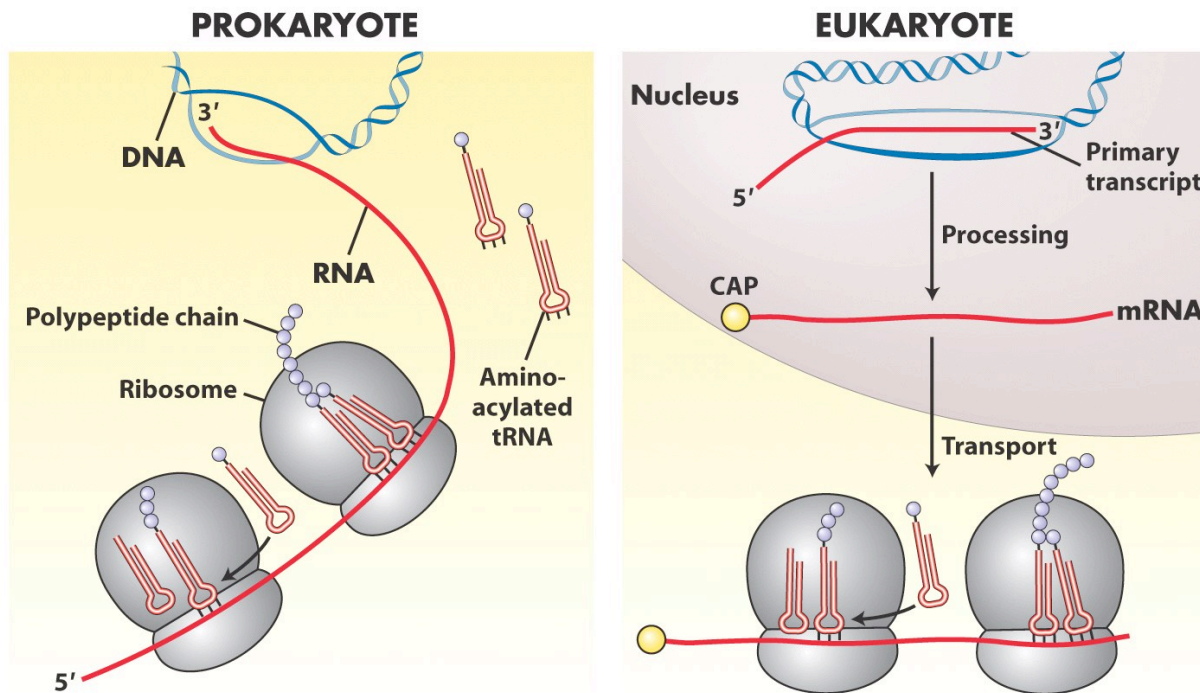


atgtcgact
100100100001
0100001110100
0100001110100

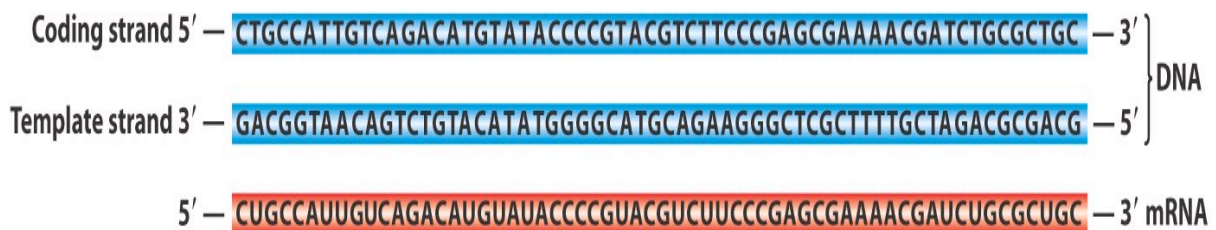
Transcription and Translation Algorithms

Transfer of genetic information from DNA to RNA to protein in prokaryotes and eukaryotes



Transcription and translation occur in separate compartments in the eukaryotic cell but in the same compartment in the prokaryotic cell. Otherwise the process of translation is similar in all organisms: An RNA called tRNA brings each amino acid to the ribosome, a molecular machine that joins amino acids into a chain according to the information provided by mRNA. Introduction to Genetic Analysis by Griffiths et al.

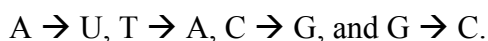
Transcription



Introduction to Genetic Analysis by Griffiths et al.

One strand of the DNA double helix is used as template (the lower strand from 3' to 5' in our example above) to synthesize RNA. RNA is complementary to the template strand.

So, given a template strand, the **algorithm** consists in changing every occurrence of :



Example: In the above example:

Template strand: 3' – GACGGTAACAGTC.....
mRNA 5' – CUGCCAUUGUCAG.....

Note that because the above DNA is from a prokaryote organism, no post-transcription processing (splicing of introns, capping, polyadenylation, etc...) are required to obtain the mature RNA (mRNA).

Translation

Assuming we have a mRNA coding sequence (CDS), translation consists of using the genetic code to translate (replace) every 3 bases (a codon) to one of the 20 amino acids. The process starts with AUG and ends with one of the 3 stop codons: UAA, UAG, and UGA. The genetic code is the dictionary for going from the world of bases (mRNA) to that of amino acids (protein). The genetic code is used as a lookup table.

So, given a CDS, the translation **algorithm** consists in substituting every 3 bases of the CDS in one of the 20 amino acids of the genetic code, starting with AUG and stopping at UAA, UAG, or UGA.

The Genetic Code

		Second letter				
		U	C	A	G	
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp	U C A G
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G
	A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G
						Third letter

[Introduction to Genetic Analysis by Griffiths et al.]

The amino acids can be represented by their full name, or 3-letter code (as in the genetic code above) or as is the usual case in bioinformatics, one-letter code.

Alanine	Ala	A	Leucine	Leu	L
Arginine	Arg	R	Lysine	Lys	K
Asparagine	Asn	N	Methionine	Met	M
Aspartic acid	Asp	D	Phenylalanine	Phe	F
Cysteine	Cys	C	Proline	Pro	P
Glutamic acid	Glu	E	Serine	Ser	S
Glutamine	Gln	Q	Threonine	Thr	T
Glycine	Gly	G	Tryptophan	Trp	W
Histidine	His	H	Tyrosine	Tyr	Y
Isoleucine	Ile	I	Valine	Val	V

Example:

mRNA 5' --**AUGUACUAUGAGCGGGAAGUGGCAAACUAG**--3'

Protein: **M Y Y E R E V A N ***

Note that in the above example the mRNA sequence to be translated starts with AUG. The translation is straightforward. In general, if we do not know where the start codon (AUG) is we have to translate the sequence in all 6 reading frames: 3 reading frames from 5' to 3' and 3 from the reverse complement 3' to 5'.