



#### Programa de Pós-Graduação em Bioquímica e Biologia Molecular

#### **BBM5002 - Bioquímica e Biologia Molecular**

**Docente:** Prof. Dr. Felipe S. Chambergo – <u>fscha@usp.br</u> https://sites.usp.br/lbbp/

Data: Segunda-feira 14 – 16 h / Sexta-feira 8 - 12 h.

USP-2021-1S

#### Controle da Expressão Gênica em Eucariotes

- 1
- Transcrição: Expressão/Ativação do gene
- 2
- **Processamento Pós-transcrição:** Remoção de introns/poliadenilação, Caping, Transporte
- **Degradação mRNA:** Estabilidade e tempo de vida
- 4
  - Tradução: Leitura correta
- 5
- **Processamento Pós-tradução:** Formação de enlaces, adição de outras moléculas.
- 6
- **Degradação de proteínas:** Proteínas inativa ou desenoveladas
- **Endereçamento e Transporte: Destino** da proteína



#### Organização e estrutura dos genes em procariotos



#### Disposição dos genes no genoma eucariótico



(Takai et al., 2004)

#### Estrutura do gene eucariótico



#### Organização e estrutura de um gene eucariótico



#### Transcrição: Expressão/Ativação do gene

Expressão gênica constitutiva: Vias metabólicas centrais
Expressão gênica regulada: Indução/Repressão, em resposta ao estado metabólico ou sinais.



#### Interação Proteínas : DNA



#### Mecanismos de regulação da iniciação da transcrição

Regulação negativa (Repressor ligado inibe a transcrição)



#### Regulação negativa (Repressor ligado inibe a transcrição)



## The sense (nontemplate) strand sequences of selected *E. coli* promoters.

Operon	-35 region	-10 region	Initiation
		(Pribnow box)	site (+1)
lac	ACCCCAGGCTTTACACTTTATGCTTCCGGCT	CG <mark>TATGTT</mark> GTGT	<b>TGGAATTGTGAGCGG</b>
lacI	CCATCGAATGGCGCAAAACCTTTCGCGGTATC	GG <mark>CATGAT</mark> AGCO	GCCCGGAAGAGAGTC
galP2	ATTTATTCCATGTCACACTTTTCGCATCTTTC	GT <mark>TATGCT</mark> ATG(	GTTATTTCATACCAT
araBAD	GGATCCTACCTGACGCTTTTTTATCGCAACTC	ГС <mark>ТАСТGТ</mark> ТТСТ	TCCATACCCGTTTTT
araC	GCCGTGATTATAGACACTTTTGTTACGCGTT	FT <mark>TGTCAT</mark> GGC1	TTTGGTCCCGCTTTG
trp	AAATGAGCTG <mark>TTGACA</mark> ATTAATCATCGAACTA	AG <mark>TTAACT</mark> AGTA	ACGCAAGTTCACGTA
bioA	TTCCAAAACGTGTTTTTTTGTTGTTAATTCGG	ГG <mark>TAGACT</mark> TGTA	AACCTAAATCTTTT
bioB	CATAATCGACTTGTAAACCAAATTGAAAAGA	ГТ <mark>ТАGGTТ</mark> ТАСА	AGTCTACACCGAAT
$t \mathbf{RNA}^{Tyr}$	CAACGTAACACTTTACAGCGGCGCGCGTCATTTC	GA <mark>TATGAT</mark> GCG(	CCCCGCTTCCCGATA
rrnD1	CAAAAAAATACTTGTGCAAAAAATTGGGATCO	CC <mark>TATAAT</mark> GCG(	CCTCCGTTGAGACGA
rrnE1	CAATTTTTCTATTGCGGCCTGCGGAGAACTCC	CC <mark>TATAAT</mark> GCG(	CCTCCATCGACACGG
rrnA1	AAAATAAATGCTTGACTCTGTAGCGGGAAGGG	CG <mark>TATTAT</mark> GCA(	CACCCCGCGCCGCTG



#### **Promoter structure in prokaryotes**



consensus sequences

#### The base sequence of the *lac* operator.



#### The nucleotide sequence of the *E. coli lac* promoter– operator region.



#### The lactose operon in E. coli

- promoter binds CAP and RNA polymerase
- operator binds the lac repressor



• the function of the lactose (lac) operon is to produce the enzymes required to metabolize lactose for energy when it is required by the cell

#### **Regulation of the lactose operon - negative control**



#### Alleviation of negative control - action of the inducer of the lac operon

- when lactose becomes available, it is taken up by the cell
- allolactose (an intermediate in the hydrolysis of lactose) is produced
- one molecule of allolactose binds to each of the repressor subunits
- binding of allolactose results in a conformational change in the repressor
- the conformational change results in decreased affinity of the repressor for the operator and dissociation of the repressor from the DNA



#### **NO TRANSCRIPTION**

• IPTG (isopropyl thiogalactoside) is also used as a (non-physiological) inducer

- repressor (with bound allolactose) dissociates from the operator
- negative control (repression) is alleviated, however...



• RNA polymerase cannot form a stable complex with the promoter

#### **Regulation of the lactose operon - positive control**

- in the presence of <u>both</u> lactose and glucose it is not necessary for the cell to metabolize lactose for energy
- in the <u>absence</u> of glucose and in the <u>presence</u> of lactose it becomes advantageous to make use of the available lactose for energy
- in the absence of glucose cells synthesize cyclic AMP(cAMP)
- cAMP<sup>1</sup> serves as a positive regulator of catabolite operons (lacoperon)
- cAMP binds the dimeric cAMP binding protein(CAP)<sup>2</sup>
- binding of cAMP increases the affinity of CAP for the promoter
- binding of CAP to the promoter facilitates the binding of RNApolymerase

 $^{1}$  cAMP = 3', 5' cyclic adenosine monophosphate



#### **NO TRANSCRIPTION**

<sup>2</sup> also termed catabolite activator protein

#### **Activation of lac operon transcription**



inactive repressor

• the function of the lactose (lac) operon is to produce the enzymes required to metabolize lactose for energy when it is <u>required</u> by the cell

#### A genetic map of the *E. coli araC* and *araBAD* operons.



# In the absence of arabinose, the araC protein inhibits the expression of the *ara* operon.



#### (a) Operon inhibited in the absence of arabinose

# With arabinose, the araC protein activates transcription.



(b) Operon activated in the presence of arabinose

The base sequence of the *trp* operator. The nearly palindromic sequence is boxed and its –10 region is overscored.

# CGAACTAGTTAACTAGTAGGCAAGGCTTGATCAATTGATCATGCGTTC-20-10+1

## A genetic map of the *E*. *coli trp* operon indicating the enzymes it specifies and the reactions they catalyze.



#### Prokaryotic RNA Polymerase: Holoenzyme Enzyme

	Subunit	Size	#/Molec	<u>Function</u>
	ţα	36.5 kD	2	chain initiation and interaction with regulatory proteins
$\bigcirc$	β	151 kD	1	chain initiation and elongation
	-β'	155 kD	1	DNA binding
	σ	70 kD	1	promoter recognition

### E. coli RNA polymerase

#### $2\alpha$ , $1\beta$ , $1\beta$ ', $1\omega$ and $\sigma$ factor



#### The function of sigma factor

- the sigma subunit of RNA polymerase is an "initiation factor"
- there are several different sigma factors in E. coli that are specific for different sets of genes
- sigma factor functions to ensure that RNA polymerase binds stably to DNA only at promoters
- sigma destablizes nonspecific binding to non-promoter DNA, sigma stabilizes specific binding to promoter DNA, this accelerates the search for promoter DNA



#### The sigma cycle

- closed promoter complex (moderately stable)
- the sigma subunit binds to the -10 region

#### - RNA polymerase holoenzyme (+ $\sigma$ factor)

- open promoter complex (highly stable)
- the holoenzyme has very high affinity for
- promoter regions because of sigma factor

- once initiation takes place, RNA polymerase does not need very high affinity for the promoter
- sigma factor dissociates from the core polymerase after a few elongation reactions

• elongation takes place with the core RNA polymerase

• sigma can re-bind other core enzymes







#### Sítio de ligação ao Ribossomo



#### **Transcription initiation in prokaryotes**:

sigma factor binds to the -35 and -10 regions and then the RNA polymerase subunits bind and begin transcription

#### (b) Initiation (a) RNA polymerase binding to promoter σ70 β β α α DNA α β' σ70 ω (1) -10-35

#### **Mechanism of RNA synthesis**



- RNA synthesis usually initiated with ATP or GTP (the first nucleotide)
- RNA chains are synthesized in a 5' to 3' direction
- Termination of some transcripts makes use of the <u>Rho protein</u>, which is a termination factor that catalyzes the dissociation of the RNA and polymerase



# **RNA chain elongation by RNA** polymerase.



## A hypothetical strong (efficient) *E. coli* terminator.




## **Classes of eukaryotic cellular RNAs**

- ribosomal RNA (rRNA)
  - 18S (small subunit)
  - 28S (large subunit)
  - 5.8S (large subunit)
  - 5S (large subunit)
- transfer RNA (tRNA)
- messenger RNA (mRNA)
- heterogeneous nuclear RNA (hnRNA) (precursors of mRNA)
- small nuclear RNA (snRNA)

<u>U1, U2, U3, U4, U5, U6, U7, U8, U9, U10...</u>

• small cytoplasmic RNA (scRNA) 7SL RNA

What are the enzymes responsible for the synthesis of these RNAs?

## DNA dependent RNA polymerase

- Cells contain 3 DNA dependent RNA polymerases:
- RNA pol I: transcribes pre-rRNA; no known viral templates
- **RNA pol II:** transcribes pre-mRNA & snRNA: polymerase for most viral DNAs.
- **RNA pol III:** transcribes pre-tRNAs, 55 rRNA, U6 snRNA; polymerase for some viral DNAs.

- The transcriptional machinery must:
  - Be directed to *initiate* transcription at the correct location on a DNA template (the *transcriptional start site*).
  - elongate through the entire gene
  - Be directed to *terminate* transcription at the correct location.

## All of these functions require the assistance of

- Cis-acting sequences along the DNA
- Trans-acting factors (accessory proteins)

## **The human RNA polymerases**

<b>Polymerase</b>	<b>Location</b>	<b>Product</b>
RNA polymerase I	nucleolus	18S, 28S, 5.8S rRNA
<b>RNA polymerase II</b>	nucleoplasm	hnRNA/mRNA, U1, U2, U4, U5 snRNA
RNA polymerase III	nucleoplasm	tRNA, 5S RNA, U6 snRNA, 7SL RNA
mitochondrial RNA polymerase	mitochondrion	all mitochondrial RNA

#### Sensitivity of the nuclear RNA polymerases to $\alpha$ -amanitin<sup>1</sup>

RNA pol I	resistant
RNA pol II	high sensitivity (binds with $K = 10^{-8} M$ )
RNA pol III	low sensitivity (binds with $K = 10^{-6} M$ )

<sup>1</sup> cyclic octapeptide from the poisonous mushroom *Amanita phalloides* 



## **Transcription and promoter elements for RNA polymerase II**



Promoter (DNA sequence upstream of a gene)

- determines start site (+1) for transcription initiation
- located immediately upstream of the start site
- allows basal (low level) transcription

Transcription element (DNA sequence that regulates the gene)

- determines frequency or efficiency of transcription
- located upstream, downstream, or within genes
- can be very close to or thousands of base pairs from a gene
- includes

enhancers (increase transcription rate)

silencers (decrease transcription rate)

- response elements (target sequences for signaling molecules)
- genes can have numerous transcription elements

# Transcription by RNA Pol. II.

- At least 40 proteins required: Pol. II itself + accessory proteins.
- Accurate transcription initiated at the promoter.
- Promoter + additional DNA sequence that controls transcription = Transcriptional control region (TCR).
  - The adenovirus type 2 major late promoter was the first TCR ever recapitulated *in vitro*.
- Initiation is a multistep process:
  - Promoter *recognition* by RNA Pol. II
  - Formation of open initiation complex (unwinding)
  - Promoter *clearance*
  - 3' movement of complex away from promoter

# Regulation of Pol. II transcription

- Transcription must be regulated: genes must be turned on and off in temporal patterns
- Viral gene expression: early and late genes
- Transcriptional regulation is controlled by:
  - Cis-acting sequences in DNA both local and distal
  - Trans-acting factors both protein and RNA
- Trans-acting factors specifically bind to cisacting sequences to either
  - activators stimulate transcription
  - repressors prevent transcription

### **Transcription and promoter elements for RNA polymerase II**



## Sequence elements within a typical eukaryotic gene<sup>1</sup>

<sup>1</sup> based on the thymidine kinase gene



#### TATA box (<u>TATA</u>AAA)

- located approximately 25-30 bp upstream of the +1 start site
- determines the exact start site (not in all promoters)
- binds the TATA binding protein (TBP) which is a subunit of TFIID

#### GC box (CC<u>GC</u>CC)

- binds Sp1 (Specificity factor 1)
- CAAT box (GGC<u>CAAT</u>CT)
  - binds CTF (CAAT box transcription factor)

Octamer (ATTTGCAT)

• binds OTF (Octamer transcription factor)

#### **Proteins regulating eukaryotic mRNA synthesis**

General transcription factors

- TFIID (a multisubunit protein) binds to the TATA box to begin the assembly of the transcription apparatus
  - the TATA binding protein (TBP) directly binds the TATA box
  - TBP associated factors (TAFs) bind to TBP
- TFIIA, TFIIB, TFIIE, TFIIF, <u>TFIIH<sup>1</sup></u>, TFIIJ assemble with TFIID

RNA polymerase II binds the promoter region via the TFII's

Transcription factors binding to other **promoter elements** and **transcription elements** interact with proteins at the promoter and further stabilize (or inhibit) formation of a functional preinitiation complex

<sup>1</sup>TFIIH is also involved in phosphorylation of RNA polymerase II.

## TRANSCRIÇÃO Formação do complexo funcional de preiniciação



Inicio de Transcrição

#### **Binding of the general transcription factors**



• TFIID (a multisubunit protein) binds to the TATA box

to begin the assembly of the transcription apparatus

- the TATA binding protein (TBP) directly binds the TATA box
- TBP associated factors (TAFs) bind to TBP
- TFIIA, TFIIB, TFIIE, TFIIF, TFIIH, TFIIJ assemble with TFIID

#### **Binding of RNA polymerase II**



- RNA polymerase II (a multisubunit protein) binds to the promoter region by interacting with the TFII's
- TFs recruit histone acetylase to the promoter



• this process is called "transactivation"

#### **Formation of a stable preinitiation complex**



- the stability and frequency with which complexes are formed determines the rate of initation of transcription
- the rate of initiation of transcription is of major importance in determining the abundance of an mRNA species

#### **Initiation of transcription and promoter clearance**



• RNA pol II is **phosphorylated** by TFIIH on the carboxy terminal domain (CTD), releasing it from the preinitiation complex and allowing it to initiate RNA synthesis and move down the gene

#### **RNA** polymerase and associated proteins to start transcription.



## **Transcription factors (partial list)**

Factor	Full name or function
CREB	Cyclic AMP response element binding protein
CTF	CAAT box transcription factor (=NF1) (binds $GGCCAATCT$ )
NF1	Nuclear factor-1 (=CTF)
AP1	Activator protein-1 (dimer of the Fos-Jun proteins)
Sp1	Specificity factor-1 (binds CCGCCC)
OTF	Octamer transcription factor (binds ATTTGCAT)
NF-κB	Nuclear factor kB
HSTF	Heat shock transcription factor
MTF	Metal transcription factor
USF	Upstream factor
ATF	Activating transcription factor
HNF4	Hepatocyte nuclear factor-4 (nuclear receptor superfamily)
GR	Glucocorticoid receptor (nuclear receptor superfamily)
AR	Androgen receptor (nuclear receptor superfamily)
ER	Estrogen receptor (nuclear receptor superfamily)
TR	Thyroid hormone receptor (nuclear receptor superfamily)
C/EBP	CAAT/enhancer binding protein
E2F	E2 factor (named for the adenovirus E2 gene)
p53	p53 (tumor suppressor protein)
Мус	Product of the c-myc protooncogene (dimerizes with Max)

Putativos Sítios de ligação de fatores de Transcrição e TATA Box em promotores de genes de celulases e xilanases de *Trichoderma reesei* 



## **Promotor para RNA pol II**



(b) Core promoter elements for RNA polymerase II

## **Promotor para RNA pol I**



(a) Promoter for RNA polymerase I

## **Promotor para RNA pol III**



(c) Two types of promoters for RNA polymerase III

#### Assembly of the pre-initiation complex (PIC)



- binding of upstream binding factor (UBF) to the upstream control elements (UCEs) and core element of the rDNA promoter, leading to the recruitment of the SL-1 initiation factor, which contains TATA-box-binding protein (TBP) and at least five TATA-box-associated factors (TAFs). The resultant stable UBF–SL-1 complex recruits an initiation-competent form of RNA Pol I, which contains RRN3 that mediates interactions between RNA Pol I and SL-1

- RNA Pol III-transcribed genes (for example, those that encode tRNAs) have internal promoters that comprise two sequence blocks (A and B) that are located in the transcribed region. The A and B blocks are recognized by TFIIIC that recruits TFIIIB, which is composed of the subunits B-related factor 1 (BRF1), BDP1 and TBP. Finally, TFIIIB recruits RNA Pol III.

- For 5S rDNA promoters the B block is replaced by a sequence, termed block C, to which TFIIIA binds and recruits and orientates TFIIIB, following which transcription initiation proceeds as for tRNA genes.

- For a small number of RNA Pol III-transcribed genes (for example, U6 snRNA (RNU6-1)) the promoters are located upstream of the gene and contain TATA boxes bound by TBP, and proximal sequence elements (PSEs) bound by a complex called small nuclear RNA-activating protein complex (SNAPC). These upstream promoters are bound by a different form of TFIIIB from tRNA genes, which is composed of BRF2, BDP1 and TBP9

#### tRNA- This RNA type is a small chain of about 80 nucleotides.



The tRNA has a clover leaf model with arms each with a specific function. The tRNA also has an anticodon region that can base pair with the codon region on the mRNA.



#### Pre-Processamento de rRNA



#### Pre-Processamento de rRNA em células humanas





**rRNA-** The rRNA is synthesized in the nucleolus. In the cytoplasm, ribosomal RNA and protein combine together to form a nucleoprotein called a ribosomes. The ribosomal RNAs form two subunits namely; the large subunit and small subunit. The Eukaryotic cells have 4 different types of rRNA namely; 28S rRNA, 18S rRNA, 5.8S rRNA and 5S rRNA.



Svedberg value = sedimentation coefficient, a measure of time (10<sup>-13</sup> sec)

# mRNA Structure

#### **Bacterial mRNA only**

Ribosome-binding sites; also called Shine-Dalgarno sequence
Eukaryotic mRNA only

- 5' Cap (methylated guanine)
- Poly A tail
- Kozak Sequence (sometimes present); enhances ribosome binding



#### a) Bacterial mRNA

#### (b) Eukaryotic mRNA

© 2012 Pearson Education, Inc.

## **RNA polII - Structure of eukaryotic mRNA**



- all mRNAs have a 5' cap and all mRNAs (with the exception of the histone mRNAs) contain a poly(A) tail
- the 5' cap and 3' poly(A) tail prevent mRNA degradation
- loss of the cap and poly(A) tail results in mRNA degradation





© 2012 Pearson Education. Inc.



# Post-transcriptional processing of eukaryotic mRNAs.



## **Types of Introns.**

Intron Type	Where Found	
GU-AG introns	Eukaryotic nuclear pre-mRNA	
AU–AC introns	Eukaryotic nuclear pre-mRNA	
Group I	Eukaryotic nuclear pre-mRNA, organelle	
	RNAs, a few bacterial RNAs	
Group II	Organelle RNAs, a few prokaryotic RNAs	
Group III	Organelle RNAs	
Twintrons (composites of two and/or more group II or III introns)	Organelle RNAs	
Pre-tRNA introns	Eukaryotic nuclear pre-tRNAs	
Archaeal introns	Various RNAs	

The consensus sequence at the exon–intron junctions of vertebrate pre-mRNAs.



Numbers = Frequency (%)

## **Splicing Alternativo**



#### The organization of the rat $\alpha$ -tropomyosin gene and the seven alternative splicing pathways that give rise to cell-specific $\alpha$ tropomyosin variants.





## Expressão gênica tecido específica







# Obrigado

fscha@usp.br

USP – 1° Semestre 2021