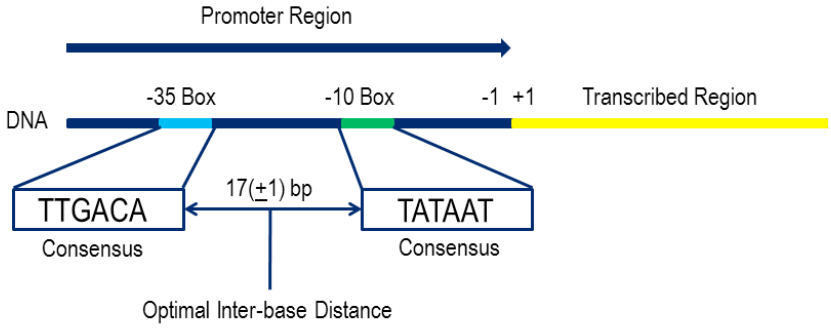
TECHNIQUES IN MOLECULAR BIOLOGY – PROMOTERS AND PLASMIDS

Class Notes Only; For detailed information and background reading ***see linked pages*** *on the class webpage*.

Definition of a Gene - no longer associated with a morphology "Green eyes" or "Wrinkled Peas". Many proteins are multimeric both hetero and homomeric (i.e. hemoglobin 2 types of polypeptides make a heterotetramer of 2 α and 2 β chains). Thus, one gene - one protein doesn't match actual function. Eukaryotes include intervening, non-polypeptide coding regions (introns)... thus a more reasonable definition of a gene might be the open reading frame (ORF) or protein coding region. The non-coding, untranslated regions generate RNA forms that are also responsible for regulating expression of the ORF and could be included in the generic term "GENE".

Basic definition: *Cis-acting elements*: The DNA that affects the expression of the adjacent gene. *Trans-acting element*: Proteins which bind the cis-acting elements, regulating (positive and negative) RNA polymerase activity and thus expression of gene.

Regulation of Gene Expression - Protein expression is controlled at both transcription and translation.



Transcription: *Promoters* - Bacterial promoters are more simple than eukaryotic promoters. Promoters, simply put, promote the binding and initiation of RNA polymerases to generate mRNA. Bacterial promoters typically consist of two short DNA sequences separated by 18-16 base pairs (most commonly 17 bp). The flanking regions, called consensus regions are centered around the -35 and -10 from the start site. NOTE there is no "0" but a -1 and +1 for the beginning of the ORF (for bacteria).

Strength of a promoter depends on the -10, -35 and ~17 bp between the consensus sequences. Promoters can be constitutive (regulated instead by RNA polymerase levels or sigma factors), positively regulated (increased promoter activity based on some transcription (trans-activating) factor, also known as an activator, increases in cellular concentration), and negatively regulated (decreased activity as a repressor protein (trans-activating element) is present).

- Promoters are often referred to as being inducible - that means there is always some repressor present to inhibit the promoter activity. Or that a promoter has low activity (RNA polymerase binding) until an activator is present. Both repressor and activator proteins are often regulated by a ligand (small molecule) created by the bacteria or exogenously added to the bacteria's environment.
- Some promoters have both activators and repressors.

RNA Polymerases - Bacterial RNA polymerases consists of five proteins/polypeptides. A sixth protein subunit is required for activity. This sixth protein is called a sigma (σ) factor. The RNA polymerase binds non-specifically to DNA, the sigma factor binds to the RNA polymerase and the promoter region between the -10 and -35 consensus region. In E. coli, there are seven different sigma factors, each expressed under different conditions. Examples:

σ^{70} holenzyme ($E\sigma^{70}$) A.K.A σ^{D}

- Used during exponential growth for the transcription of most genes. σ^{70} is the most common sigma factor in *E. coli*.
- **Consensus promoter sequence:** TTGACA__(17)__TATAAT

σ^S holenzyme ($E\sigma^S$) A.K.A σ^{38}
- Used during late exponential and stationary growth phases. Is also is a major regulator of the response of *E. coli* to starvation.

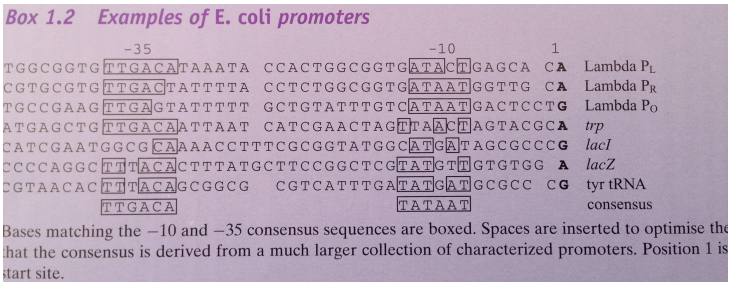
Consensus promoter sequence: TTGACA__(17)__TATAAT

σ^{28} holenzyme ($E\sigma^{28}$) A.K.A σ^F
- This sigma factor is regulates the transcription of genes necessary for motility of *E. coli*.

Consensus promoter sequence: TAAAGWWY__(11/12)__RYCGAWRN

These promoters are divided into housekeeping and environmentally responsive promoters. Housekeeping genes like sigma 70 is constitutively expressed, while sigma S or 28 are under expression promoters responsive to conditions of food/nutritional source, and thus activate a select set of genes via their cis-activating elements.

Consensus and Promoter Sequence - The actual sequence of the -10, -35 and 17 intervening bp are clearly not the same for all promoters and reading frames. The term consensus gives us an indication of the true nature of what the sequence is. Consensus can mean a general agreement. That is a good explanation of a promoter consensus sequence. The sequence shown here is not the exact sequence found nearly anywhere. A consensus sequence is a statistical creation. The linked page on Bacterial promoters gives an exercise demonstrating how such a sequence is generated.


Box 1.2 Examples of E. coli promoters

	−35	−10	1	
	TGGCGGTGTTGACATAAATA	CCACTGGCGGTGATACTGAGCA	CA	Lambda P_L
	CGTGCGTGTTGACTATTTTA	CCTCTGGCGGTGATAATGGTTG	CA	Lambda P_R
	TGCCGAAGTTGAGTATTTTT	GCTGTATTTGTCATAATGACTCCTG		Lambda P_O
	ATGAGCTGTTGACAATTAAT	CATCGAACTAGTTAACTAGTACGCA		*trp*
	CATCGAATGGCGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCG			*lacI*
	CCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGG		A	*lacZ*
	CGTAACACTTTACAGCGGCG	CGTCATTTGATATGATGCGCC	CG	tyr tRNA
	TTGACA	TATAAT		consensus

Bases matching the −10 and −35 consensus sequences are boxed. Spaces are inserted to optimise the hat the consensus is derived from a much larger collection of characterized promoters. Position 1 is tart site.

(from the handout...) In 1987, **Calvin Harley** and **Robert Reynolds** of McMaster University published an analysis of 263 **phage**, **plasmid** and **bacterial** promoters:

"In the final compilation, all bases in the -35 (TTGACA) and -10 (TATAAT) hexamers were highly conserved, 92% of promoters had inter-region spacing or 17±1 bp, and 75% of the uniquely defined start points initiated 7±1 bases downstream of the -10 region."

The degree of conservation (%) of each base in their compilation of the consensus hexamers was:

$$T_{78}T_{82}G_{68}A_{58}C_{52}A_{54} - - 16_{21}\ 17_{52}\ 18_{19} - - T_{82}A_{89}T_{52}A_{59}A_{49}T_{89}$$

- A follow-up examination of other promoters that are not as strong found a slightly different probability

$$T_{69}T_{79}G_{61}A_{56}C_{54}A_{54} - - 16_{17}\ 17_{43}\ 18_{17} - - T_{77}A_{76}T_{60}A_{61}A_{56}T_{82}$$

- When one considers the various sigma factors, a consensus sequence is pretty unique and should only be considered for a particular RNA polymerase and the associated sigma subunit.

Not all RNA needs to be produced at the same amount or all of the time, thus a control at the transcription (promoter and sigma factor) level and transcriptional level (see next section) is how most bacteria control protein expression. Thus there are strong and weak promoters.

A classic strong promoter is the recA promoter:

recA	TTGA**T**A -- **16** -- TATAAT
Std consensus sequence	TTGA**C**A-- **17** -- TATAAT

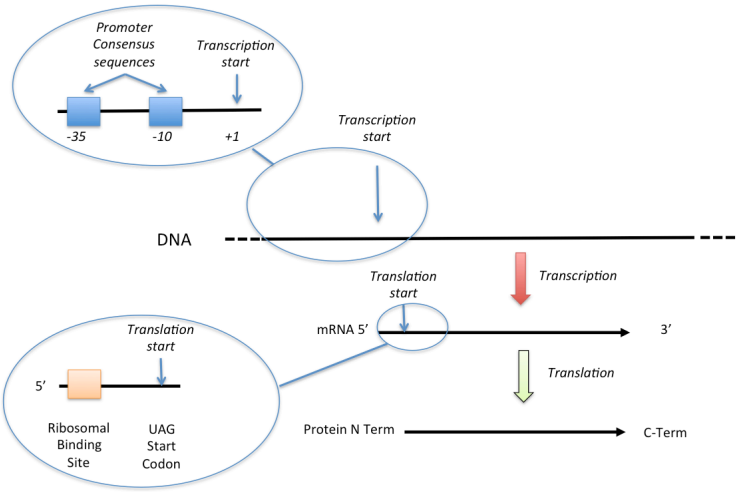
Notice the two small changes that shift the promoter strength.

The *araBAD* promoter is a well-known weak promoter:

araBAD	**C**TGACA -- **18** -- TA**CTG**T
Std consensus sequence	**T** TGACA-- **17** -- TA **TAA**T

Again, notice the changes driving a weakened promoter.

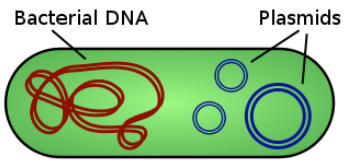
In addition to the -10 and -35 sequences, other regions can impact gene regulation. These sequences are gene/promoter specific and it is important to be aware of such things. Some strong promoters have a heavy A/T rich region upstream of the promoters, especially those involved with promoters of housekeeping ribosomal RNA. These sequences are typically -47 to -57 bp and help promoter binding by allowing the C-terminal domain of the alpha subunit of RNA polymerase to bind to DNA with several fold greater affinity, thus needing less RNA polymerase to bind and process the gene and increasing transcription.

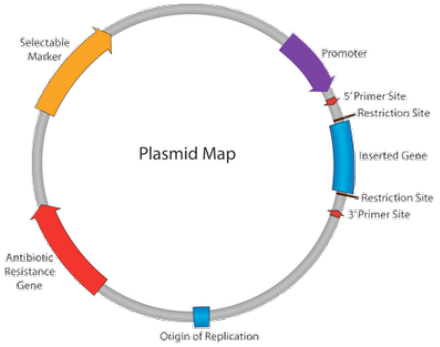


Translation - Ribosomal Binding Site (RBS): An additional step of regulation is during translation. In translation, the ribosome binds to a specific site of the mRNA produced in transcription. Translation requires BOTH the RBS and a start codon. This site, also known as the Shine-Dalgarno sequence, is bound by the 16s rRNA. Like the promoter sequences, there is a significant variation in the RBS sequence. Sequences rich in AG (purine) bp will lead to high rates of initiation of translation... i.e. rRNA binding to start site on mRNA. The more complimentary the RBS is to the 16s rRNA the better the binding and higher the rate of translation. The RBS also aids in processing the movement of the rRNA along to the start site. An excellent detailed explanation can be found on the linked pages (Registry of Standard Parts). The rate of translation is also impacted by the start codon, there are several. The bacterial start codon, UAG is not the same as found in other organisms (GUG, UUG, CUG and so on...). Thus when cloning a protein from a foreign organism into bacteria, the RBS and start codon should be carefully considered. Finally, the stability of the mRNA also may impact the rate of translation. The more stable the mRNA the more rRNA can process through expressing proteins. The RBS can affect the mRNA by increasing degradation by RNase enzymes (some RBS sites also code for mRNA cleavage by RNase enzyme) and thus the amount of protein synthesized.

...AUAAAGGAGGUAAAUAAUG→
RBS start codon

A typical RBS sequence is located about 6 nucleotides upstream of a start codon in an mRNA. The ribosomal holoenzyme binds to both the RBS and the start codon. The start codon and everything downstream are translated by the ribosome.





Plasmid and Vectors - Plasmids are double stranded, extra-chromosomal circular DNA. Bacteria can contain a few (low copy number) to hundreds (high copy number) of plasmids. Plasmids replicate using the bacterial machinery if the propter cis-acting elements are in place. A vector is a plasmid and the two terms are often used interchangeably. However, some vectors are specifically designed to contain DNA (also called a cloning vector) while other vectors are specifically engineered to express protein. For the most part, plasmids and vectors are the same, one should pay attention to the elements of the plasmid/vector to use the vector most appropriately.

Elements of a Plasmid: Most plasmids are small from 1-200 bp in size with common eukaryotic proteins ranging from 4-12 kB. The basic role for plasmids are to carry genes for altering DNA or for expressing proteins. Plasmids are replicated by bacteria but unless there is a bacterial promoter will not express protein. Plasmids are often used to transiently transfect cells of other organisms. Phages are also used for this purpose. We will discuss cloning an insert in and out of a plasmid later. However, depending on the use, it is important to have the correct RNA promoter for expression in the appropriate host cell. i.e. a bacterial promoter will make plenty of protein in bacteria but not in mammalian cells. See links of common plasmids and interpret their maps.

Plasmid Element	Description *(Map and table from addgene)*
Origin of Replication (ORI)	DNA sequence which allows initiation of replication within a plasmid by recruiting transcriptional machinery proteins
Antibiotic Resistance Gene	Allows for selection of plasmid-containing bacteria.
Multiple Cloning Site (MCS)	Short segment of DNA which contains several restriction sites allowing for the easy insertion of DNA. In expression plasmids, the MCS is often downstream from a promoter.
Insert	Gene, promoter or other DNA fragment cloned into the MCS for further study.
Promoter Region	Drives transcription of the target gene. Vital component for expression vectors: determines which cell types the gene is expressed in and amount of recombinant protein obtained.
Selectable Marker	The antibiotic resistance gene allows for selection in bacteria. However, many plasmids also have selectable markers for use in other cell types.
Primer Binding Site	A short single-stranded DNA sequence used as an initiation point for PCR amplification or sequencing. Primers can be exploited for sequence verification of plasmids