiating gene transcription at specific DNA positions. Bacterial RNA polymerase (RNAP) is a multi-subunit enzyme, consisting of five subunits: α2, β, β', and ω. The core RNAP contains the active site that catalyzes the formation of the phosphodiester bond of nascent RNA.

The α subunits of RNAP interact with the upstream **promoter** (UP) element, which consists of two distinct subsites located upstream of the –35 element. The RNAP core enzyme interacts specifically with the template-strand positions -4 to +2, forming the core recognition element (CRE).
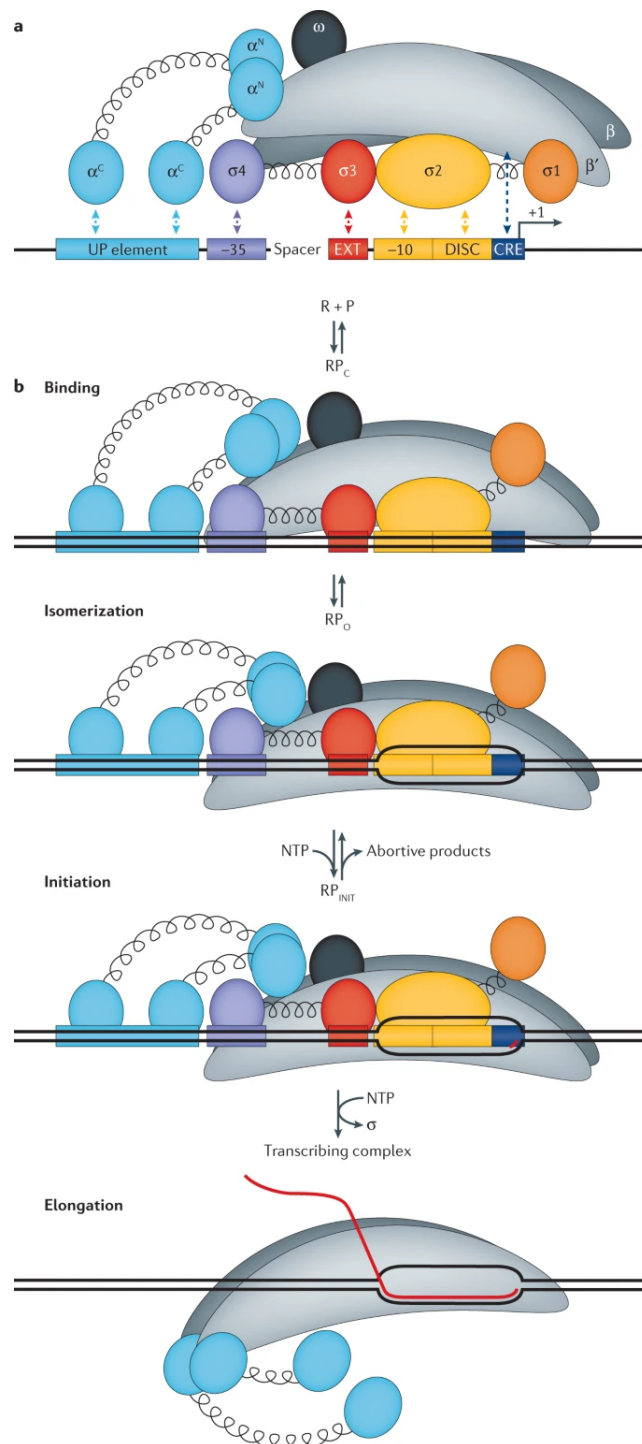
Figure. The regulation of bacterial transcription initiation. Source: https://doi.org/10.1038/s41576-020-0254-8

To form the holoenzyme Eσ, RNAP is associated with a **σ factor**. There are two structurally and evolutionary distinct families of σ factors: σ54 and σ70. σ70-related factors contain up to four functional domains (σ1–4). The σ2 domain recognizes and interacts with the −10 element, and the σ4 domain interacts with the −35 element. The extended −10 element interacts with σ3, and this interaction is crucial in promoters whose −35 and −10 elements show a poor match to

consensus sequences.

Some promoters, particularly the ones that respond to amino acid starvation, have an element called discriminator, which is recognized by and interacts with the σ2 domain (conserved region σ1.2). σ70 bound to the nontemplate strand captures the –10 region in an open complex and allows the single-strand template DNA to enter the active site. For this, σ70 does not need an energy source such as ATP or GTP15.

σ interacts with different promoter elements to position the Eσ to unwind the double-stranded DNA in the region of the transcription start site (TSS), which corresponds to the first base of the transcript6. Most bacteria rely on different σ factors to take Eσ to different sets of promoters in response to changes in growth conditions. Alternative σ factors are classified into two evolutionary distinct families: σ54 and σ70. σ54 is a single member family, whereas σ70 normally has several members, whose number varies among bacterial species. For example, σ24, σ28, σ32, σ38, and σ70 are members of the σ70 family in E. coli.

Initially, Eσ binds to promoters in a closed complex (RPc), covering a DNA region from approximately -55 bp to approximately +15 bp relative to the transcription start site (TSS), as determined by DNA footprints18. This binding triggers conformational changes in both the DNA and Eσ, leading to the formation of an open complex (RPo). In the open complex, the DNA strands are separated from positions -11 to +3 bp, creating a region known as the 'transcription bubble'. Within the transcription bubble, the base on the template strand marked as +1 serves as the template for the first nucleotide of the transcript. Transcription begins with a short, unstable region that may undergo abortive transcription, where Eσ synthesizes short RNA products before transitioning to a stable elongation complex. The transcription cycle then proceeds with elongation and termination steps, which have been extensively reviewed elsewhere.

The primary role of promoter elements is to interact with Eσ and facilitate the formation of a DNA–Eσ complex that is capable of initiating transcription. Therefore, promoter elements determine the intrinsic activity of a promoter. The activity of different promoters can vary by several orders of magnitude, ranging from those that produce only a few RNA copies per cell generation to those that generate tens of thousands of RNA copies.

Promoters are generally subject to regulation, either indirectly through changes in Eσ concentration or substrate concentrations, or directly through specific regulators. **Transcription factors (TFs)**, including **activators and repressors**, are a subclass of these regulators that act by binding to specific DNA targets. Activators compensate for the intrinsic weakness of promoters prone to activation by recruiting Eσ, while repressors prevent Eσ from transcribing by directly obstructing the promoter or inhibiting specific steps of the transcription process.

Transcription factors (TFs) are usually allosteric proteins that bind specifically to their operator DNA sites, which are usually located near promoters, either in the holo or apo conformation to regulate gene expression. We refer to a functional holo conformation when the TF binds to DNA as a complex bound to an effector that can be either a noncovalently bound small molecule, or after a covalent modification, such as phosphorylation by a two-component system; whereas a TF binds in an apo conformation when the protein binds alone. For instance, CRP binds to its specific binding sites once bound to cAMP, its allosteric small ligand; whereas the LacI repressor binds to DNA as a protein in apo conformation, and unbinds in the presence of allolactose, its allosteric modifier.
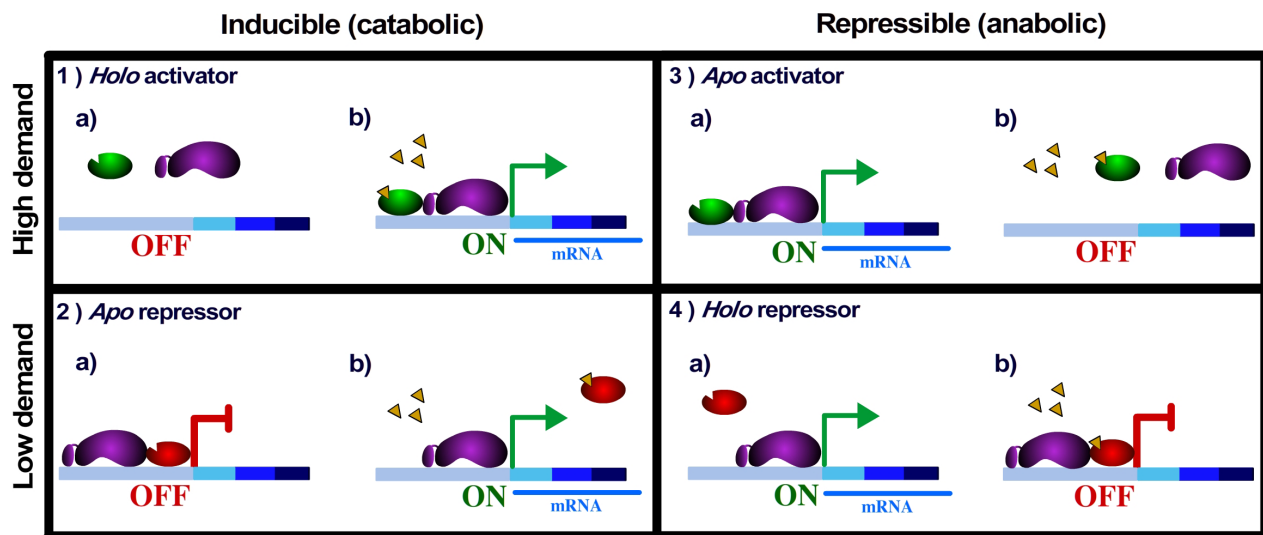

Figure. Holo and Apo conformations of Transcription Factors. Source: https://doi.org/10.1371/journal.pone.0065723

The activity of most DNA-binding transcription regulators is tightly linked to external signals through various mechanisms, allowing for rapid responses and making gene expression highly sensitive to environmental changes.

Elongation is the second stage of transcription in bacteria. Once the RNA polymerase has successfully bound to the promoter region and initiated transcription, it begins the elongation phase, during which it synthesizes the RNA molecule.

After initiation, RNA polymerase moves along the DNA template strand in the 3' to 5' direction, creating an RNA molecule in the 5' to 3' direction. It "reads" the DNA template strand and synthesizes the complementary RNA strand. As RNA polymerase moves along the DNA template, it adds complementary RNA nucleotides (adenine, cytosine, guanine, and uracil) to the growing RNA chain. RNA polymerase selects the appropriate ribonucleotide triphosphate (NTP) based on base pairing rules (A-U and C-G).

RNA polymerase has proofreading capabilities to correct errors in RNA synthesis. If an incorrect

nucleotide is added, RNA polymerase can remove it and replace it with the correct one. During elongation, a transcription bubble forms, where the DNA double helix temporarily unwinds, allowing RNA polymerase access to the template strand.

Termination is the final stage of transcription in bacteria, where the newly synthesized RNA molecule and RNA polymerase disengage from the DNA template. In bacteria, there are specific DNA sequences called terminator sequences that signal the end of transcription. There are two main types of terminators: intrinsic terminators and rho-dependent terminators.

Intrinsic Terminators have a distinct structure that includes a hairpin loop formed in the nascent RNA transcript. This hairpin loop causes RNA polymerase to pause and eventually dissociate from the DNA template. In rho-dependent terminators, a protein called rho (ρ) is involved. Rho binds to the nascent RNA and helps RNA polymerase to dissociate from the DNA template.

Once the terminator sequence is encountered, it triggers the release of the RNA transcript and RNA polymerase from the DNA template. This marks the end of the transcription process.
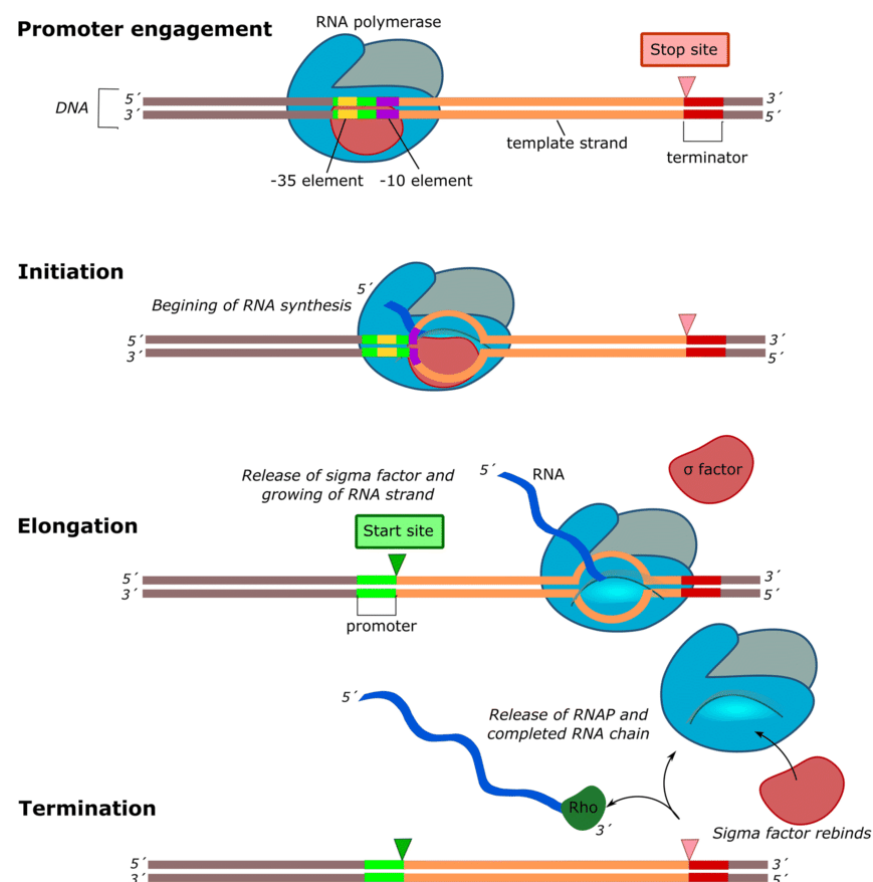

Figure. Bacterial transcription process. Source: DOI:10.1007/s00253-020-10577-0

Operons are a fundamental concept in the field of bacterial genetics, serving as a central mechanism for the regulation of gene expression in prokaryotic organisms like bacteria. These operons consist of a cluster of functionally related genes that are typically governed by a single

promoter and operator region. Their role is to facilitate the efficient coordination of multiple genes involved in a common biological pathway or process.

This concept has evolved over time, as discussed in the operon section, particularly in the context of RegulonDB. This evolution has been driven by the recognition of complex operons featuring multiple transcripts, often due to the presence of internal promoters and/or terminators."

Figure. deoCABD operon region, Genes:4 Promoters:4 Transcription Units:4

References

[1] Mejia-Almonte, C., Busby, S.J.W., Wade, J.T., van Helden, J., Arkin, A.P., Stormo, G.D., Eilbeck, K., Palsson, B.O., Galagan, J.E. and Collado-Vides, J. (2020) Redefining fundamental concepts of transcription initiation in bacteria. Nat Rev Genet, 21, 699-714. PMID: 32665585

[2] Gene Action – Operon Hypothesis. https://www.biologyonline.com/tutorials/gene-action-operon-hypothesis

Gene

n., [dʒiːn]

Definition

1. A region of DNA that encodes at least one functional product.
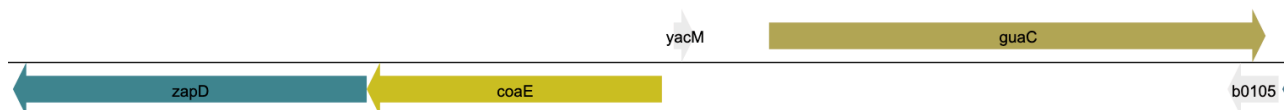2. "DNA region" defined as an extent of DNA greater than zero nucleotides.

Figure. Graphical representation of DNA region with genes with different size and orientations.

Acronym and synonyms

Properties

Property	Description		Example
position left	Absolute left position in the genome	required	350
position right	Absolute right position in the genome	required	1020
strand	DNA strand where the gene is coded	required	forward /reverse
dna sequence	dna gene sequence	required	ATGGTGTATG...TAA
name	gene name	optional	araC
synonyms	other names associated to the gene	optional	

Other optional properties for genes can be queried in the [database model](#) of RegulonDB.

Example

- Gene name: **apaG**
- Synonyms: EG10047
- Genome position: 51229 - 51606
- Size: 378 base pairs
- Strand: reverse
- DNA sequence :

```
>RDBECOLIGNC00043|apaG gene|product:DUF525 domain-containing protein ApaG|s
ATGATCAATTCGCCCCGAGTGTGTATTCAGGTTCAAAGCGTCTACATTGAGGCTCAATCT
TCACCTGATAATGAACGTTACGTTTTTGCTTATACCGTAACCATACGCAATCTGGGGCGA
GCGCCAGTGCAGTTGTTGGGGCGTTACTGGCTGATCACCAATGGCAATGGCCGTGAAACC
GAAGTCCAGGGCGAAGGAGTGGTTGGCGTCCAGCCACTTATCGCGCCTGGCGAAGAGTAC
CAGTACACCAGCGGTGCAATCATTGAAACCCCGCTGGGCACCATGCAGGGTCACTACGAA
ATGATCGATGAAAATGGCGTCCCTTTCAGCATCGACATTCCCGTATTCCGACTCGCCGTT
CCCACACTCATTCATTAA
```

Comments

The gene definition implies continuos DNA segment that carries the information for the structure

of at least one functional-RNA or polipeptide chain. If the gene carries the information for more than one product, it means that the gene is composed of overlapping CDSs or overlapping functional-RNA bearer regions.

The definition also is intended to include genes that produce more than one product, e. g. dnaX produces by frameshifting both subunit tau and subunit gamma of the DNA polymerase.

The definition is intended to regard independent cistrons that code for subunits of heteromultimeric proteins as different genes. For example, the cistrons that code for subunits of IHF, ihfA and ihfB, are differente genes.

In the database, genes of the pseudo and phantom types are included, indicating that they were previously considered functional genes but are now categorized as pseudo or phantom for various reasons.

Useful links

- Gene from [Biology Online](#).
- Gene from [Wikipedia](#)

References

[1] Mejia-Almonte, C., Busby, S.J.W., Wade, J.T., van Helden, J., Arkin, A.P., Stormo, G.D., Eilbeck, K., Palsson, B.O., Galagan, J.E. and Collado-Vides, J. (2020) Redefining fundamental concepts of transcription initiation in bacteria. Nat Rev Genet, 21, 699-714. [PMID: 32665585](#)

Product

Definition

1. A product is the RNA or protein produced based on the gene template.
2. "A gene product is the biochemical material, either RNA or protein, resulting from expression of a gene." [2]
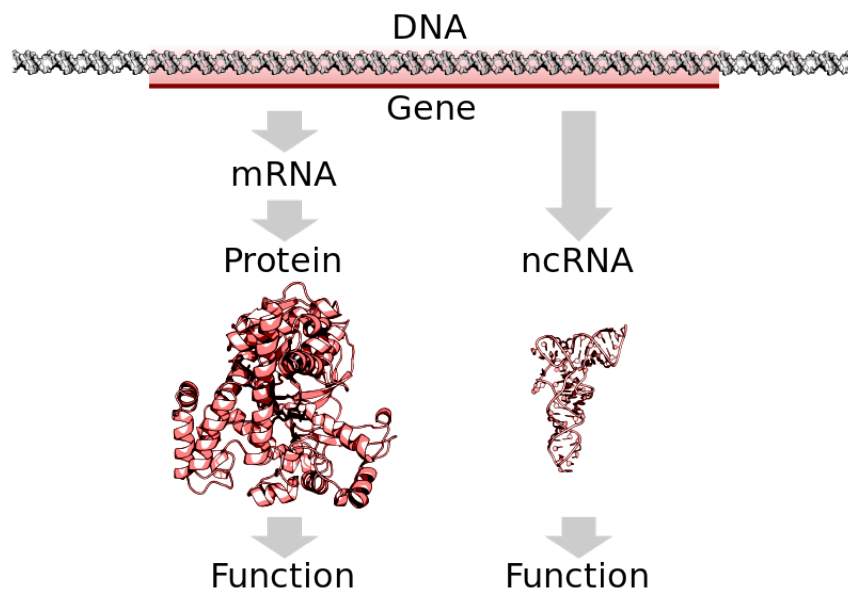
Source: Biology Online

Acronym and synonyms

Properties

Property	Description		Example
product name	Name of the product	optional	DUF525 domain-containing protein ApaG
gene id	Identifier of the gene that codes for the product	required	RDBECOLIGNC00043
sequence	Product sequence (RNA o AminoAcids)	required	
product type	Type of produch such as small RNA, tRNA, Transcription Factor..	optional	small RNA

Other optional properties for product can be queried in the database model of RegulonDB.

Example

- Product name: DUF525 domain-containing protein ApaG
- Synonyms: ApaG

- Molecular Weight: 13.867
- Cellular Locations: periplasmic space
- Sequence:

```
>DUF525 domain-containing protein ApaG|size: 125 aa
MINSPRVCIQVQSVYIEAQSSPDNERYVFAYTVTIRNLGRAPVQLLGRYWLITNGNGRET
EVQGEGVVGVQPLIAPGEEYQYTSGAIIETPLGTMQGHYEMIDENGVPFSIDIPVFRLAV
PTLIH
```

Comments

The relationship between a gene and its product was originally 1-to-1. However, this relationship was modified to allow a gene to have more than one product, as in the case of infB, copA, and other genes that generate protein isoforms due to alternative or internal start codons.

Useful links

- Gene product from wikipedia

References

[2] Article: Gene product Wikipedia. URL: https://en.wikipedia.org/wiki/Gene_product.

Promoter

noun, plural: promoters

Definition

1. A promoter is the DNA sequence where RNA polymerase binds and initiates transcription.
2. "A promoter as a DNA segment essential for the specific initiation of transcription at a defined location in a DNA molecule, although this location might not be one single base. It is recognized by a specific Eσ, and this recognition is not necessarily autonomous." [1]
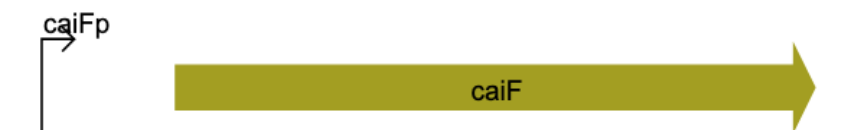

Figure: Graphical representation of a genome region showing one promoter and the caiF gene.

Acronym and synonyms

Properties

Property	Description		Example
name	Promoter name		caiFp
sigma factor	Sigma Factor that recognize the promoter		sigma70
tss	Genome position of Transcription Start Site (+1)		34218
sequence	Promoter Sequence (+1/TSS in upper case)		
strand	DNA strand where the promoter is located		
boxes	the sigma factor-binding motifs, known as -10/-12 and -35/-24 boxes		

Other optional properties for promoter can be queried in the [database model](#) of RegulonDB.

Example

- Promoter name: caiFp
- Confidence Level: Strong
- Transcription start site (+1): 34218
- Sigma Factor: Sigma70
- The promoter sequence in RegulonDB consists of 80 bp, with 60 bp upstream and 20 bp downstream from the precise initiation of transcription (+1). The binding motifs for sigma factors are marked as boxes

```
                                                                            +1
gatgacataagcaggatttag ctcaca cttatcgacggtgaagttgca tactat cgatatAtccacaattttaatatggcc
                        ─35                                   ─10
```

Comments

- Promoter sequences are specific to the different sigma factors associated to the RNA polymerase core. A promoter is represented as a stretch of 60 upstream and 20 downstream upper-case nucleotide sequences from the precise initiation of transcription, also called +1.
- By naming convention, all promoter names inherit the name of the first transcribed gene,

followed by 'p' to indicate that it is the promoter.

- A single promoter could have multiple TSSs within a 5-bp region, keeping a main TSS when known and the other like alternative. This change was made to take into account the flexibility in TSS selection due to the conformational dynamics of the transcription bubble of the open complex.
- Different σ factors initiating at the same TSS define different promoters. Overlapping promoters can initiate transcripts at the same TSS. For example, glmY expression in E. coli is controlled by two overlapping promoters, one recognized by σ54 and the other by σ70. RNAP holoenzyme is using different recognition elements even when selecting the same base as TSS.

Useful links

- Promoter from [Wikipedia](#)

References

[1] Mejia-Almonte, C., Busby, S.J.W., Wade, J.T., van Helden, J., Arkin, A.P., Stormo, G.D., Eilbeck, K., Palsson, B.O., Galagan, J.E. and Collado-Vides, J. (2020) Redefining fundamental concepts of transcription initiation in bacteria. Nat Rev Genet, 21, 699-714. PMID: 32665585

Terminator

noun

Definition

1. A **Terminator** is the region where transcription ends and RNAP unbinds from DNA.

2. A process in which the mRNA synthesis (i.e. transcription) or protein synthesis (i.e. during translation) stops at the terminator site [3].

3. A transcription terminator is a section of nucleic acid sequence that marks the end of a gene or operon in genomic DNA during transcription.

Acronym and synonyms

Properties

Property	Description		Example
lefEndPosition	Terminator left end position in the genome		274
rightEndPosition	Terminator right end position in the genome		310
strand	DNA strand where the terminator is located		forward
sequence	Terminator sequence		ggaaacacagAAAAAAG..GGCTTTTTTTTTcgaccaaagg
type	Terminator type		rho-independent
transcriptionUnit	Transcription unit(s) related to the terminator		thrLABC

Comments

"Two classes of transcription terminators, Rho-dependent and Rho-independent, have been identified throughout prokaryotic genomes. These widely distributed sequences are responsible for triggering the end of transcription upon normal completion of gene or operon transcription, mediating early termination of transcripts as a means of regulation such as that observed in transcriptional attenuation, and to ensure the termination of runaway transcriptional complexes that manage to escape earlier terminators by chance, which prevents unnecessary energy

expenditure for the cell."[4]
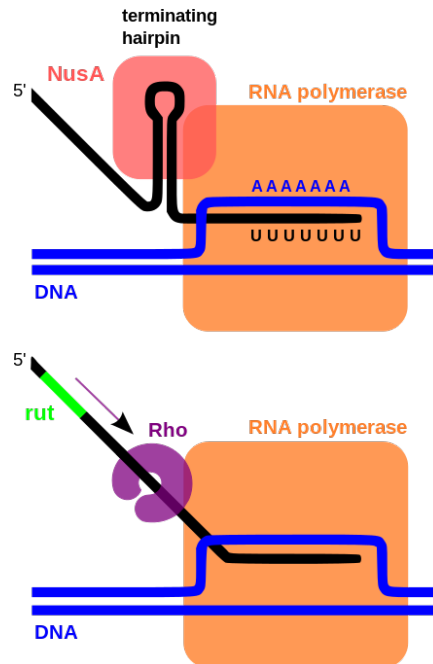
"Rho-dependent terminators. Rho-dependent transcription terminators require a large protein called a Rho factor which exhibits RNA helicase activity to disrupt the mRNA-DNA-RNA polymerase transcriptional complex. Rho-dependent terminators are found in bacteria and phages. "[4]

"Intrinsic transcription terminators or Rho-independent terminators require the formation of a self-annealing hairpin structure on the elongating transcript, which results in the disruption of the mRNA-DNA-RNA polymerase ternary complex. The terminator sequence in DNA contains a 20 basepair GC-rich region of dyad symmetry followed by a short poly-A tract or "A stretch" which is transcribed to form the terminating hairpin and a 7–9 nucleotide "U tract" respectively. "[4]

References

[3] https://www.biologyonline.com/dictionary/termination

[4] Article: Terminator (Genetics) https://en.wikipedia.org/wiki/Terminator_(genetics)

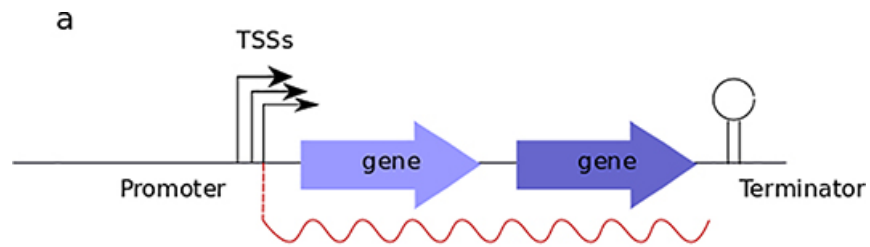# Transcription Unit

Definition

1. A Transcription Unit is as a DNA segment that starts from a TSS and ends at a transcription

termination site.

2. DNA segment delimited by different nonspurious TSS–TTS pairs

3. A Transcription unit is a set of one or more genes transcribed from a single promoter. A TU may also include regulatory protein binding sites affecting this promoter and a terminator.
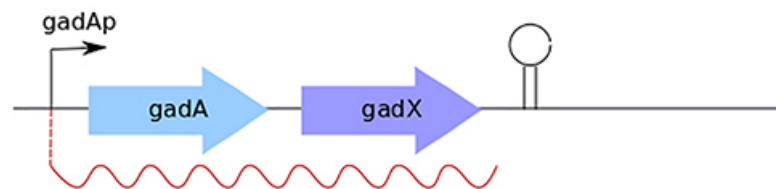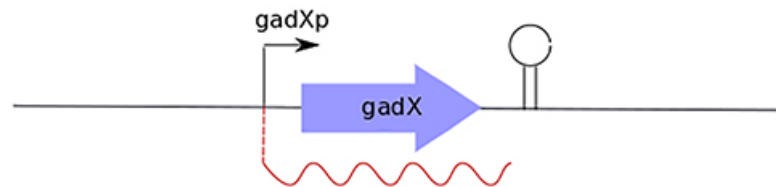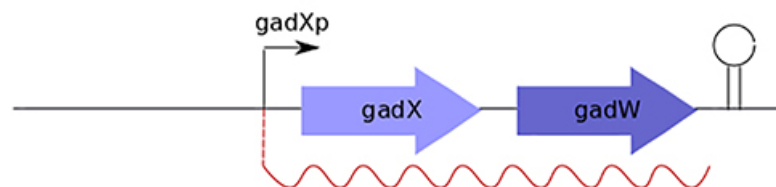
Figure: Transcription unit and operon schematic. When different transcription start sites (TSSs) from the same promoter are not differentially regulated, they form a single transcription unit that is limited by, but not including, a single promoter and a single terminator. b | Schematic of the

gadAXW complex operon. Several internal promoters and terminators enable different sets of genes to be co-transcribed in different Transcription Units.
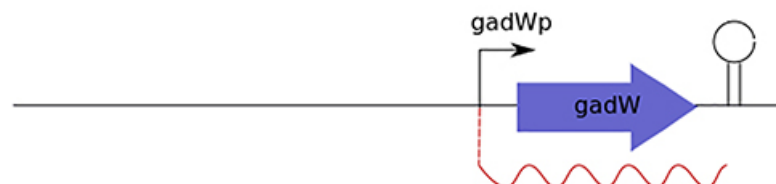
Acronym and synonyms

TU: Transcription Unit

Transcript

Properties

Property	Description		Example
name	transcription unit name	required	degQS
promoter id	promoter identifier		
terminator id	terminator identifier		
genes	transcribed genes		gedQ, gedS

Comments

References

[1] Mejia-Almonte, C., Busby, S.J.W., Wade, J.T., van Helden, J., Arkin, A.P., Stormo, G.D., Eilbeck, K., Palsson, B.O., Galagan, J.E. and Collado-Vides, J. (2020) Redefining fundamental concepts of transcription initiation in bacteria. Nat Rev Genet, 21, 699-714. PMID: 32665585

Operon

noun, plural: operons

Definition

1. An operon refers to a set of adjacent genes whose transcription is coordinated by one or several overlapping TUs transcribed in the same direction and that share at least one gene. A simple operon is one in which gene transcription is coordinated through a single TU, while a complex operon is one in which transcription is coordinated through several overlapping TUs, transcribed in the same direction and sharing at least one gene [1]

2. An operon is a set of one or several genes and their associated regulatory elements, which are transcribed as a single unit.

3. A group of genes or a segment of DNA that functions as a single transcription unit. It is comprised of an operator, a promoter, and one or more structural genes that are transcribed into one polycistronic mRNA [2].

Acronym and synonyms

Properties

Property	Description		Example
name	Operon name	required	abgABT-ogt
strand	DNA strand where the operon is coded	required	reverse
genes	Genes contained in the operon	required	abgA,abgB,abgT,ogt
transcriptionUnit(s)	Transcripts in the operon		

Example
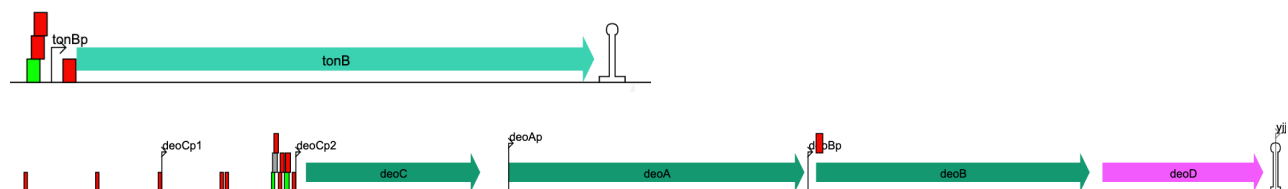
Figure. Examples of simple and complex operon.

Comments

The classical definition of operon is a group of two or more genes transcribed as a polycistronic unit (Jacob and Monod, JMB, 1961). For database purposes only, we extended the definition in order to include the possibility of operons with only one gene. Note: An operon is, therefore, one or more contiguous genes transcribed in the same direction. Please note that, according to this definition, an operon must contain a promoter upstream of all genes and a terminator downstream. It is also relatively common to find operons with several promoters, some of them internally located, thus, transcribing a partial group of genes. In all cases so far, one gene belongs to only one operon.

A complex operon with several promoters contains, therefore, several transcription units. According to the definition of operon, at least one transcription unit must include all the genes in the operon. See for instance the rpsU-dnaG-rpoD, glnALG, focA-pflB.

References

[1] Mejia-Almonte, C., Busby, S.J.W., Wade, J.T., van Helden, J., Arkin, A.P., Stormo, G.D., Eilbeck, K., Palsson, B.O., Galagan, J.E. and Collado-Vides, J. (2020) Redefining fundamental concepts of transcription initiation in bacteria. Nat Rev Genet, 21, 699-714. PMID: 32665585

[2] Operon. https://www.biologyonline.com/dictionary/operon

Regulon, transcription factors, and regulatory interactions

Regulatory gene products (Regulators)

noun, plural: regulators

Definition

1. Any gene product that increases or decreases the expression of a specific set of genes [1].

2. A gene that is involved in the production of a substance that controls or regulates the expression of one or more genes, such as the gene that codes for a repressor protein that inhibits the activity of an operator gene [2].

3. (general) A substance or process that regulates or controls another, as in a growth regulator that regulates the growth of an organism [3].
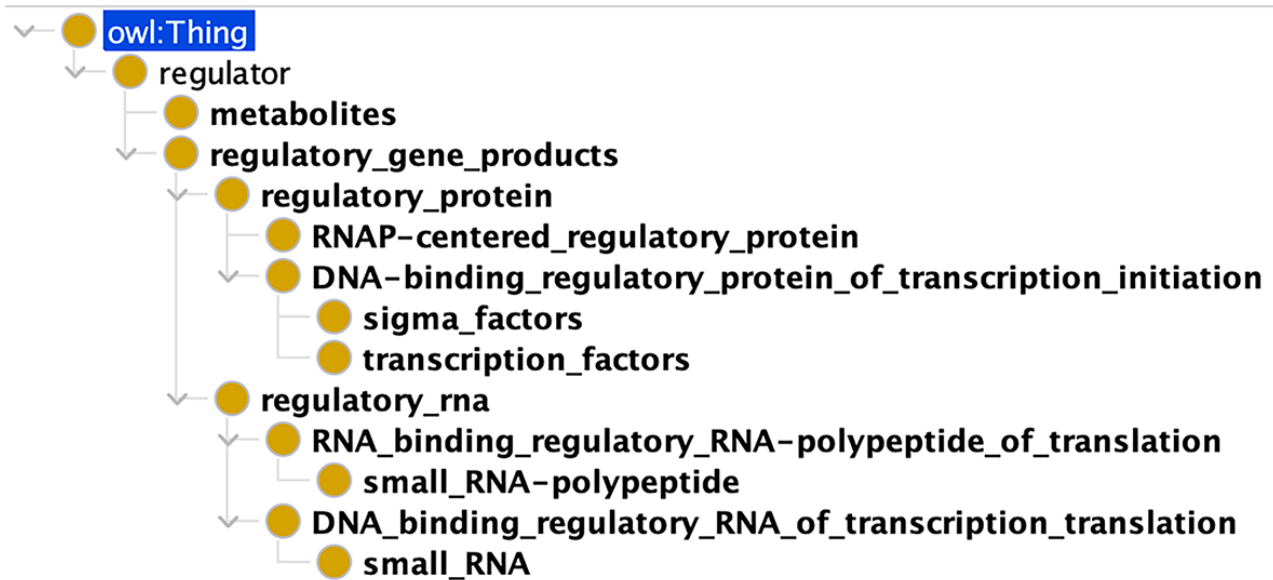
Figure. Classification of regulators.

Acronym and synonyms

- Regulatory gene

- TF: Transcription Factor.
- sRNA: small RNA

Properties

Property	Description		Example
name	regulator name	required	DNA-binding transcriptional dual regulator AraC
function	Gene expression effect caused by the regulator	required	activator, repressor
abbreviateName	regulator short name		AraC
synonyms	regulator synonyms		
type	regulator type		Transcription Factor, small RNA

Comments

Regulatory gene products can be classified into two main categories: RNA regulators and regulatory proteins. Moreover, these categories can be further subdivided based on the level at which they act to regulate gene expression or the specific mechanism of regulation. Specifically, the "DNA-binding transcription initiation regulatory protein" class is proposed as a subclass of regulatory proteins that bind to specific DNA sequences to regulate transcription initiation. This class includes TFs, which are regulatory proteins that bind near a promoter and affect transcription initiation at that promoter. It also includes sigma factors, which are regulatory proteins that bind to DNA during transcription initiation and are part of the RNA polymerase holoenzyme, which is essential for specific transcription initiation.

Additionally, another proposed subclass of regulatory proteins involving RNA polymerase, "RNAP-centered regulatory proteins of transcription initiation." These proteins, like DksA, regulate transcription initiation through interaction with the RNAP holoenzyme, also known as σE. Unlike DNA-binding regulatory proteins, the specificity of these proteins is not directly determined by the recognition of specific DNA sequences.

In our conceptual model, we address this diversity of regulators by their modeling into three main classes: gene products, regulatory elements not encoded in the DNA, and their combinations in regulatory complexes. These are the components of RIs and their active and inactive conformations. In this way, we can describe a variety of forms, such as: a) a regulatory gene product that does not require any other element for regulation; b) a regulatory complex formed by one or more regulatory gene products, either of the same or different gene product type or different type (some examples of these complexes are the dimeric AraC, the IHF complex composed of IhfA and IhfB, CRP complexed with cAMP, or CsrA complexed with CsrB RNA); and c) a regulatory effector, such as ppGpp, a metabolite that directly binds to the RNAP holoenzyme.

References

[1] Mejia-Almonte, C., Busby, S.J.W., Wade, J.T., van Helden, J., Arkin, A.P., Stormo, G.D., Eilbeck, K., Palsson, B.O., Galagan, J.E. and Collado-Vides, J. (2020) Redefining fundamental concepts of transcription initiation in bacteria. Nat Rev Genet, 21, 699-714. PMID: 32665585

[2] Regulatory Gene. https://www.biologyonline.com/dictionary/regulatory-gene

[3] Regulator. https://www.biologyonline.com/dictionary/regulator

Binding Site and Regulatory Site

noun

Definition

1. TFBS is defined as a DNA site where a TF binds specifically and that the term transcription factor regulatory sites (TFRS) be defined as the subset of TFBSs that are involved in transcription regulation [1].

2. In addition to TFs, there are other kinds of regulators, such as sRNAs and metabolites, whose specific binding sites are defined as regulator binding sites (R-BSs), and those with regulatory evidence are defined as regulator regulatory sites (R-RS).

Figure. Graphical representation of Binding Site(s).

Acronym and synonyms

- BS : Binding Site.
- RS: Regulatory Site.
- R-BS: Regulator Binding Site.

  - TFBS: Transcription Factor Binding Site.
  - sRNA-BS: small RNA Binding Site.

- R-RS: Regulator Regulatory Site.

  - TFRS: Transcription Factor Regulatory Site.
  - sRNA-RS: small RNA Regulatory Site.

Properties

Property	Description		Example
leftEndPosition	Absolute Left Genome Position		
rightEndPosition	Absolute Right Genome Position		
sequence	binding site sequence		
length	binding site length		

Comments

References

[1] Mejia-Almonte, C., Busby, S.J.W., Wade, J.T., van Helden, J., Arkin, A.P., Stormo, G.D., Eilbeck, K., Palsson, B.O., Galagan, J.E. and Collado-Vides, J. (2020) Redefining fundamental concepts of transcription initiation in bacteria. Nat Rev Genet, 21, 699-714. PMID: 32665585

Regulatory Interactions

Definition

1. RI is a quadruple interaction (T, R, R-RS, F), where T represents the target, R represents the regulator, R-RS represents the site where the regulator binds and participates in gene regulation, and F represents the function or regulatory effect on the target [1]
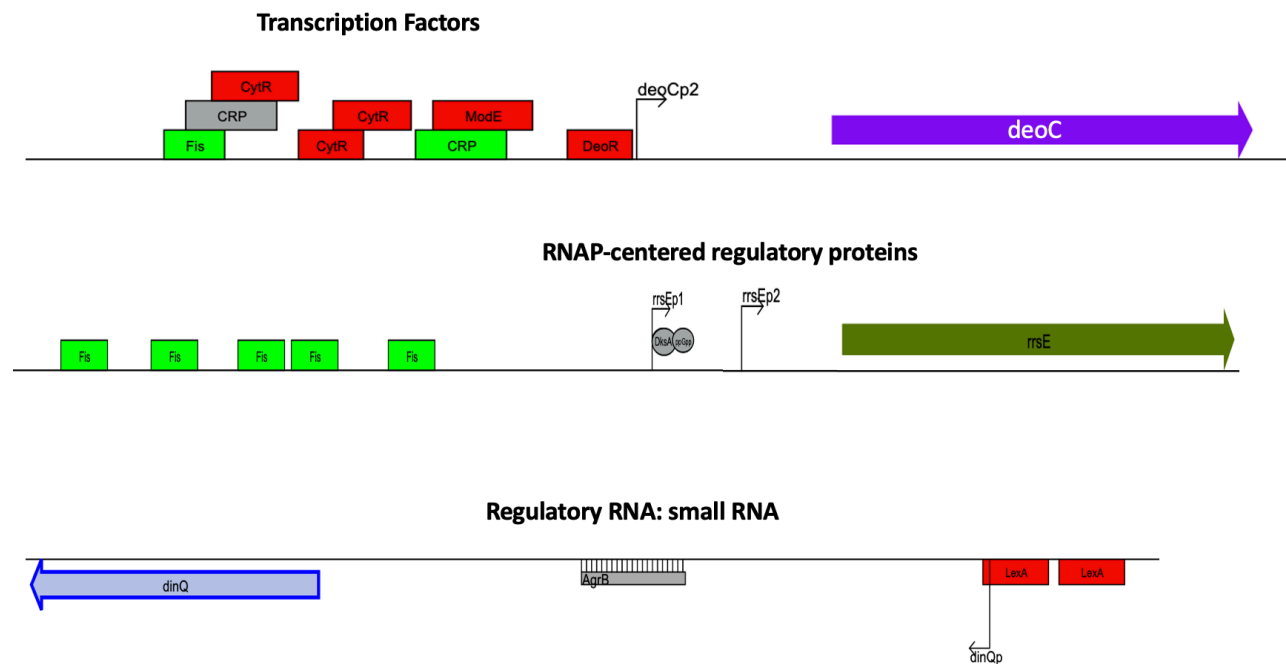
Figure. Three types of regulators: Transcription Factors, RNAP centered regulatory gene product and small RNA regulator.

Acronym and synonyms

RI: Regulatory Interaction

Properties

Property	Description		Example
regulator		required	
regulated entity	Transcription unit, gene (id:name) regulated by the Regulator	required	
function	Gene expression effect caused by the Regulator bound to the RS		activator, repressor
regulatory site	Regulator Regulatory site (R-RS)		
relativeDistRSPromoter	Relative distance from the center position of RS to the Transcription Start Site		-20, +45

Comments
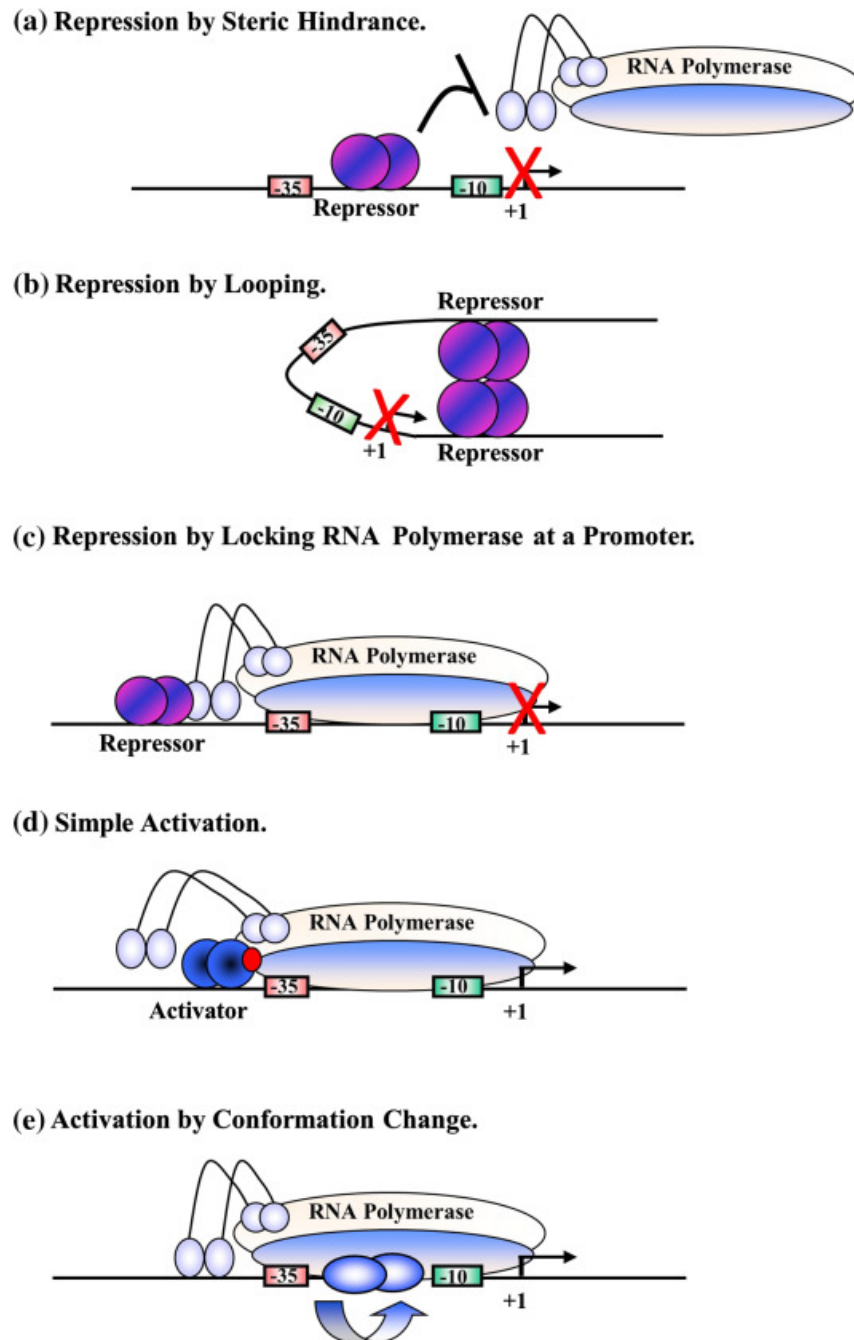
Figure. Mechanisms of repression and activation by transcription factors at bacterial promoters.
Source: https://doi.org/10.1016/j.jmb.2019.04.011

In RegulonDB version 12.0, these TF binding sites were renamed as transcription factor regulatory sites (TFRSs), since they exert an effect on the regulated promoter, unlike transcription factor binding sites (TFBSs), that may or may not create an effect. This distinction is relevant, since current HT binding methods identify TFBSs as lacking evidence of their regulatory role. In addition to TFs, there are other kinds of regulators, such as sRNAs and metabolites, whose specific binding sites are defined as regulator binding sites (R-BSs), and those with regulatory evidence are defined as regulator regulatory sites (R-RS).

References

[1] Mejia-Almonte, C., Busby, S.J.W., Wade, J.T., van Helden, J., Arkin, A.P., Stormo, G.D., Eilbeck, K., Palsson, B.O., Galagan, J.E. and Collado-Vides, J. (2020) Redefining fundamental concepts of transcription initiation in bacteria. Nat Rev Genet, 21, 699-714. PMID: 32665585

Attenuator

noun

Definition

1. A control region at the promoter end of repressible amino acid operons that exerts transcriptional control based on the translation of a small leader peptide gene [1].

Figure. Attenuator graphical representation.

Acronym and synonyms

Properties

Example

The attenuator is adjacent to the structural genes of the trp (tryptophan) operon; in the presence of tryptophan this region acts to reduce the rate of transcription from the structural genes.

High level of tryptophan

complete leader peptide

ribosome

Stem loop: transcription termination signal

RNA pol

5' mRNA 1 2 3 4 UUUU 3'

Trp codons

DNA trpL

Low level of tryptophan

incomplete leader peptide

ribosome stalls

Alternate stem loop: transcription continues

RNA pol

5' 1 2 3 4 trp regulated genes

Trp codons

DNA trpL
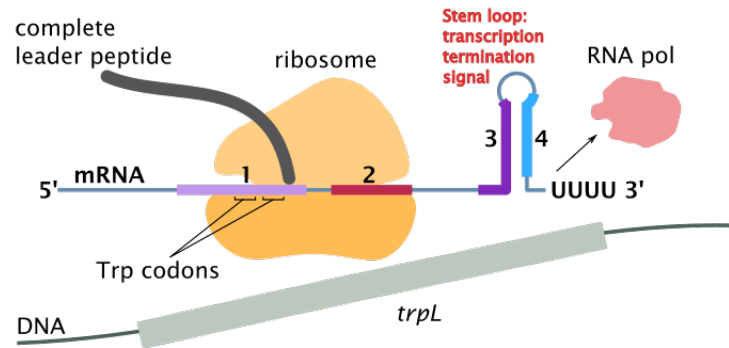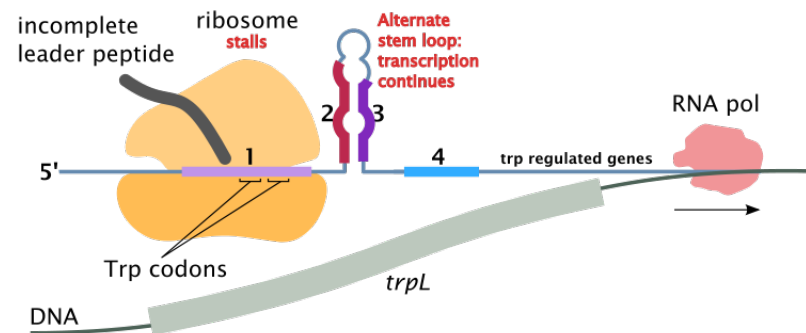
Comments

Attenuation is a regulatory mechanism for some bacterial operons that results in premature termination of transcription.

For each predicted operon, the upstream region of the first gene is taken (Figure 1a). For every run of Us present in this region (Figure 1b), a stable structure in the adjacent region is searched for (Figure 1c). If a terminator is found, an anti-terminator is searched for it must be overlapping with the terminator (Figure 1d). An anti-antiterminator can be analogously located by finding a structure that overlaps with the anti-terminator (Figure 1e). For the particular case of translational attenuators, a terminator is searched for that overlaps with the Shine-Dalgarno site.
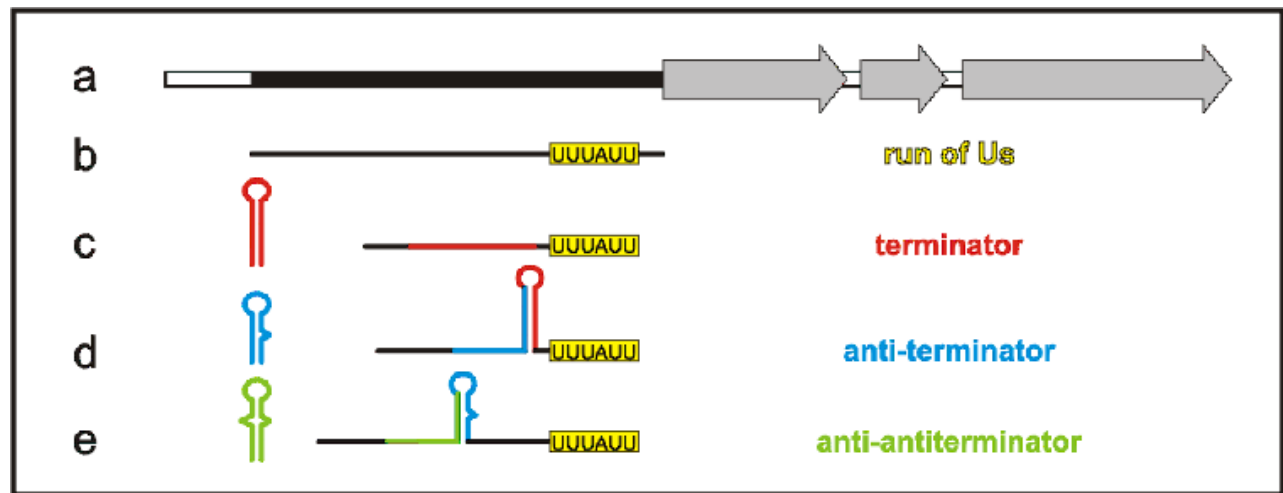
Useful links

References

[1] Attenuator. https://groups.molbiosci.northwestern.edu/holmgren/Glossary/Definitions/Def-A/attenuator.html

GENSOR Unit

noun

Definition

1. The GENSOR units show the cascade of information processing that begins with the appearance of a signal, its transformation into the metabolite that binds the TF, called an effector, the genetic switch set of the TF binding to its TFRSs, and the response that includes all of the activated or repressed target genes.

Figure. GENSOR unit components and its graphical representation.

Acronym and synonyms

GU : Gensor Unit

Properties

Property	Description		Example
GU name	GENSOR Unit name (inherited from the name of the Regulator)	required	AraC, AdiY
Description			
Signal		required	
Reactions		required	
Components			
Group			

Example

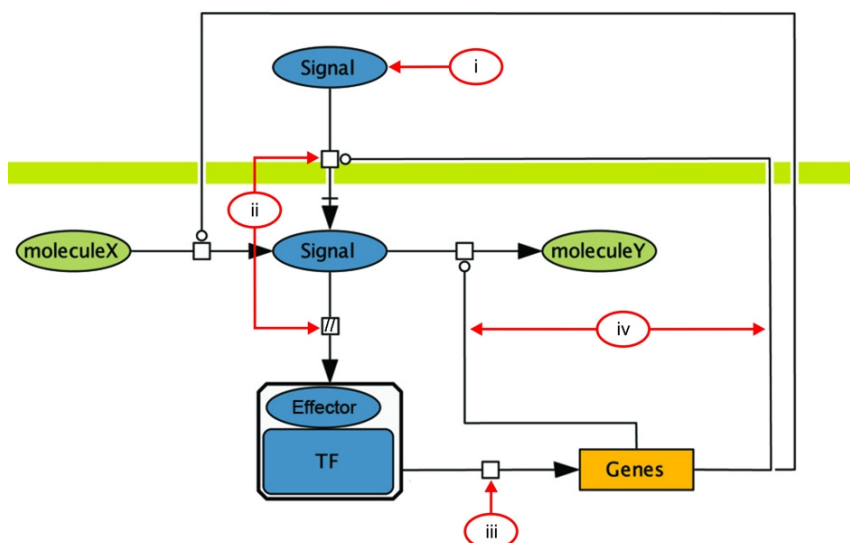Figure. AraC-arabinose GU. In the presence of l-arabinose, AraC activates transcription of genes that code for proteins necessary for utilization and transport of l-arabinose. In the absence of glucose CRP coregulates with AraC. The four GU elements are represented by different colors in the image: blue, the signal (i) and signal transduction (ii); yellow, genetic switch (iii); green, the response (iv) [2].

Comments

The ability of a cell to respond to changes inside the cell or in the environment initiates when the new signal or stimulus is sensed and transmitted through a series of molecular concatenated reactions, called signal transduction or transduction pathways. These events bring into action genetic switches that modify gene expression to produce the adequate response to the cell. We call these processes genetic Sensory Response Units, or Gensor Unit. Gensor Unit is defined by four components: i)the signal or stimulus, ii) the signal transduction pathway, iii) the regulatory mechanisms governing the expression of genes, and iv) the adequate response resulting from the modified gene expression of the affected set of target genes.

References

[1] RegulonDB v12.0

[2] Gama-Castro S, Salgado H, et al. RegulonDB version 7.0: transcriptional regulation of Escherichia coli K-12 integrated within genetic sensory response units (Gensor Units). Nucleic Acids Res. 2011 Jan;39(Database issue):D98-105. doi: 10.1093/nar/gkq1110. Epub 2010 Nov 4. PMID: 21051347; PMCID: PMC3013702.

Riboswitch

Definition

1. Riboswitches are gene control elements that directly bind to specific ligands to regulate gene expression without the need for proteins. [1]

2. Riboswitches are functional RNA elements that regulate gene expression by directly detecting metabolites. [2]

Figure. Riboswitch graphical representation in RegulonDB.

Acronym and synonyms

Properties

Example

Comments

"When the mechanism for riboswitch control of gene expression is done via prevention of translation, the ribosome-binding site (RBS) is blocked upon ligand binding to the aptamer domain and the RNA cannot be translated into a protein. In the absence of the ligand, the RBS is unblocked and therefore the mRNA is translated. In the case of transcriptional termination, alternative terminator or anti-terminator structures are formed that either terminate or promote transcription, respectively (Penchovsky and Stoilova, 2013; Penchovsky and Traykovska, 2015). In the presence of specific ligand that binds to the aptamer domain, a terminator structure is usually formed into the 5′-UTR region, which stops the RNA transcription. In the absence of the ligand, an anti-terminator structure is formed instead of the terminator and, because of that, the complete mRNA is transcribed and translated into a protein. In addition, there is one more cis-acting mechanism of riboswitch control of gene expression, which functions via RNA destabilization. [1]"

"Riboswitches are RNA elements mostly found in the 5′-UTR of bacterial mRNAs, where they sense concentration change of specific metabolites, usually called ligands, and regulate
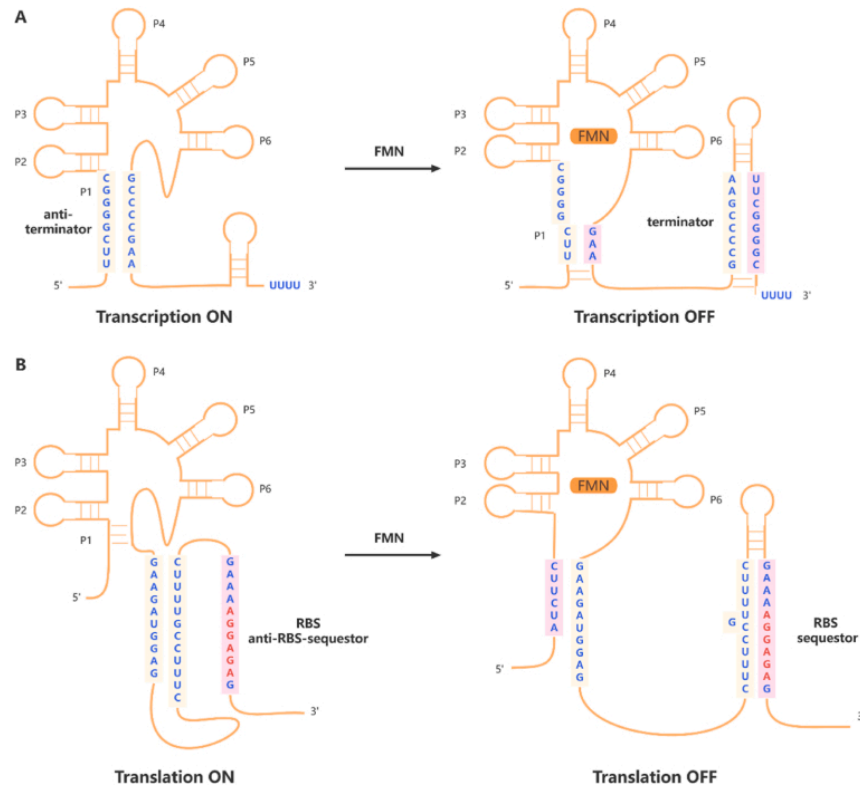
downstream gene expression in cis through tertiary structure changes of mRNAs, the controlled genes are always related to metabolic reaction of ligands [2]".

Figure. Two main regulatory mechanisms of riboswitches. (A) Transcriptional level, ribD FMN riboswitch of Bacillus subtilis, (B) Translational level, ypaA FMN riboswitch of B. subtilis. Blue sequence in beige box represents anti-terminator sequence or anti-RBS-sequestor sequence, in pale pink box represents sequence paired with anti-terminator sequence or anti-RBS-sequestor sequence, red sequence represents RBS. Riboswitches regulate gene expression through influencing terminator or RBS structure when bound to ligands [2].

Useful links

References

[1] Nikolet Pavlova, Dimitrios Kaloudas, Robert Penchovsky. Riboswitch distribution, structure, and function in bacteria, Gene, 2019, doi: 10.1016/j.gene.2019.05.036.

[2] Xu J, Hou J, Ding M, Wang Z, Chen T. Riboswitches, from cognition to transformation. Synth Syst Biotechnol. 2023 Jun 3;8(3):357-370. doi: 10.1016/j.synbio.2023.05.008. PMID: 37325181; PMCID: PMC10265488.

Shine Dalgarno

Definition

1. Shine-Dalgarno or Ribosome binding site is the polypurine sequence and its consensus is AGGAGG. It is located on bacterial mRNA, just prior to an AUG initiation codon; it is complementary to the sequence at the 3' end of 16S rRNA; involved in binding of ribosome to mRNA.

Acronym and synonyms

RBS: Ribosome Binding Site

Properties

Example

Comments

Useful links

References

Regulon

Definition

1. Regulon is defined as a set of genes subject to regulation of one and only one regulator (Maas WK , 1964, PMID:14168690). Note: The initial definition was derived from studies of the arginine biosynthetic genes, which were, contrary to operons, found to be scattered (non contiguously located) in the chromosome of E.coli. To better describe the alternative groups of co-regulated genes, we now call this a simple regulon, as opposed to a complex regulon. Complex regulon A group of genes subject to regulation by two or more regulators, where all genes are subject to the regulation of exactly the same transcription factors. Strict complex regulon: Complex regulons can still be subdivided into strict complex regulons. A strict complex regulon is a set of genes subject to regulation by two or more transcription factors, where the effect of each regulator (activator or repressor) is the same for all the regulated genes.

Properties

Example

Comments

Useful links

References

Matrix

Definition

1. A matrix, weight matrix, or positional weight matrix represents a collection of aligned binding sequences for the same transcriptional regulator. It is a derivative of a multiple alignment of such sites. Each row corresponds to one of the letters of the relevant alphabet -e.g., 4 rows in the case of DNA. Each column corresponds to one of the positions within the aligned sites. A frequency matrix contains the frequency of the four nucleotides at each position.

Properties

Example

Comments

Useful links

References

Alignment

Definition

1. A multiple alignment with the collection of binding sites for a regulatory protein is generated by using initially extended binding sites. RegulonDB contains such mutiple alignments,

generated by using the Wconsensus program (Hertz GZ, Comput Appl Biosci. 1990 Apr;6(2):81-92.)

Properties

Example

Comments

Useful links

References

Growth Condition

Definition

1. Growth conditions are the experimental conditions in which a strain is grown in particular experiments performed to study changes in gene expression.

Properties

Example

Comments

Useful links

References

Effector

Definition

1. We call effector in RegulonDB the precise metabolite that binds to the TF, altering its conformation and involved in the switch for the binding-unbinding specifically to its TFBSs

sites. Most effectors are metabolites that bind non-covalently to an allosteric TF site. We include as effectors covalent modifications, i.e. phosphorylations for the two component TFs. The literature may be confusing since effectors are also called "signals". A signal in RegulonDB is distinguished clearly from effectors. See GUs for more detail.

Properties

Property	Description		Example

Example

Comments

Useful links

References

References

References

[1] Mejia-Almonte, C., Busby, S.J.W., Wade, J.T., van Helden, J., Arkin, A.P., Stormo, G.D., Eilbeck, K., Palsson, B.O., Galagan, J.E. and Collado-Vides, J. (2020) Redefining fundamental concepts of transcription initiation in bacteria. Nat Rev Genet, 21, 699-714. PMID: 32665585

[2] Article: Gene product Wikipedia. URL: https://en.wikipedia.org/wiki/Gene_product

[3] https://www.biologyonline.com/dictionary/termination

[4] Article: Terminator (Genetics) https://en.wikipedia.org/wiki/Terminator_(genetics)