Synthetic life

(continuation of „The molecular origins of life" SoSe 2018)

WiSe 2018/19

Zbigniew Pianowski

7 lectures (90 min. each) in English
Heidelberg: Mon. 13:30-15:00 INF 270, Seminarraum 335

1st lecture: 5th Nov. 2018 (HD)

Following lecture terms:
Heidelberg: 12.11, 19.11., 03.12., 10.12., 17.12., 07.01.2019.

NO LECTURE ON: 26. Nov., nor 24., 31. Dec!

The most current dates, handouts – on the website:
http://www.ioc.kit.edu/pianowski/ and by Moodle

Mailing list for changes and supplementary information

Overview of the course

Artificial genetic polymers and oligonucleotide analogues;

unnatural base pairing – expansion of the genetic alphabet;

artificial ribozymes for efficient catalysis and recognition (SELEX, DNAzymes, foldamers);

biosynthetic incorporation of **unnatural aminoacids (UAAs)** into proteins;

enzyme engineering – production of enzymes with unknown or unnatural properties, *ab initio* protein design, directed evolution, theozymes;

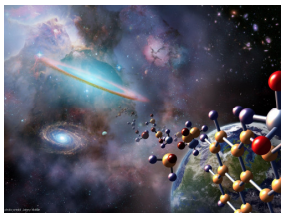
Artificial lipid vesicles as models for protocell multiplication;

synthetic biological circuits – riboswitches, time-delay circuits, oscilators, optogenetics;

design of artificial organisms – minimal genome project, **Synthia** – fully artificial genome resulting in living bacterial species

The molecular origins of life

Life is a self-replicating chemical system capable of evolution (NASA, 2009)



Origin of the Universe – stars, planets, elements

Origin of biorelevant monomers – primordial soup

Complex chemical processes on the way to living systems

Protocells and LUCA

What is Life? What makes it different from just matter?

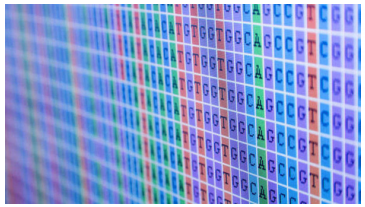
Everything – living or not – is just chemicals made of atoms.

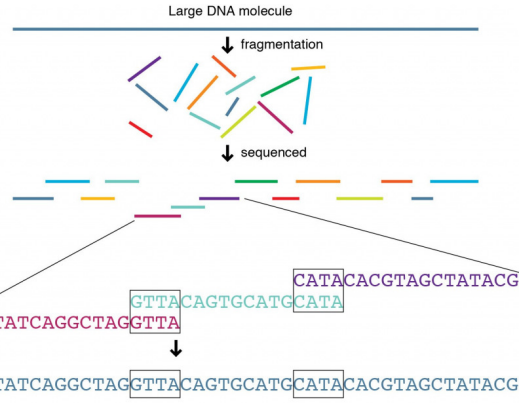
Every living creature has its code, that makes it grow, reproduce, and change.

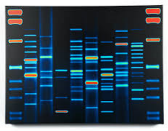
DNA turns dust into life.

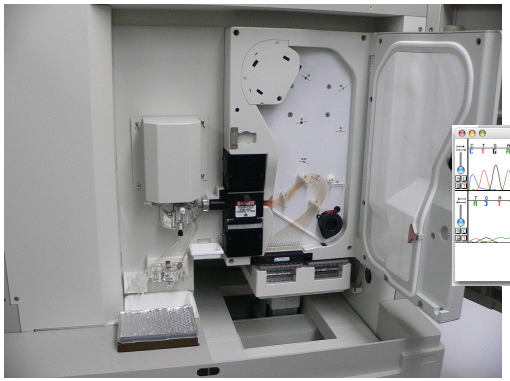
Fishes swim in water. But what makes fishes alive and not water is the way how the atoms are organized – By the special kind of molecules: DNA – the double helix molecule that houses the genetic alphabet of A, C, G and T, which, in different combinations, can make a flower, or a frog, or you...

20 years ago, Scientists learned to read the creatures' entire DNA sequence, from beginning to the end – the genome





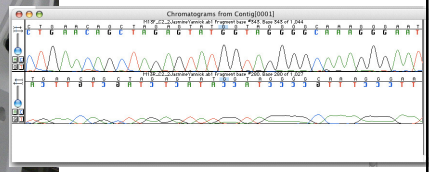




Mark Pellegrini

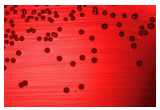
An ABI PRISM 3100 Genetic Analyzer. Such capillary sequencers automated the early efforts of sequencing genomes.

Electropherograms are commonly used to sequence portions of genomes

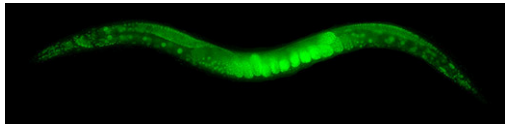


Tom David

Whole genome sequencing was initially achieved for simple organisms: bacteria, nematodes, flies and plants...



Haemophilius influenzae
1995

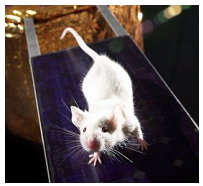


Caenorhabdis elegans
1998

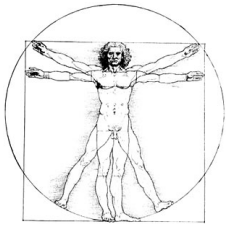
Drosophila melanogaster
2000

Arabidopsis thaliana
2000

... and way up to mammals and human



Mus musculus
2002



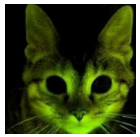
Homo sapiens
2004
Human Genome Project (NIH)
Craig Venter – Celera Genomcs (private)

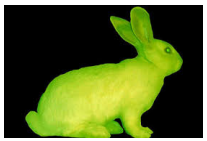
With that knowledge, scientists begun to tinker...

... to take a glow from a jellyfish...

... and transfer it to a cat...

... or to a rabbit...





To make creatures do what they never did before.

As biologists got better in this, a new kind of science was born – synthetic biology

Definition: *Synthetic Biology*

(also known as *Synbio*, *Synthetic Genomics*, *Constructive Biology* or *Systems Biology*)

„the design and construction of new biological parts, devices and systems that do not exist in the natural world and also the redesign of existing biological systems to perform specific tasks"

Advances in nanoscale technologies – manipulation of matter at the level of atoms and molecules – are contributing to advances in synthetic biology.

What can we do with new tools of synthetic biology?

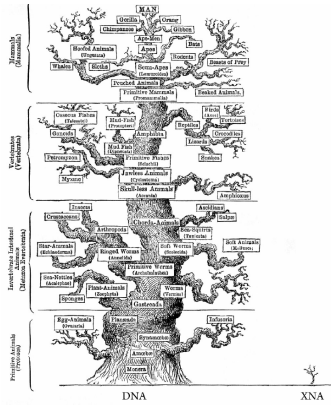
We can improve what was spelled out for the 3,5 Billion years of evolution.

We can take it beyond reading genomes or editing genomes...

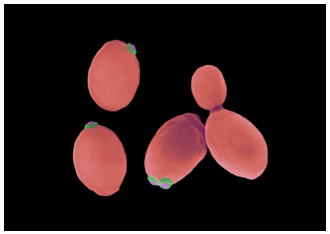
...and start writing genomes.
Our own ideas of
what life should be like.

Making creatures drastically different from any that have ever existed.

How could it be done?



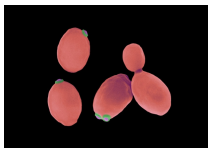
Already a group of scientists have re-written and rebuild the entire instruction kit for yeast. And they plan to put their recipe into a cell. And if everything goes according to the plan, It will come alive and begin making baby yeast exactelly like their most unusuall parent. And they will be new to the world.



Synthetic yeast project – „Synthetic yeast 2.0"

Synthetic biologists have been engineering chromosomes from scratch, sticking them into yeast and seeing whether the modified organisms can still function normally

Could yeasts could have evolved through alternate routes?
How much can you change a genome and still have a working organism?



Jef Boeke „Synthetic yeast 2.0"

Each of *S. cerevisiae*'s 16 chromosomes were assigned to teams of collaborators worldwide.

Each was to create a chromosome that was stable yet evolvable, and would keep yeast functioning as usual.

The teams used computer programs to design the codes of their respective chromosomes. They omitted some sequences found in naturally occurring yeast chromosomes, such as repetitive parts of the genome, in hopes of increasing the stability of the synthetic versions.

And they endowed their creations with a mechanism that mimics the random variation that drives evolution. When this scrambling system is triggered, it can shuffle, duplicate and delete genes at random.

This synthetic yeast will break the continuous chain of evolution that links every creature back to the first living cell.

It will be discontinuous in a way, a thing onto itself.
And then, where do we take this new technology?

The work that the consortium has already done could help to optimize the creation of microbes to pump out alcohol, drugs, fragrances and fuel. And it serves as a guide for future research on how genomes evolve and function.

Overview of the course

Artificial genetic polymers and oligonucleotide analogues;

unnatural base pairing – expansion of the genetic alphabet;

artificial ribozymes for efficient catalysis and recognition (SELEX, DNAzymes, foldamers);

biosynthetic incorporation of **unnatural aminoacids (UAAs)** into proteins;

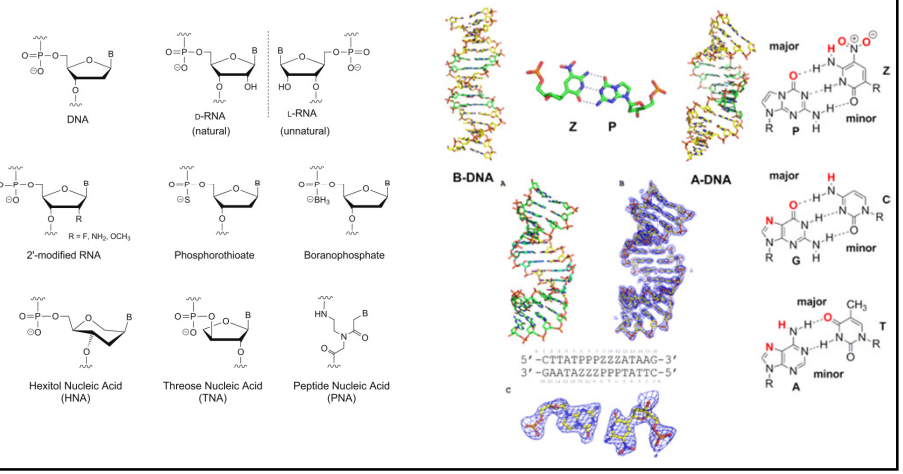
enzyme engineering – production of enzymes with unknown or unnatural properties, *ab initio* protein design, directed evolution, theozymes;

Artificial lipid vesicles as models for protocell multiplication;

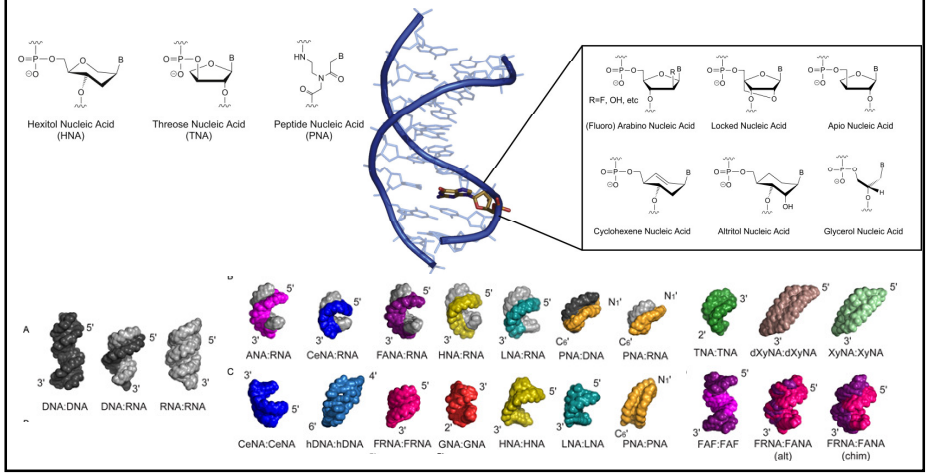
synthetic biological circuits – riboswitches, time-delay circuits, oscilators, optogenetics;

design of artificial organisms – minimal genome project, **Synthia** – fully artificial genome resulting in living bacterial species

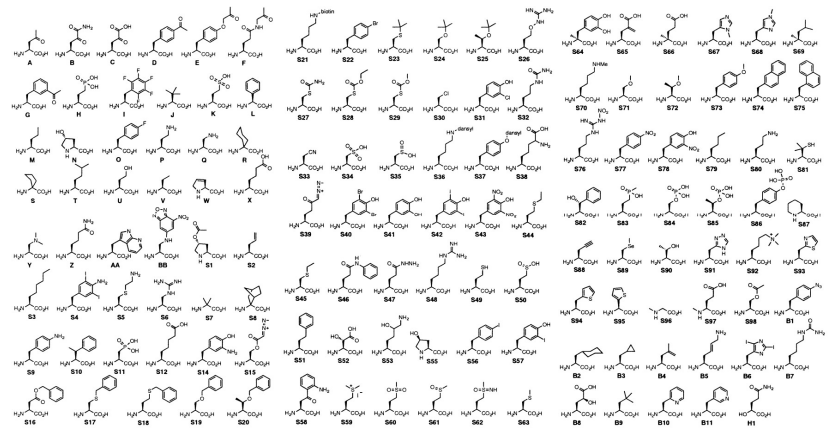
Artificial genetic polymers



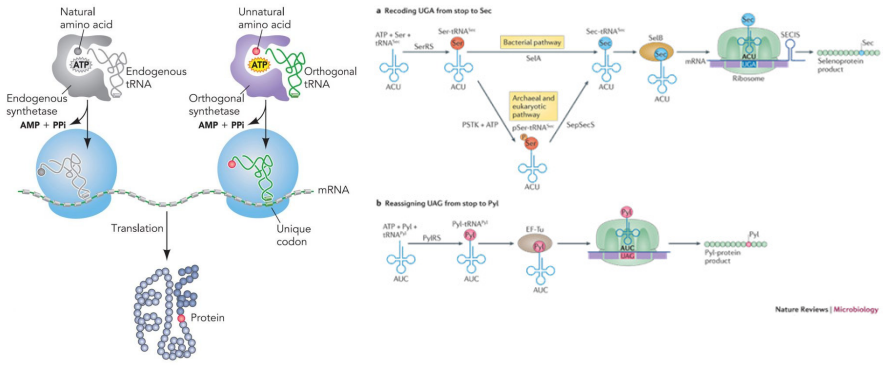
Artificial genetic polymers



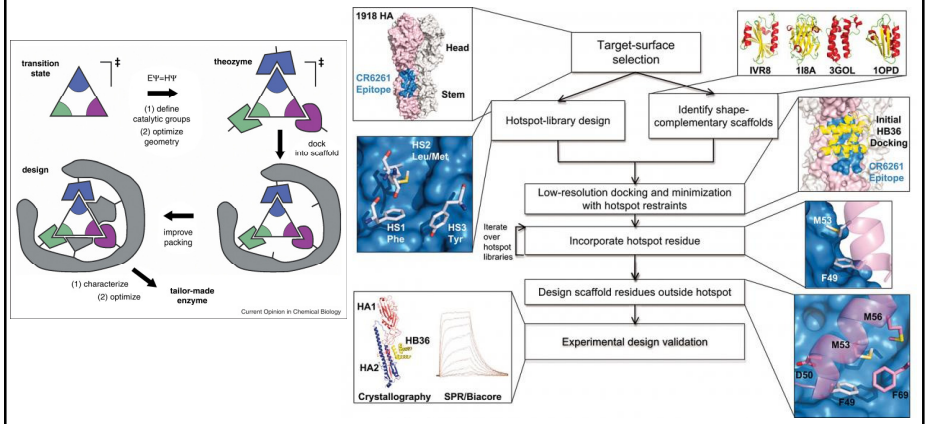
Expanded genetic code



Expanded genetic code



de novo enzyme design



Introduction

Basic bio-vocabulary
methods of molecular biology

Greek mythology – introduction to modern molecular biology – chimera, centaur

Introduction

The Central Dogma of the molecular biology – DNA → RNA → proteins

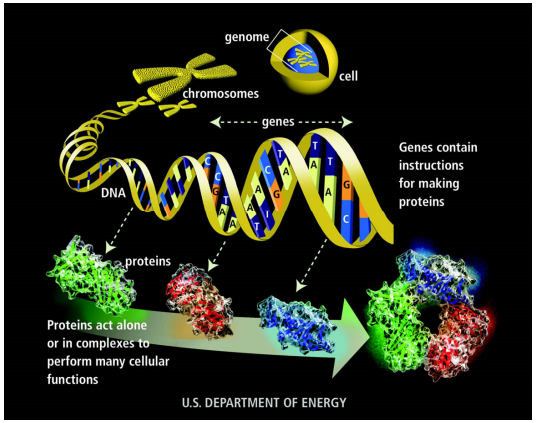
Polymerases and ribosomes - the molecular machines of life

PCR – Polymerase chain reaction – *in vitro* DNA amplification

Recombinant protein production – how to produce a protein in another organism

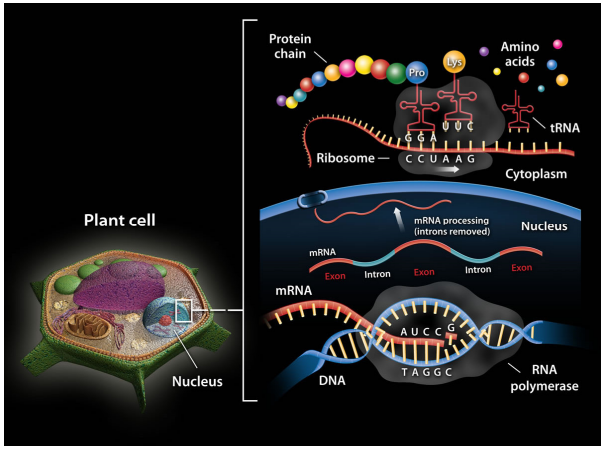
Protein engineering – how to make desired modifications in proteins

From DNA to proteins

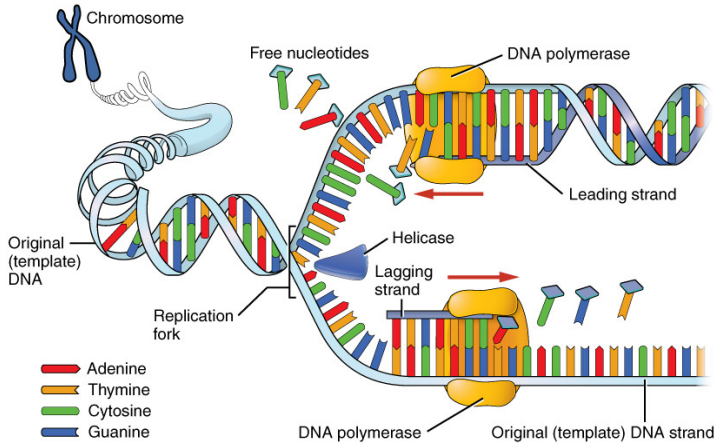


https://www.youtube.com/watch?v=gG7uCskUOrA

From DNA to proteins

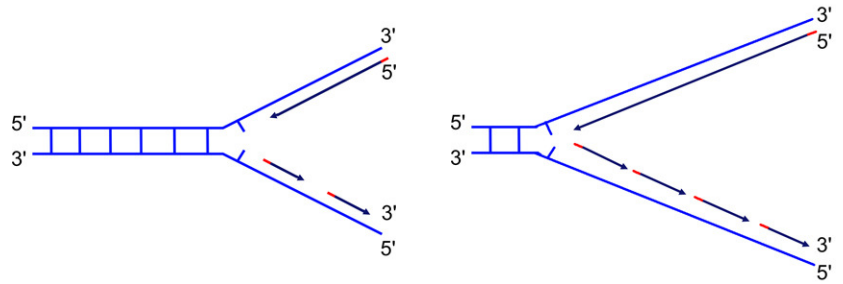


From DNA to proteins



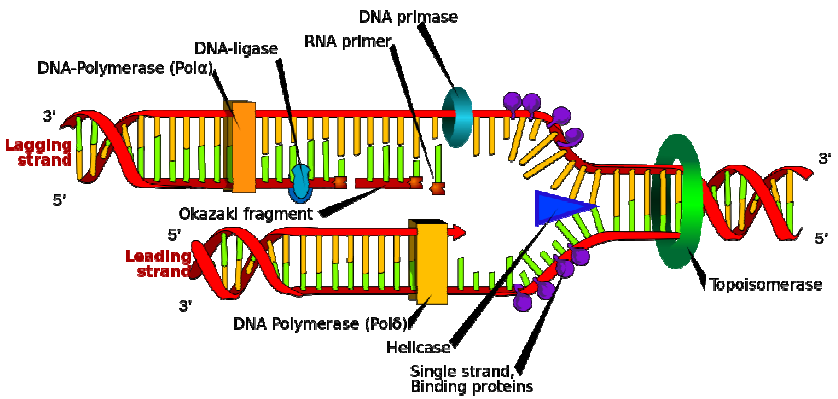
From DNA to proteins

Because DNA can only be synthesized from the 5' to 3' direction, the two newly-synthesized strands are synthesized differently. They are called the "continuous" (on the top in the figure) and "discontinuous" (on the bottom in the figure) strand, and also the "leading" (top) and "lagging" (bottom) strand.



As the replication fork moves down the helix, you can see the leading strand synthesized continuously, while the lagging strand must initiate new sites of synthesis.

From DNA to proteins



From DNA to proteins

Prokaryotic DNA-polymerases

Polymerase	Polymerase activity (for all enzymes 5′ → 3′)	Exonuclease activity
DNA polymerase I	Filling if gap after removal RNA primer, DNA repair, removal of RNA primers	5′→3′ and 3′→5′
DNA polymerase II	DNA repair	3′→5′
DNA polymerase III*	Replication, proofreading and editing	3′→5′

*The main enzyme of replication

70

Eukaryotic enzymes:

Five common DNA polymerases from mammals.

1. Polymerase α (alpha): nuclear, DNA replication, no proofreading
2. Polymerase β (beta): nuclear, DNA repair, no proofreading
3. Polymerase γ (gamma): mitochondria, DNA replication, proofreading
4. Polymerase δ (delta): nuclear, DNA replication, proofreading
5. Polymerase ε (epsilon): nuclear, DNA repair, proofreading

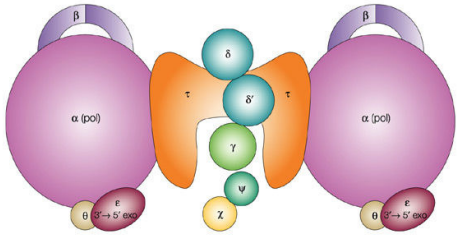
✓ Polymerases vary by species.

From DNA to proteins

Subunits of the DNA Polymerase III Holoenzyme

Subunit	Function	Groupings
α ε θ	5′–3′ polymerization 3′–5′ exonuclease core assembly	Core enzyme: Elongates polynucleotide chain and proofreads
γ δ δ′ χ ψ	Loads enzyme on template (Serves as clamp loader)	γ complex
β	Sliding clamp structure (processivity factor)	
τ	Dimerizes core complex	

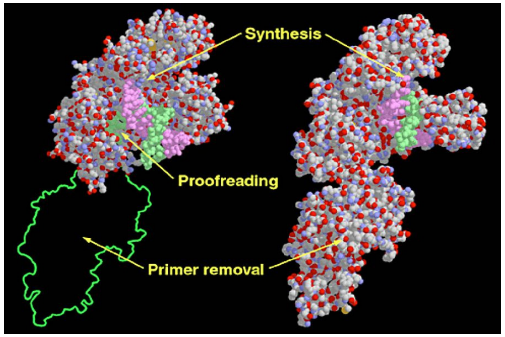
Procaryotic DNA Polymerase III



Nature Reviews | Molecular Cell Biology

From DNA to proteins

Procaryotic DNA Polymerase I



Transcription: DNA → RNA

Transcription requires the opening of the DNA helix, similar to replication

An RNA transcript is synthesized in the 5' to 3' direction, just like a new strand of DNA.

Transcription: DNA → RNA

Here the RNA transcript continues elongation until the transcript is complete

There is a signal in the DNA sequence that indicates a transcription start site and a transcription termination site. The transcription machinery must be able to read these signals, which implies that proteins can read the sequence of DNA

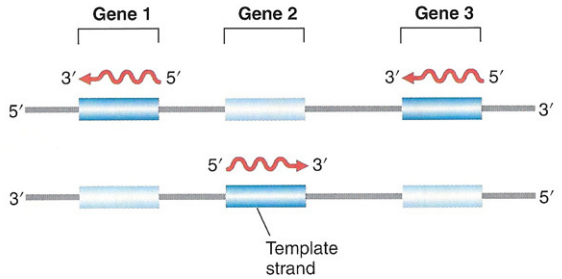
Transcription: DNA → RNA

It is possible to transcribe the other strand in this example

Here the RNA transcript continues elongation until the transcript is complete. Only one strand of any gene is transcribed, so the transcription initiation and termination signals must be specific to one strand

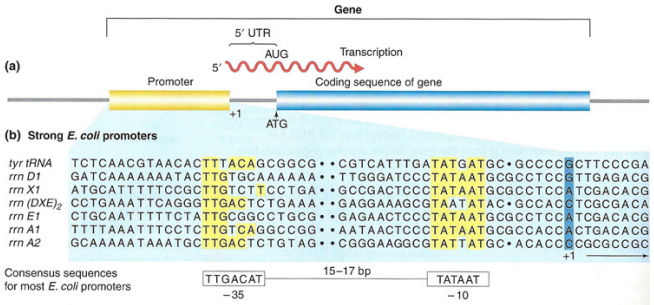
Transcription: DNA → RNA

Because transcription start and stop signals are specific to one strand, different genes on the same chromosome can be oriented in different directions. Only one particular strand is ever transcribed for each gene



Transcription: DNA → RNA

How do the start/stop signals look like? For procaryotes (like bacteria *E. coli*):



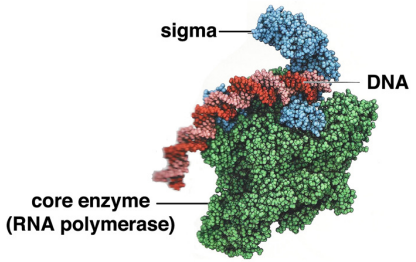
the transcription start site does not exactly correspond to the codon that is used to initiate translation – base 1 (+1). Transcription begins upstream of this sequence. The RNA between the „base 1" and the start of translation some distance downstream is called the 5' untranslated region or 5' UTR. There is also an untranslated region at the 3' end (3' UTR).

There are two conserved elements in promoter regions of *E. coli*: an element around -10 and an element around -35. These elements can be identified in most *E. coli* promoters.

Transcription: DNA → RNA

How does the transcription machinery recognize these sequences?

RNA polymerase itself does not carry out recognition of the promoter. In *E. coli*, there is a separate protein called sigma factor that specifically recognizes the promoter. Sigma factor and RNA polymerase form a complex that initiates transcription at the promoter. Once RNA elongation has begun, sigma factor dissociates from the complex.



The RNA polymerase, sigma factor, and DNA in a complex.

Transcription: DNA → RNA

Initiation in eukaryotes is more complex.

Most eukaryotic promoters have a "TATA box" at position -30, and typically have other classes of promoter sequences that are shared by groups of genes.

Eukaryotes use a set of general transcription factors that bind to the promoter region, then recruit other protein factors including RNA polymerase.

Once the preinitiation complex is formed, RNA polymerase is phosphorylated and released from the complex to begin RNA strand elongation.

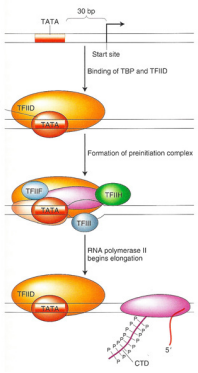
eukaryotic promoter

general transcription factors bind

TBP

preinitiation complex

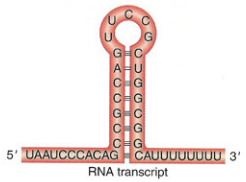
RNA pol II begins elongation



Transcription: DNA → RNA

In *E. coli*, there are two mechanisms for transcription termination, an intrinsic mechanism and a mechanism that depends on a specific protein called rho factor.

Bacterial transcription terminator



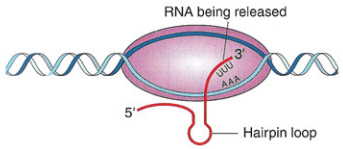
In the intrinsic mechanism, show above, there is a self-complementary sequence past the end of the coding sequence that forms a hairpin loop once it is transcribed. The base-paired part of the loop is very GC-rich, so the hairpin is stable. The presence of this structure interferes with RNA eleongation, and transcription terminates.

Transcription: DNA → RNA

In rho-dependent transcription termination, a specific sequence at the end of the gene binds rho factor. When RNA polymerase encounters rho, the polymerase dissociates from the template, terminating transcription.

Eukaryotic transcription termination is similar to that of prokaryotes (polyT, or recruiting CPSF/CTSF).

Bacterial termination: intrinsic mechanism

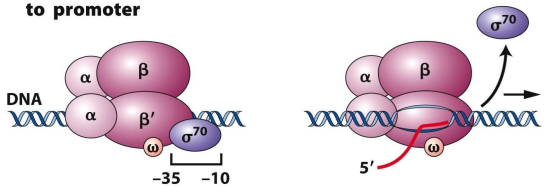


Transcription: DNA → RNA

RNA Polymerase of prokaryotes

Subunit	Function
α,α	Determine the DNA to be transcribed
β	Catalyze polymerization
β′	Bind & open DNA template(unwinding)
ω	Function is not known
σ	Recognize the initiation sites called promoter

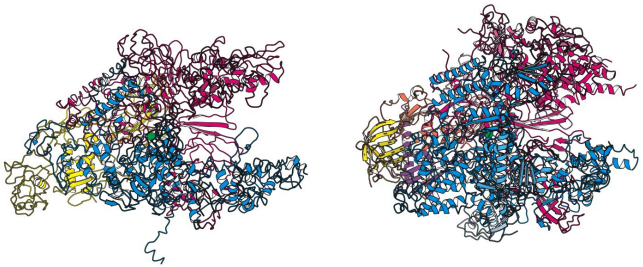
(a) RNA polymerase binding to promoter **(b) Initiation**



Transcription: DNA → RNA

RNA Polymerases in Eukaryotes

Form	Product	Location
I	rRNA	Nucleolus
II	mRNA, snRNA	Nucleoplasm
III	5S rRNA, tRNA	Nucleoplasm



Prokaryotic RNA polymerase **Eukaryotic RNA polymerase**

Figure 29.1
Biochemistry, Seventh Edition
© 2012 W. H. Freeman and Company

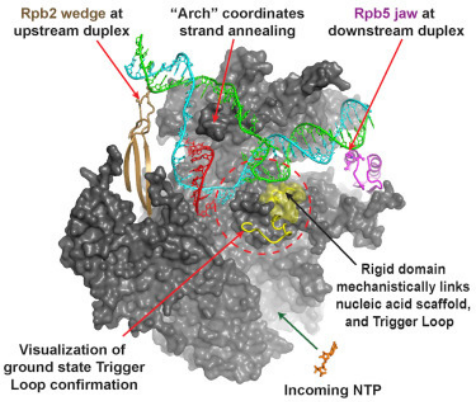
Transcription: DNA → RNA

Prokaryotic versus Eukaryotic Transcription

4) RNA polymerases

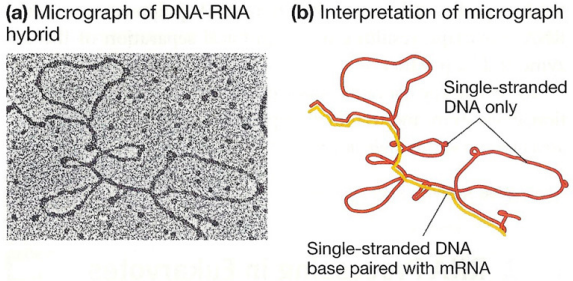
- There are three distinct classes of RNA polymerases in eukaryotic cells. All are large enzymes with multiple subunits. Each class of RNA polymerase recognizes particular types of genes.
- RNA polymerase I- Synthesizes the precursor of the large ribosomal RNAs (28S, 18S and 5.8S).
- RNA polymerase II - Synthesizes the precursors of messenger RNA and small nuclear RNAs(snRNAs).
- RNA polymerase III- Synthesizes small RNA, including t RNAs, small 5S RNA and some snRNAs.

Large Network of Interactions for DNA binding



mRNA processing

In 1977, Phil Sharp (Nobel Prize 1993) hybridized an mRNA to its DNA template and prepared the hybrid molecule for electron microscopy by coating the nucleic acid with a basic protein, then using rotary shadowing to coat the nucleic acid-protein complex.

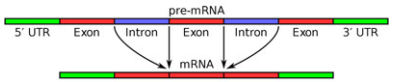


Eucaryotic genes are discontinuous!!!

mRNA processing

The transcript is discontinuous.
There are parts of the DNA template that are not represented in the mRNA.

Eukaryotic mRNA processing

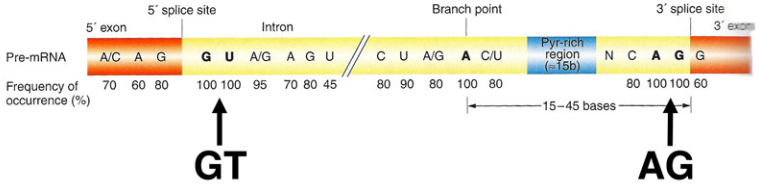


introns are removed (splicing)

When a eukaryotic gene is transcribed, the primary transcript is processed in the nucleus in several ways. The most striking modification is splicing. Parts of the primary transcript, called introns, are spliced out of the mRNA. The remaining segments of mRNA are called exons.

mRNA processing

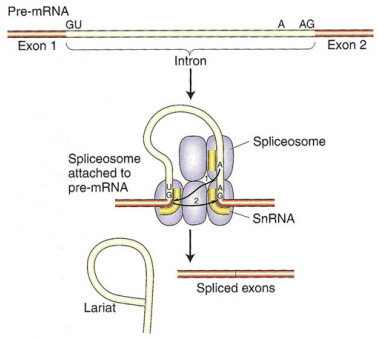
Sequencing of many eukaryotic genes reveals a consensus sequence for splice sites to remove introns



The 5' end of the intron begins with a splice donor site that almost always inclues GT as the first two bases of the intron (very rarely, it's GC).
The 3' end of the intron ends with a splice acceptor site that always includes AG as the last two bases of the intron. Around the 5' GT and the 3' AG are short consensus sequences that allow us to identify likely splice sites in genomic DNA.

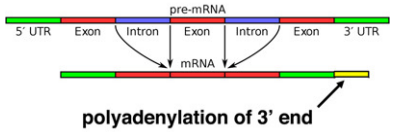
mRNA processing

Splicing is facilitated by a ribonucleoprotein complex called the spliceosome. The spliceosome carries out the removal of introns as RNA lariats, joining exons together to make a mature mRNA, as shown below.

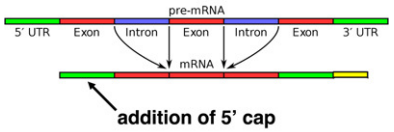


mRNA processing

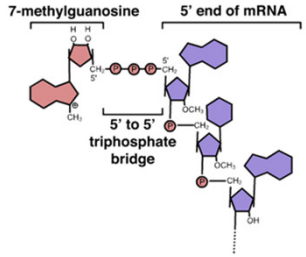
Eukaryotic mRNAs are also modified at the 3' end by the addition of a poly-A tail.



Finally, eukaryotic mRNAs have a chemical modification of the 5' end, called a cap. The cap is added to the first base of the 5' UTR

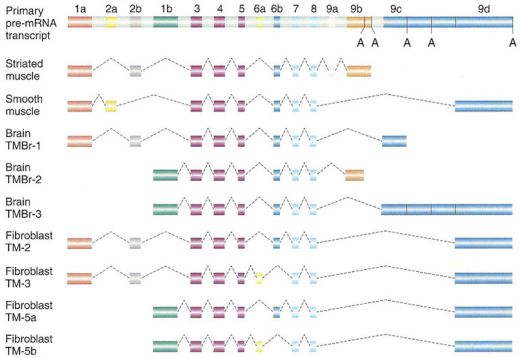


Eukaryotic 5′ cap is a modified GTP

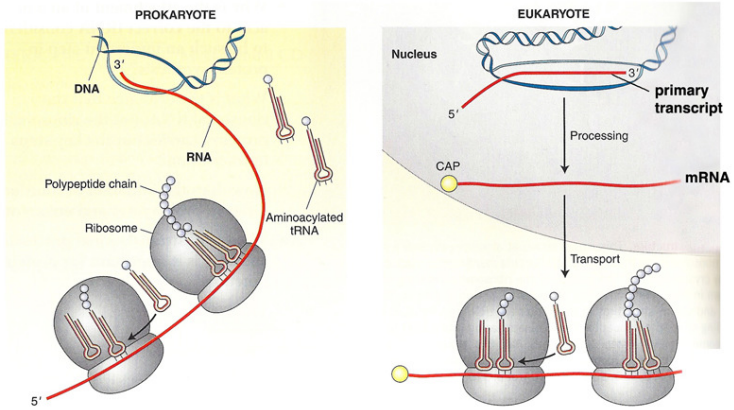


Alternative splicing - tropomyosin

Splicing is regulated, with many genes producing multiple isoforms of the same protein that can differ considerably in their amino acid sequence due to alternative splicing. Isoforms of the muscle protein tropomyosin derived from alternative splicing are shown below. While there are some exons common to all isoforms, some isoforms have large protein segments entirely missing from other isoforms

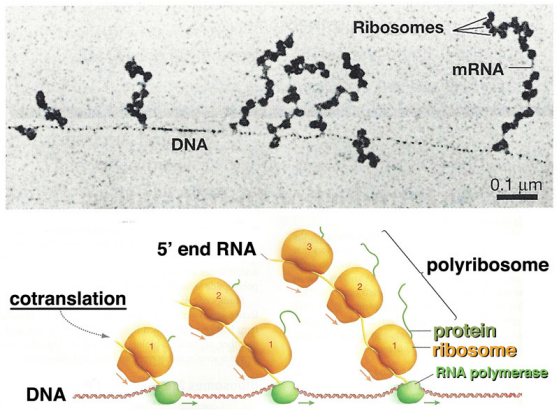


Translation: RNA → proteins



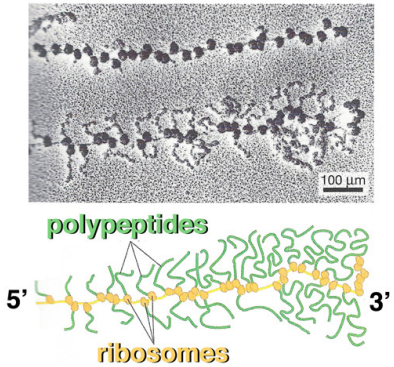
Translation: RNA → proteins

The coupling of transcription and translation in bacteria



Translation: RNA → proteins

Eukaryotic polyribosomes

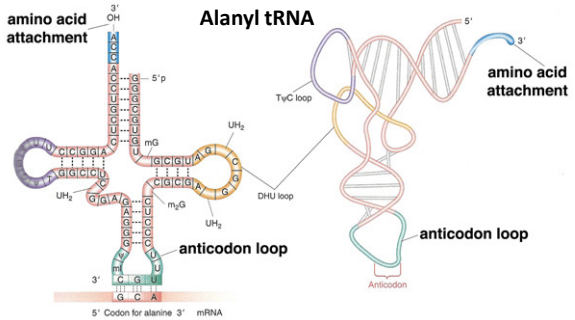


One can see the newly-synthesized peptide, and can tell which is the 5' end of the mRNA, because there are shorter protein tails on the ribosomes at that end.

Translation: RNA → proteins

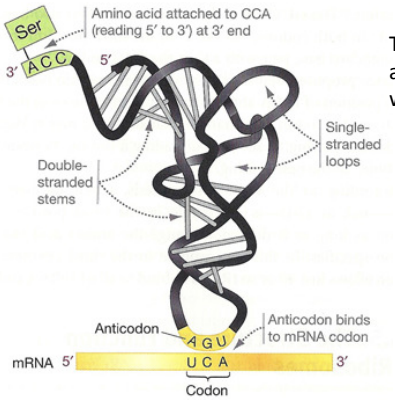
A transfer RNA has a cloverleaf structure with regions of base pairing.
A tRNA has the structure shown here both as a flat cloverleaf and in its folded form.
Two important parts of a tRNA:

- the anticodon, which participates in base pairing with a codon in the mRNA
- the site of amino acid attachment at the 3' end of the tRNA



Translation: RNA → proteins

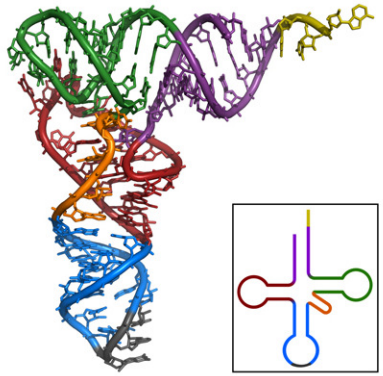
Charged serine tRNA



This shows a "charged" serine tRNA, covalently attached to the amino acid serine at its 3' end, with the anticodon paired to a serine codon

Translation: RNA → proteins

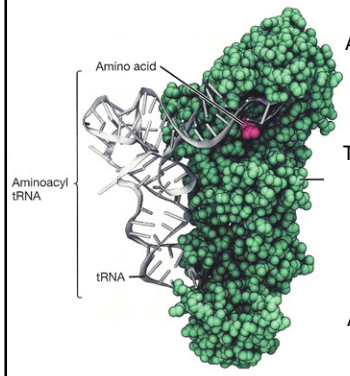
tRNA in 3D



This is a better representation of the 3D structure of a tRNA. The model is color-coded to the flat cloverleaf representation in the lower right

Translation: RNA → proteins

Aminoacyl tRNA synthethase



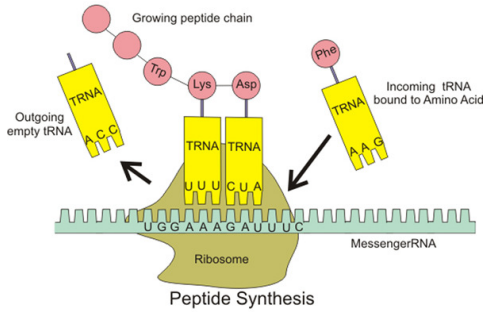
A special set of enzymes "charges" tRNAs, attaching the correct amino acid to particular tRNAs.

A charged tRNA is called an aminoacyl tRNA, so the charging enzymes are more properly called aminoacyl tRNA synthetases.

There is only one aminoacyl tRNA synthetase for each amino acid, even though there can be multiple tRNAs for that amino acid. Each aminoacyl tRNA synthetase is able to recognize all of the tRNAs that need to be charged with the one amino acid that is their specialty.

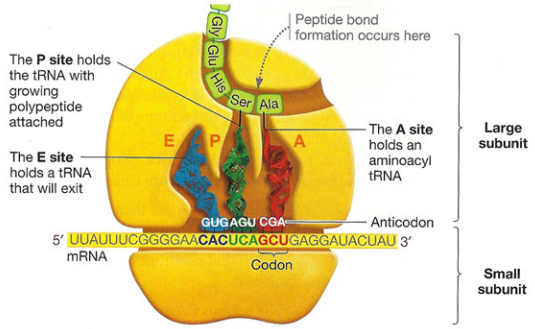
Amino acids are attached to the hydroxyl (-OH) group at the 3' end of the tRNA through their carboxyl (-COOH) group

Translation: RNA → proteins



Proteins are synthesized by ribosomes that read the sequence of mRNA and write it as protein. Translation is accomplished with the help of charged tRNAs that allow individual codons to specify the next amino acid added to the growing polypeptide. The mRNA is read from the 5' end to the 3' end, with the protein being synthesized from the amino terminus to the carboxy terminus

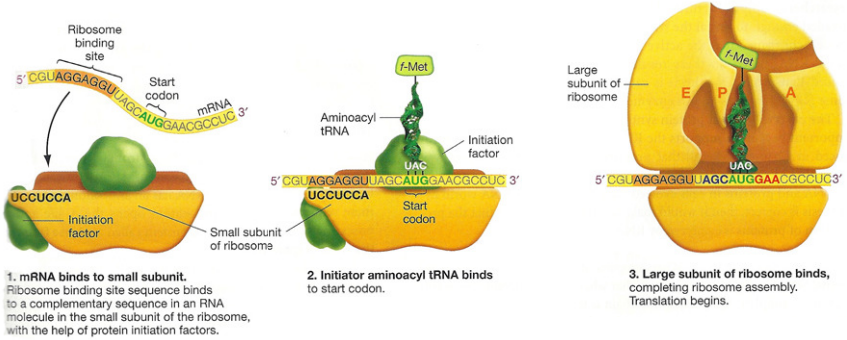
Translation: RNA → proteins



Three sites are associated with tRNAs: the A (aminocyl) site, that accepts a new aminoacyl tRNA; the P (polypeptide) site, that holds a tRNA with the growing polypeptide chain; and the E (exit) site that holds an uncharged tRNA ready to exit the ribosome

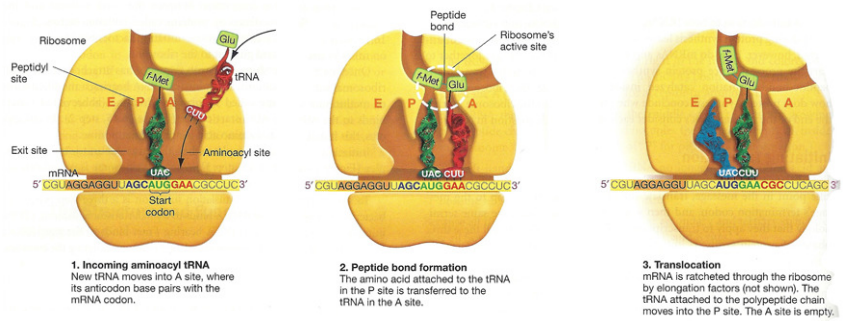
Translation: RNA → proteins

Translation initiation



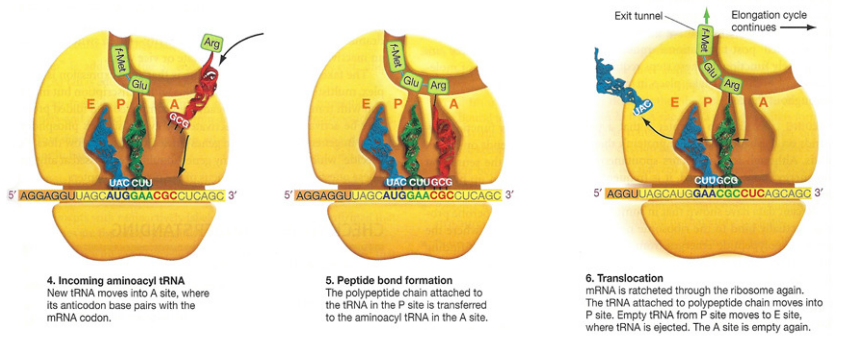
Translation: RNA → proteins

Translation elongation



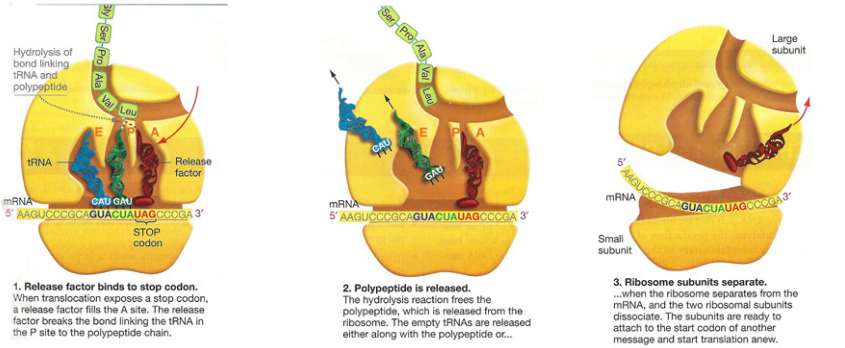
Translation: RNA → proteins

Translation elongation



Translation: RNA → proteins

Translation termination



Translation: RNA → proteins – the genetic code

nonpolar polar basic acidic (stop codon)

Standard genetic code

1st base	2nd base U		2nd base C		2nd base A		2nd base G		3rd base
U	UUU	(Phe/F) Phenylalanine	UCU	(Ser/S) Serine	UAU	(Tyr/Y) Tyrosine	UGU	(Cys/C) Cysteine	U
	UUC		UCC		UAC		UGC		C
	UUA	(Leu/L) Leucine	UCA		UAA	Stop (*Ochre*)	UGA	Stop (*Opal*)	A
	UUG		UCG		UAG	Stop (*Amber*)	UGG	(Trp/W) Tryptophan	G
C	CUU	(Leu/L) Leucine	CCU	(Pro/P) Proline	CAU	(His/H) Histidine	CGU	(Arg/R) Arginine	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	(Gln/Q) Glutamine	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	(Ile/I) Isoleucine	ACU	(Thr/T) Threonine	AAU	(Asn/N) Asparagine	AGU	(Ser/S) Serine	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	(Lys/K) Lysine	AGA	(Arg/R) Arginine	A
	AUG[A]	(Met/M) Methionine	ACG		AAG		AGG		G
G	GUU	(Val/V) Valine	GCU	(Ala/A) Alanine	GAU	(Asp/D) Aspartic acid	GGU	(Gly/G) Glycine	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	(Glu/E) Glutamic acid	GGA		A
	GUG		GCG		GAG		GGG		G