

PLNT 2530 2024

Unit 4 **NUCLEAR GENE STRUCTURE**

Plant Biotechnology by A Slater, N Scott, M Fowler
Analysis of genes and Genomes by Richard J Reece
Plant Biotechnology and Genetics by C Neal Stewart JR

Rationale

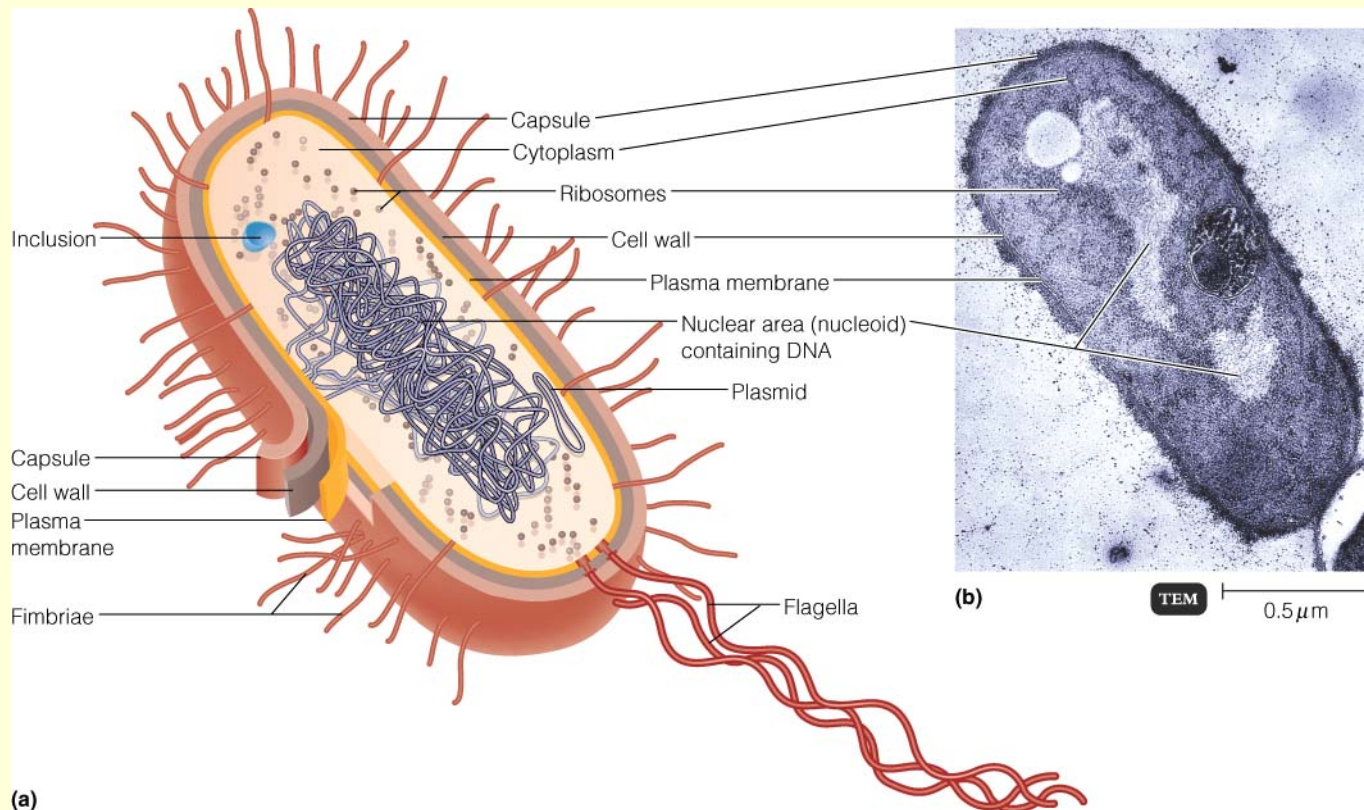
The properties of plants are in part derived from their genes, and the genomes which express those genes. If you want to make the plant do something, you have to understand the genome that underlies the biology of the plant.

Prokaryotes: Bacteria and Archea

All functions in a single cellular compartment

Membrane functions on plasma membrane

Transcription, translation and replication can all occur simultaneously

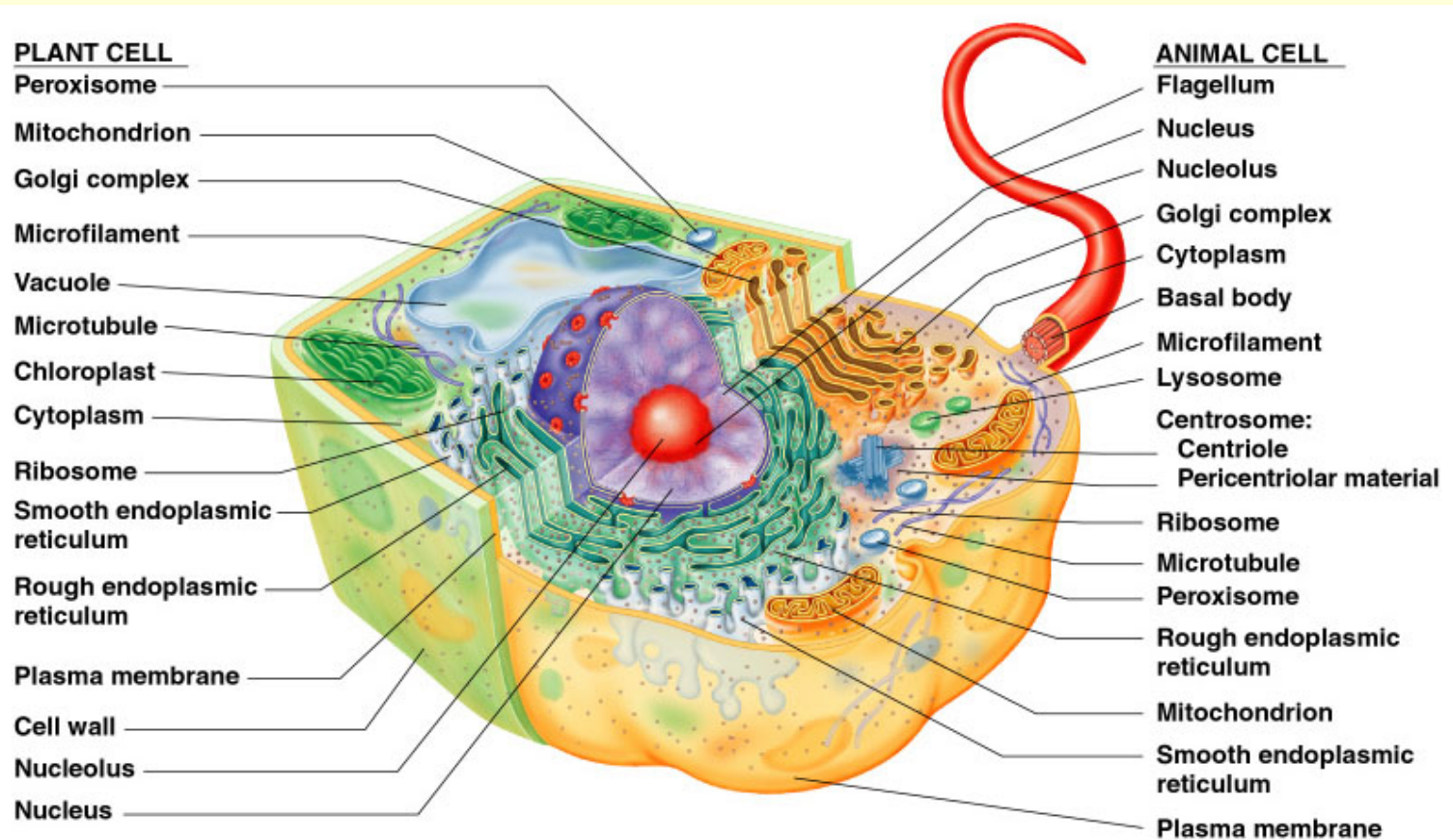


Eukaryotes: Protists, Plants, Fungi, Animals

Cellular functions compartmentalized in membranous organelles

Transcription in nucleus; **translation** in cytoplasm

Interphase: Gene expression; **Mitosis**: DNA replication and cell division



(a) Highly schematic diagram of a composite eukaryotic cell, half plant and half animal

Copyright © 2004 Pearson Education, Inc., publishing as Benjamin Cummings.

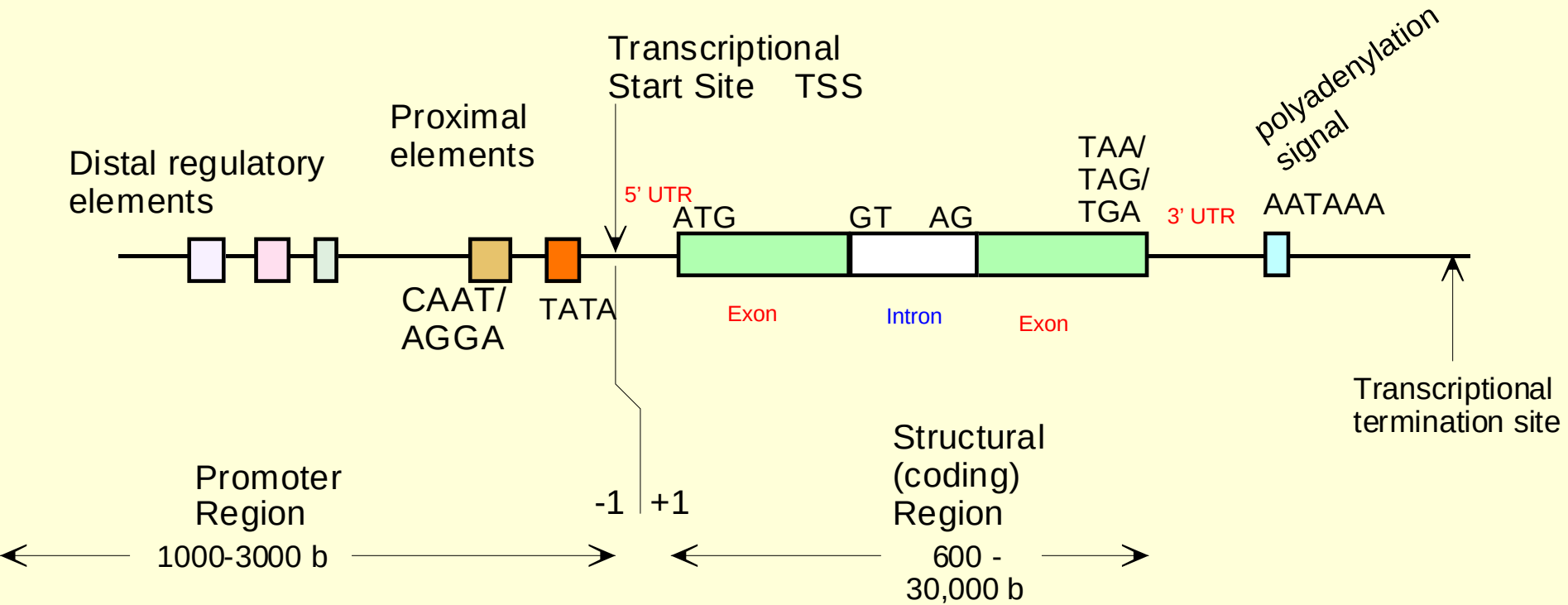
Eukaryotic genes

- 1) ribosomal RNA (rRNA)** - 4 ribosomal RNA genes from 2 transcripts
- 2) transfer RNA (tRNA)** - carry amino acids that are incorporated into proteins during translation.
- 3) messenger RNA (mRNA)** - translated into proteins
- 4) heterogeneous nuclear RNA (hnRNA)** - an umbrella term that encompasses a wide variety of non-coding RNAs that carry out diverse functions, many involved in gene expression or processing of RNA.

RNA polymerase	location	Products
I	Nucleolus	28S rRNA, 18S rRNA and 5.8S rRNA
II	Nucleus	mRNA
III	Nucleus	tRNA, 5S rRNA

rRNA - requires transcription of the gene into RNA by RNA polymerase I (Pre-rRNA 18S, 5.8S, 28S genes) and RNA polymerase III for 5S RNA gene.

Elements of a nuclear gene



Gene structure

Terminology: **downstream** -in the 3' direction
 upstream -in the 5' direction

Regulatory region of gene: is promoter present in the upstream region.
-Can be proximal and distal regions.

Proximal region responsible for the basal transcription and tissue specificity

This region consists of:

i) **Transcription Start Site** (-1/+1)-gene location where the first ribonucleotide of the RNA is being synthesized (+1 first base of the transcript). (**This is NOT the same as the start codon!**)

ii) **TATA box** ~(-25 to -30) in eukaryotes called because of the prominence of A and T's. RNA polymerase and general TF's bind to TATA and form pre-initiation complex.

iii) Other sequences, '**CAAT**' **box** (-70 to -80 bases), not required for transcription but influence the level, rate, timing or tissue specificity of transcription.

Gene structure

Distal sequences modulate gene expression (amplify or suppress)

-influence developmentally and environmentally regulated genes and may determine tissue specificity

Enhancers are a class of distal sequence elements act in a position- and orientation-independent manner to elevate gene transcription

Can be upstream (100 b -5 kb), downstream or even in an intron of the gene

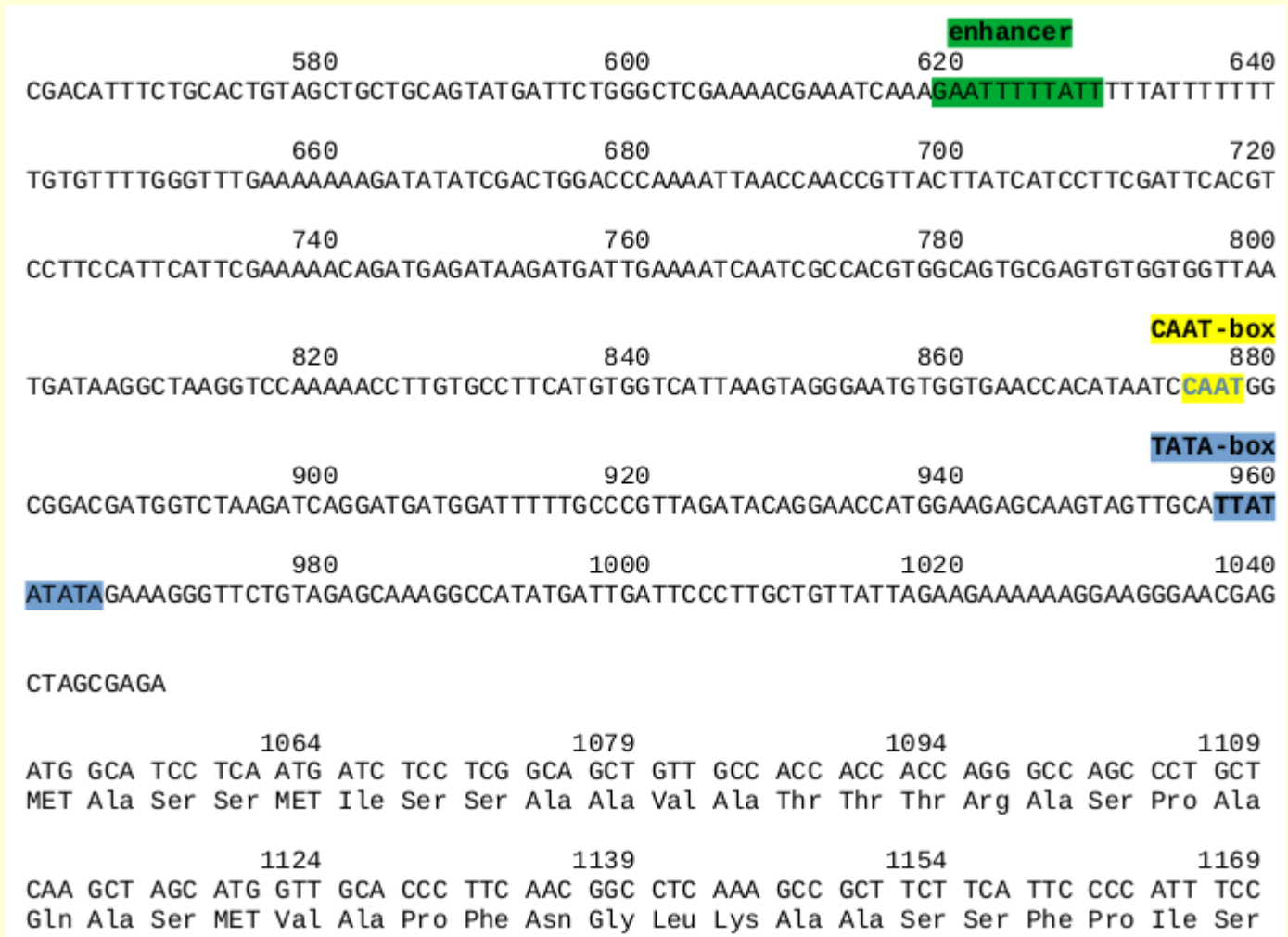
Gene structure

Example: RBCS1 gene (AJ419827) from *Coffea arabica*

RUBISCO small subunit gene. Only part of the gene is shown, beginning in the promoter region and ending in the first exon.

The transcription start site has not been determined experimentally. It would be somewhere between the TATA box and the start codon.

The enhancer (green) is light-regulated.



Gene structure

regulatory elements – sequences (known or unknown) that influence gene expression

- **cis-elements** - regulatory elements that are directly attached to the promoter region and on the same strand as the coding region
- **trans-acting factors** - normally protein transcription factors which interact with cis-elements to affect the transcription process. They are trans in the sense that they are created on genes far from the gene being affected

Protein encoding genes - transcribed from TSS and processed into mRNA

Features are related to:

i) transcription

ii) post-transcriptional modification of the RNA transcript (in the nucleus)

iii) export of mRNA to the cytoplasm

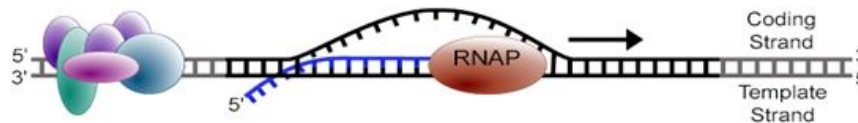
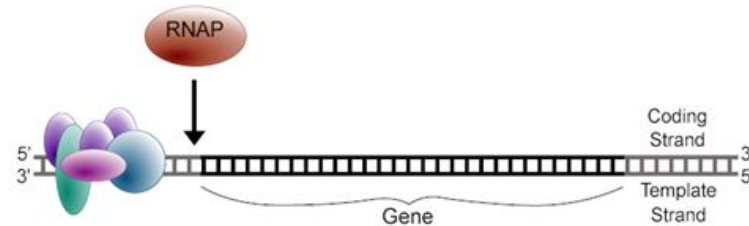
iv) translation of the mRNA into a polypeptide

Computer-generated video of Chromosomes,
DNA replication, Transcription, and Translation

<https://www.youtube.com/watch?v=4PKjF7OumYo>

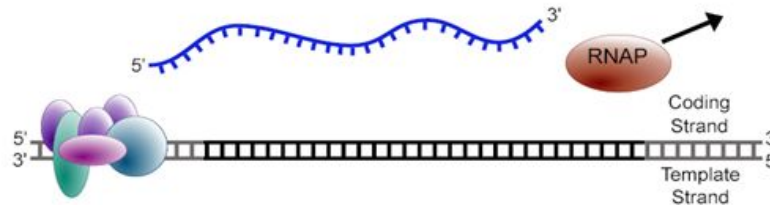
The Three Steps of Transcription

•Initiation



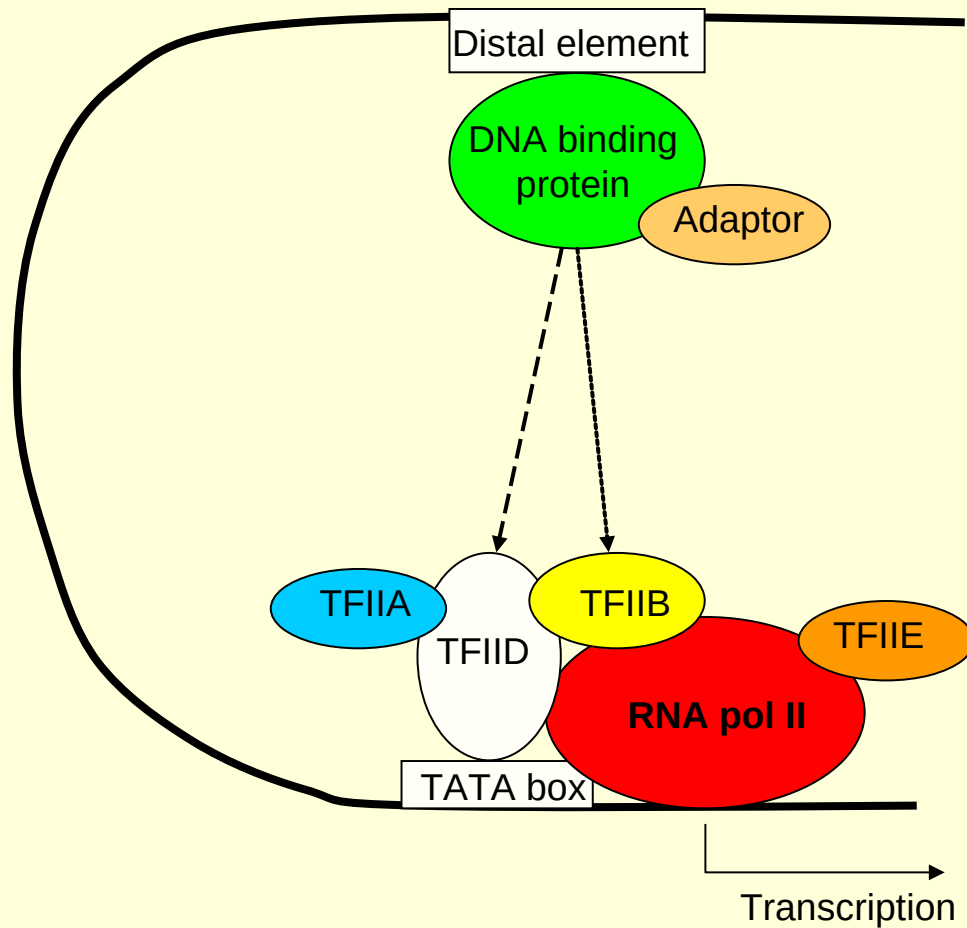
•Elongation

•Termination



RNAP - RNA polymerase

Transcription initiation: interactions of cis-elements and trans-acting DNA binding proteins



RNA processing

- The **primary transcript** is not the mRNA found in the cytoplasm but a product called or **pre mRNA**
 - Plant primary transcripts typically 1-3 kb
 - Mammalian transcripts can be much larger
 - Largest known (2,000 kb) Duchenne muscular dystrophy gene with more than 50 large introns, mRNA 18 kb
 - pre mRNA includes all introns
 - pre-mRNA is included in the broad category of heterogeneous nuclear RNA (hnRNA), which encompasses many types of RNAs in the nucleus.

Involves :

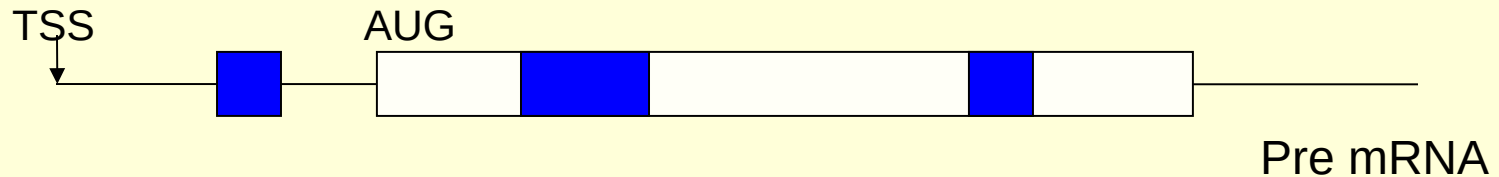
5'Capping

3' polyadenylation

Intron splicing

- **Terminology**

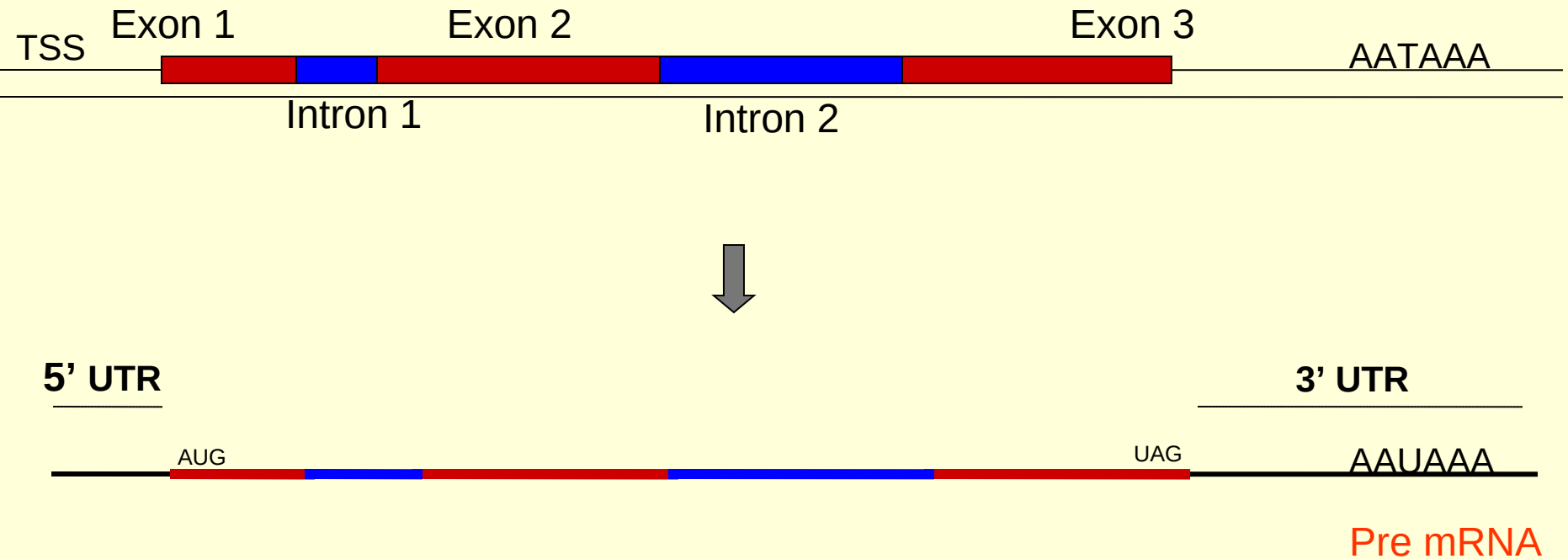
Intron is an intervening section of non-coding sequence which interrupts the coding/non-coding sequence of a gene and is removed during processing of the gene primary transcript.



Exon any stretch of sequence from the primary transcript (and related region of gene) which reaches the cytoplasm as part of the mRNA.

Gene structure

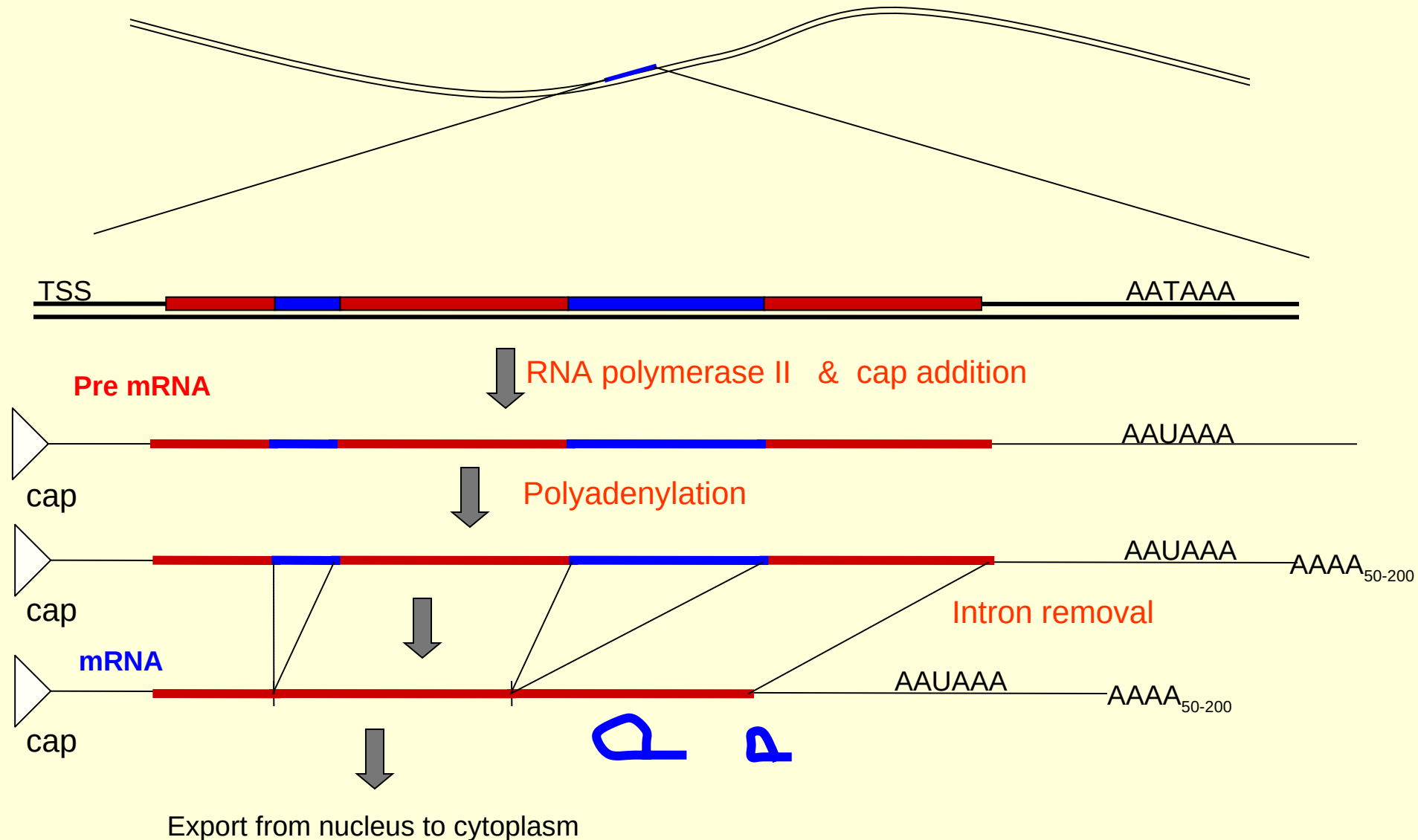
Transcription of gene with introns into pre mRNA



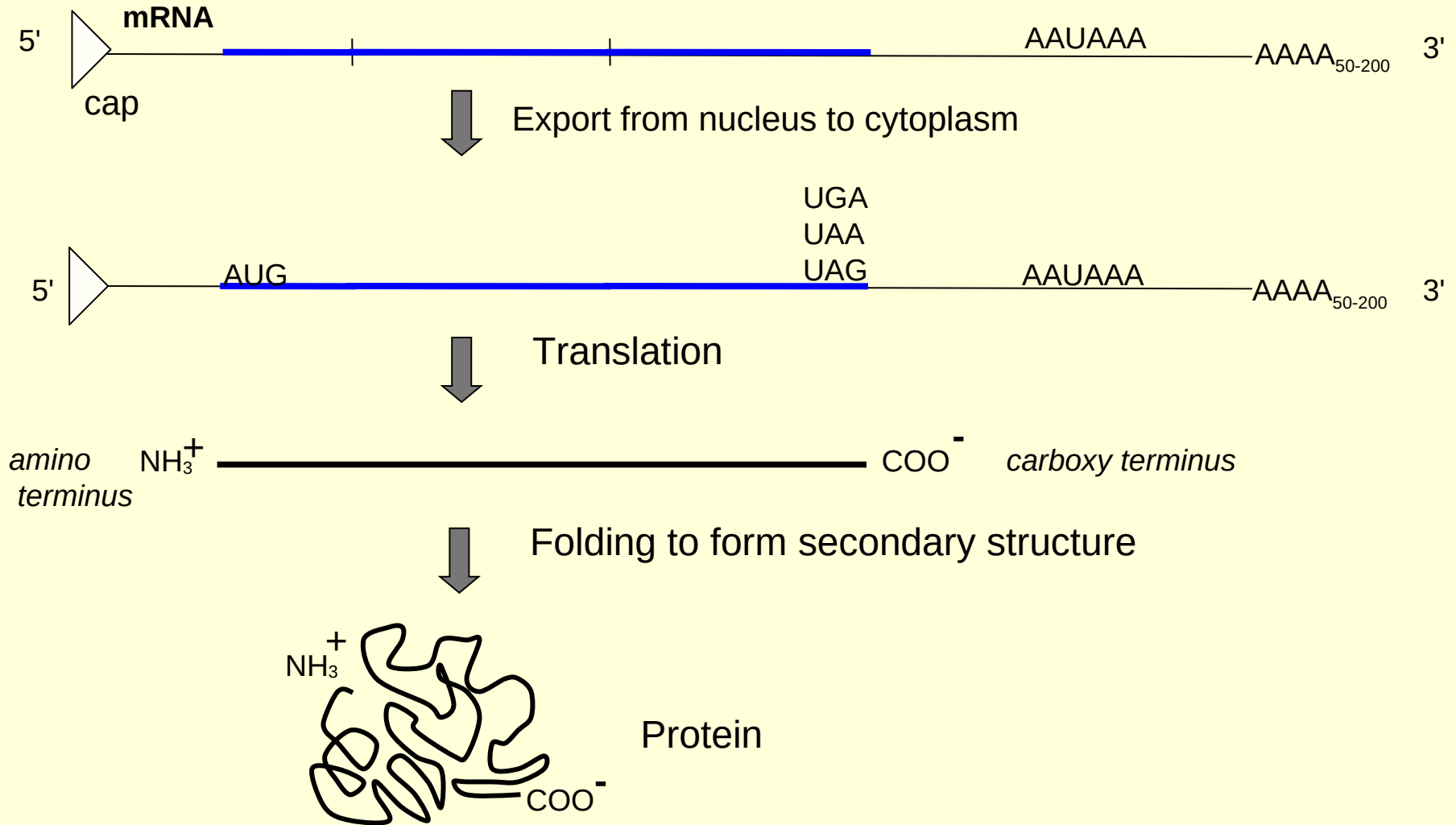
5' UTR - 5' untranslated region - Upstream of translation start site

3'UTR - 3' untranslated region - Downstream of translation stop site

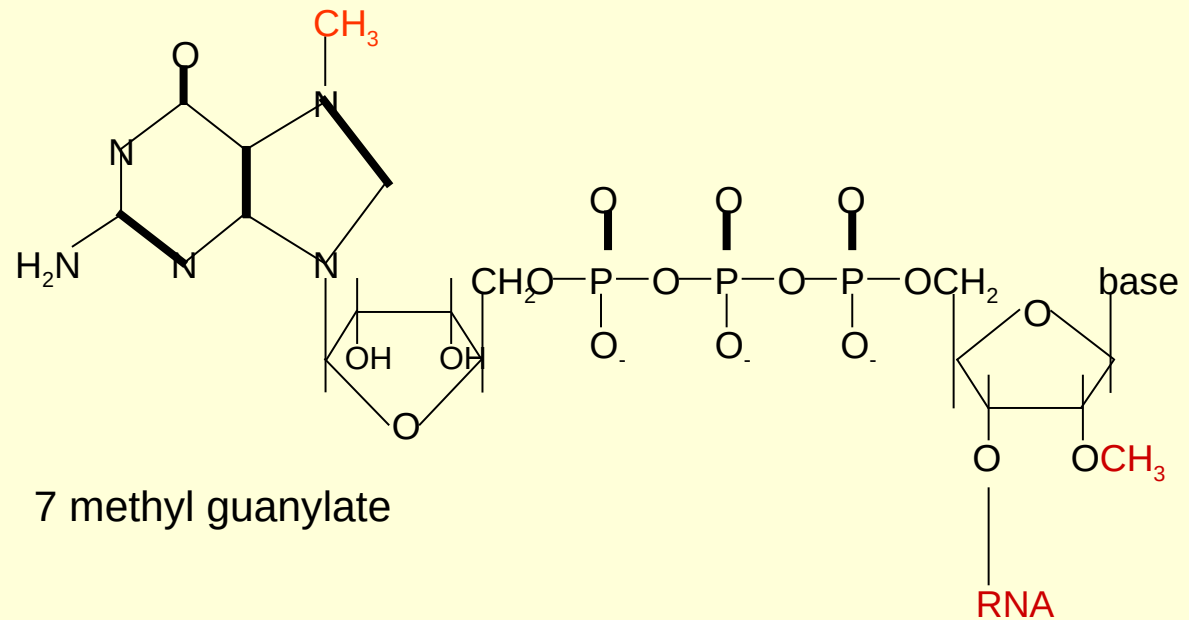
Transcriptional processing of a nuclear gene



Post-translational steps of gene expression



1. Cap addition



-occurs rapidly during transcription by addition of GTP in a 5'-5' manner

-methylation of the 2'-site of the second ribose occurs in all animals and higher plants

Cap Function: Recognition by cap binding protein which couples with 40S ribosomal subunit for formation of translation pre-initiation complex

2. Polyadenylation

- Cleavage of pre mRNA 10-30 bases downstream of AAUAAA signal by specific RNA endonuclease
- Addition of 20-250 A ribonucleotide residues by a template-independent RNA polymerase –**poly A polymerase**

Signal Consensus sequence	A	A	U	A	A	A
Frequency %Animals	97	98	100	100	100	97
Plants	90	100	100	87	91	65

Polyadenylation doesn't occur in rRNA, tRNA genes BUT occurs for all plant mRNAs except for histone genes

Function of Poly A tail:

- i. intron processing,
- ii. transport of mRNA to cytoplasm
- iii. protection of transcript in cytoplasm from exonuclease degradation

Technological use: Separation of mRNA from other RNA species

3. Intron splicing

- Several classes of introns with different removal mechanisms
- Only consider introns of nuclear, protein-encoding genes (plastid and mitochondrial protein genes are similar)
- Plant introns typically 70-1000 b long and rich in A and U's (mammalian introns are larger (to 100kb) and more abundant)
- Introns have characteristic boundary sequences

Intron-exon junction consensus sequence

	exon	intron		intron	exon
	--A G	G U A A G U	-----	Py n PY PY Pu A Py-----	N C A G
---Frequency %	62 77	100 100 60 74 84 50			78 100 100 55



Intron splicing (cont'd)



AUGCAUUCAGGACAGGUA-----CAGGAUGGGCCUACU

Correct splicing and translation

AUGCAUUCAGGACAGGUA-----CAGGAUGGGCCUACU
Met His Ser Gly Gln Asp Gly Pro Thr

Incorrect splicing and translation

AUGCAUUCAGGACAGGUA-----CAGGAUGGGCCUACU
Met His Ser Gly Gln Met Gly Leu Leu

Critical that introns be precisely removed **or** an incorrect protein is made

Translation and post-translation elements of genes

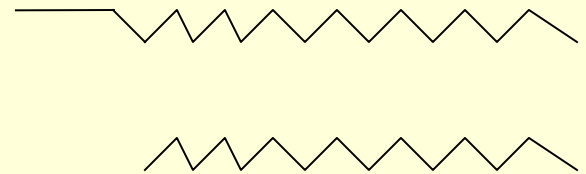
Protein processing into a biologically active form

conversion of mRNA into linear series of amino acids

Post-translational modifications

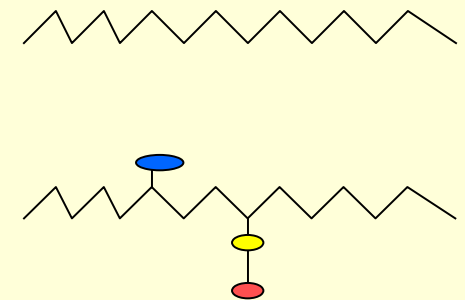
-Proteolytic processing

target (signal) sequence removal
protein cleavage



-modification of amino acids

hydroxylation (Pro)
phosphorylation (Ser, Thr, Tyr)
glycosylation (Ser, Thr, Asn)
acetylation (Lys)



Information encoded in gene

Targeting of proteins to a cellular destination

Nuclear-encoded proteins are translated in the cytoplasm but are targeted to an organelle in the cell based on signal sequences in the protein.

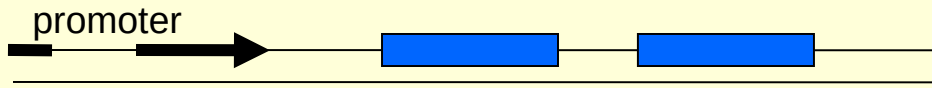
- a) endoplasmic reticulum
- b) vacuole
- c) nucleus
- d) chloroplast
- e) mitochondria
- f) export from the cell

internal sequences

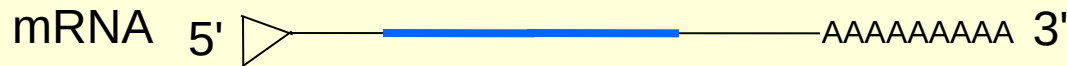
15-60 N-terminal amino acids which are cleaved off at the destination site

Computer programs exist which can predict localization site from the sequence

Summary

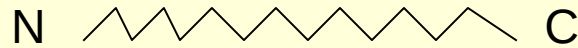


Flow for expression

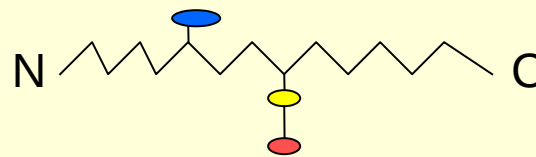


All the information is encoded in gene

Preprotein



Post-translational modification



Folded protein at organelle destination

