Synthetic life

(continuation of „The molecular origins of life" SoSe 2021)

NaturalNews.com

WiSe 2021/22

Zbigniew Pianowski

7 lectures (90 min. each) in English

1st lecture: 18th Oct. 2021 (Mon.)
Zoom:

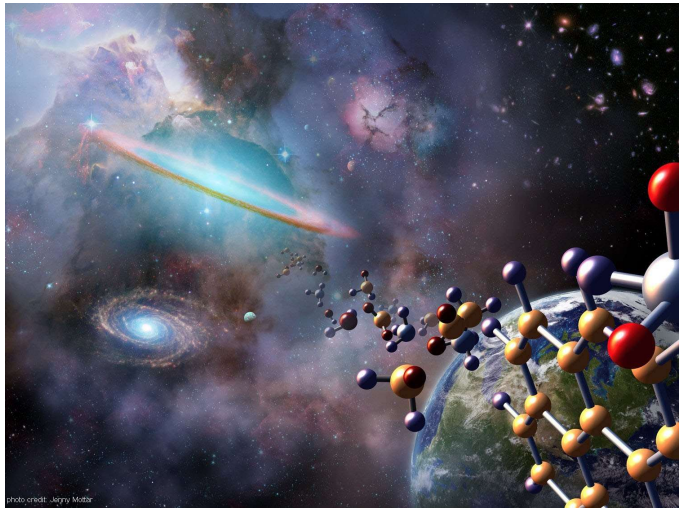
Following lecture terms 15:00-16:30, Mon.:
25.10, 8.11., 15.11., 22.11., 6.12., 13.12.2021.

The most current dates, handouts – on the website:

and by ILIAS (KIT)

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

Synthetic life

www.genome.gov/about-genomics/policy-issues/Synthetic-Biology

How new functions can emerge from known biological building blocks?

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

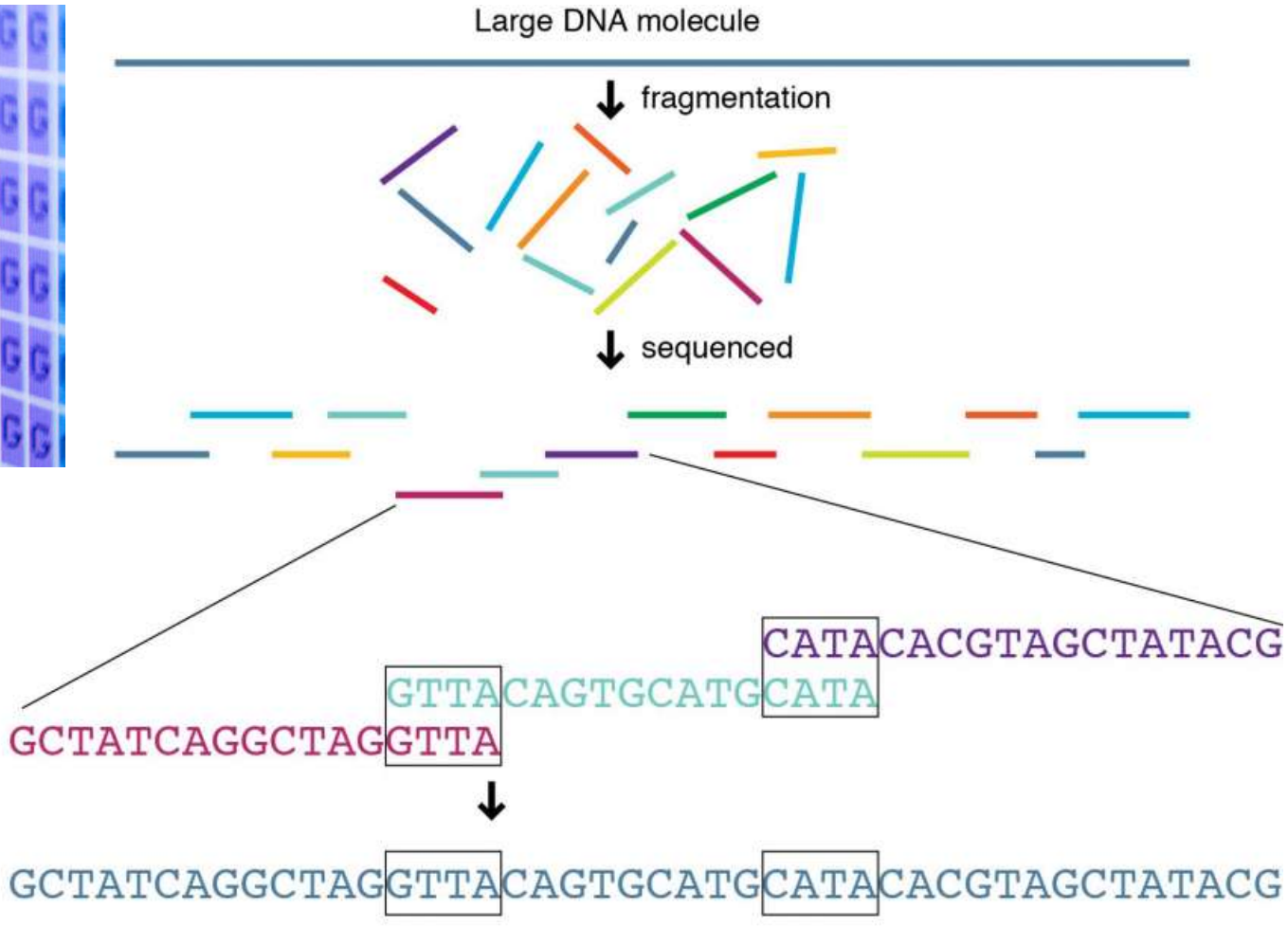
Everything – living or not – is just chemicals made of atoms.

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

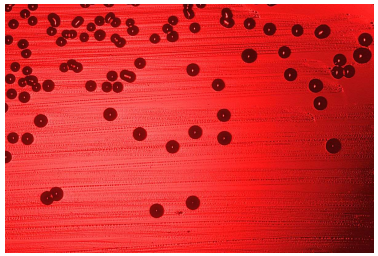
DNA turns dust into life.

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...

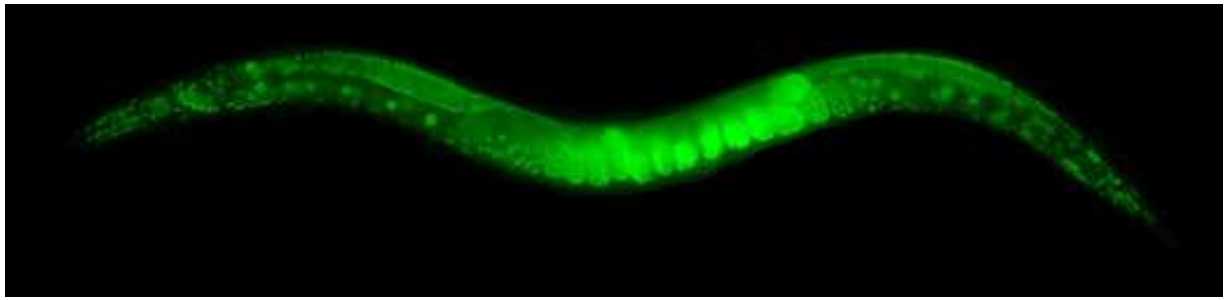
20 years ago, Scientists learned to read the creatures' entire DNA sequence, from beginning to the end – the genome



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



Haemophilius influenzae
1995



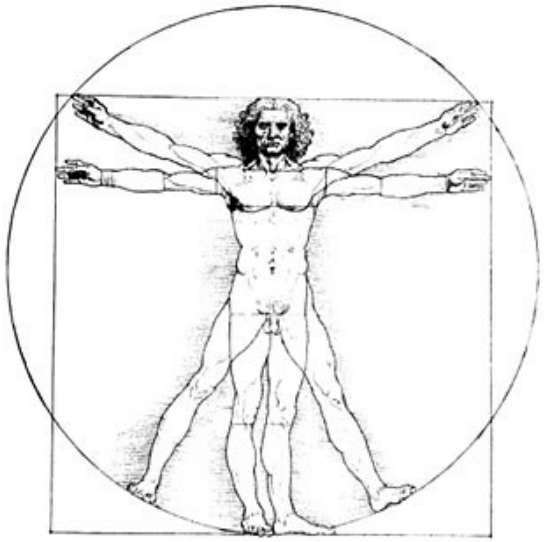
Caenorhabdis elegans
1998

Drosophila melanogaster
2000

Arabidopsis thaliana
2000

... and way up to mammals and human

Mus musculus
2002



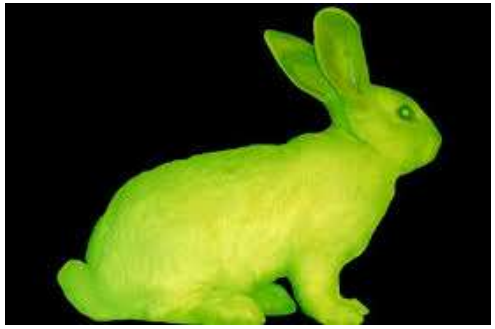
Homo sapiens
2004
Human Genome Project (NIH)
Craig Venter – Celera Genomcs (private)

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?

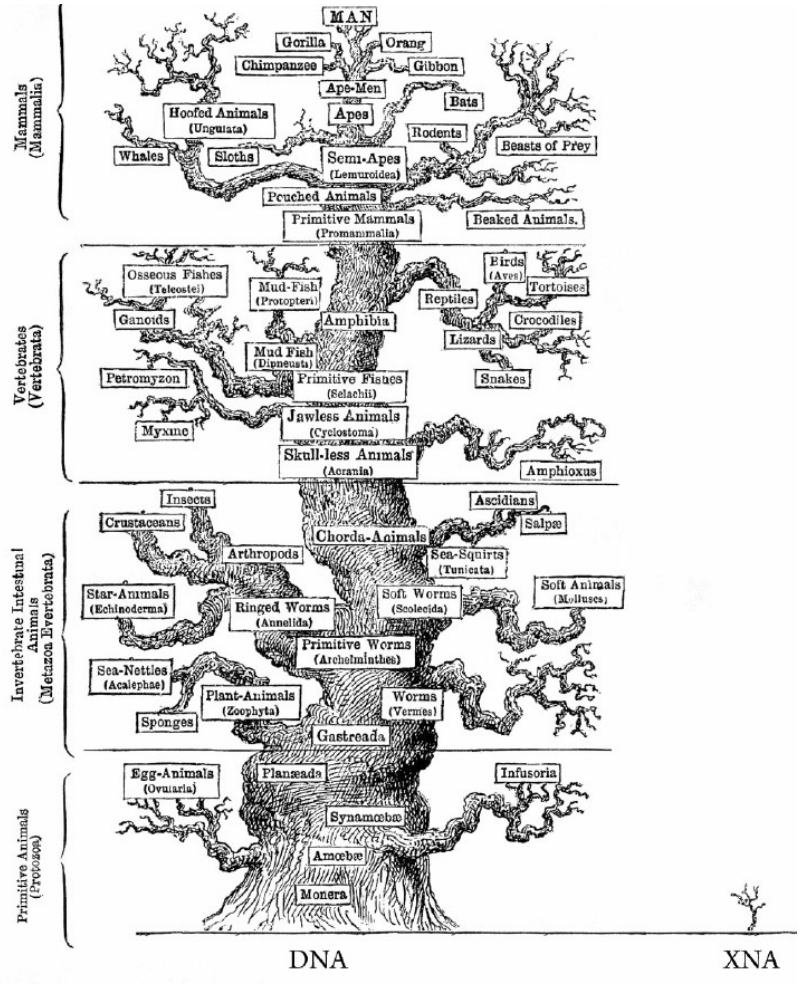
We can improve what was spelled out
for the 3,5 Billion years of evolution.

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?



Overview of the course

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

unnatural base pairing – expansion of the genetic alphabet;

artificial genetic polymers and oligonucleotide analogues (XNA);

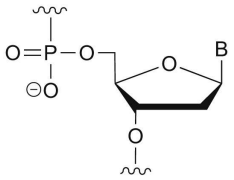
biosynthetic incorporation of **unnatural aminoacids (UAAs)** into proteins;

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

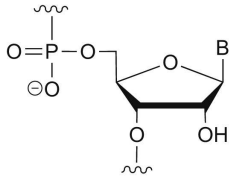
Artificial lipid vesicles as models for protocell multiplication;

design of artificial organisms

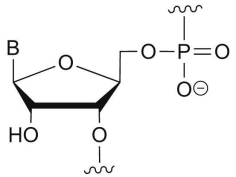
Artificial genetic polymers



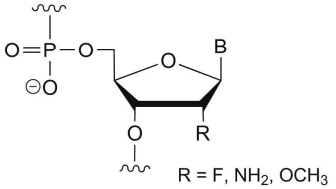
DNA



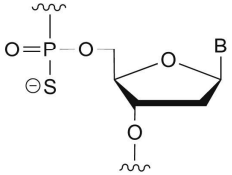
D-RNA
(natural)



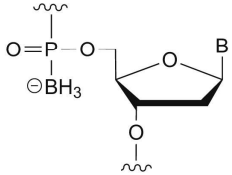
L-RNA
(unnatural)



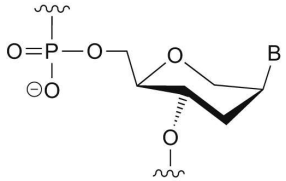
2'-modified RNA



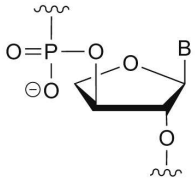
Phosphorothioate



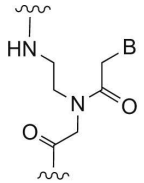
Boranophosphate



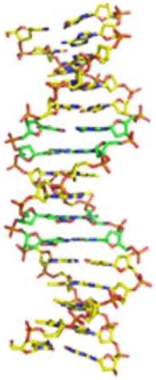
Hexitol Nucleic Acid
(HNA)



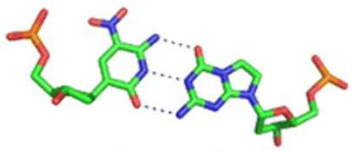
Threose Nucleic Acid
(TNA)

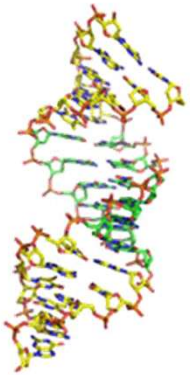


Peptide Nucleic Acid
(PNA)

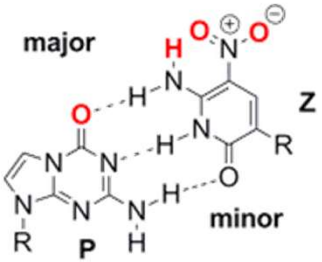


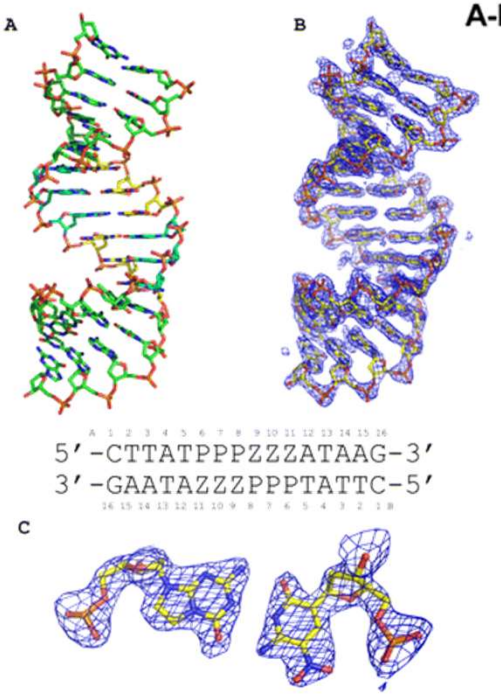
B-DNA

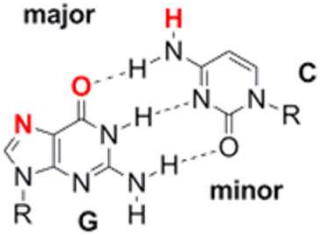


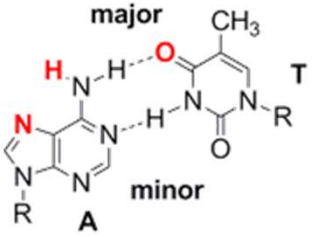


A-DNA

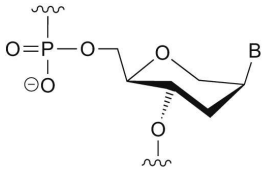




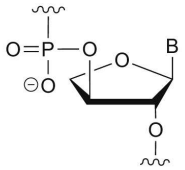




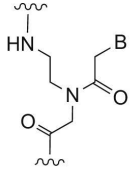
Artificial genetic polymers



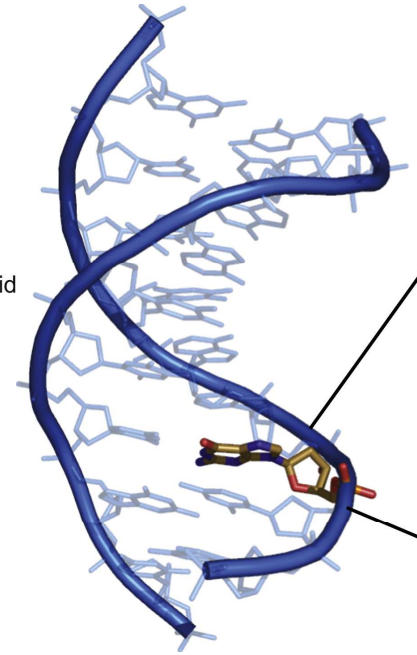
Hexitol Nucleic Acid
(HNA)

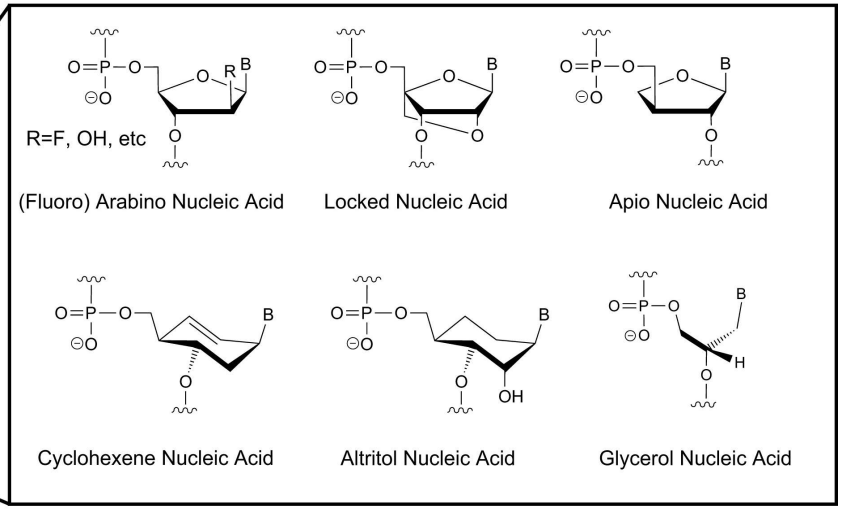


Threose Nucleic Acid
(TNA)



Peptide Nucleic Acid
(PNA)





(Fluoro) Arabino Nucleic Acid

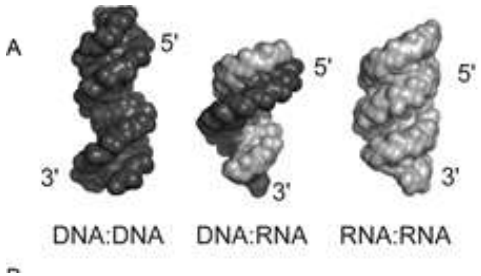
Locked Nucleic Acid

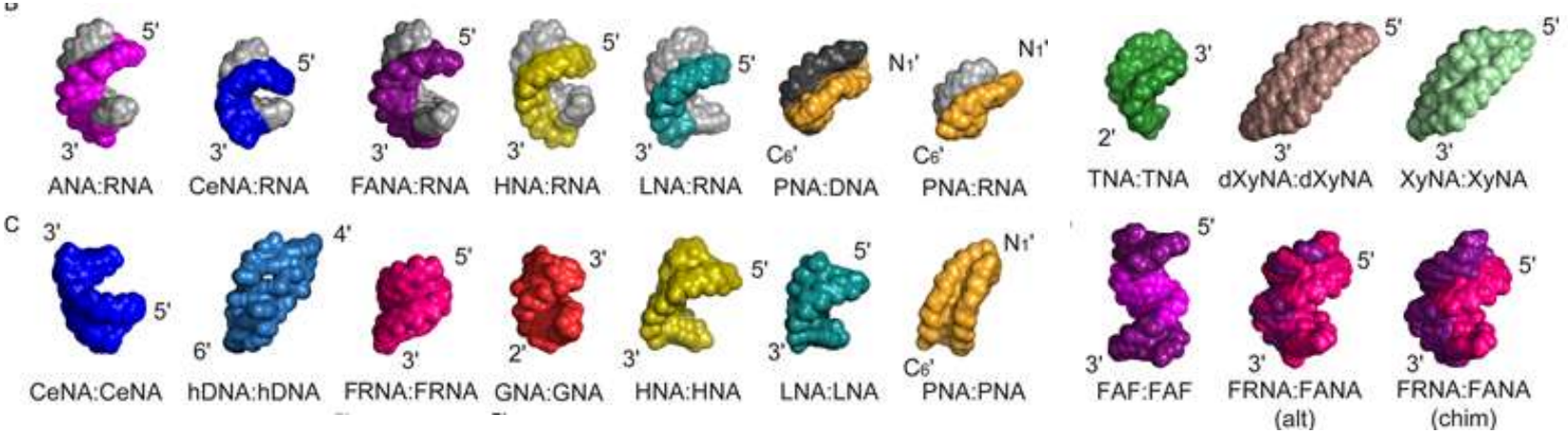
Apio Nucleic Acid

Cyclohexene Nucleic Acid

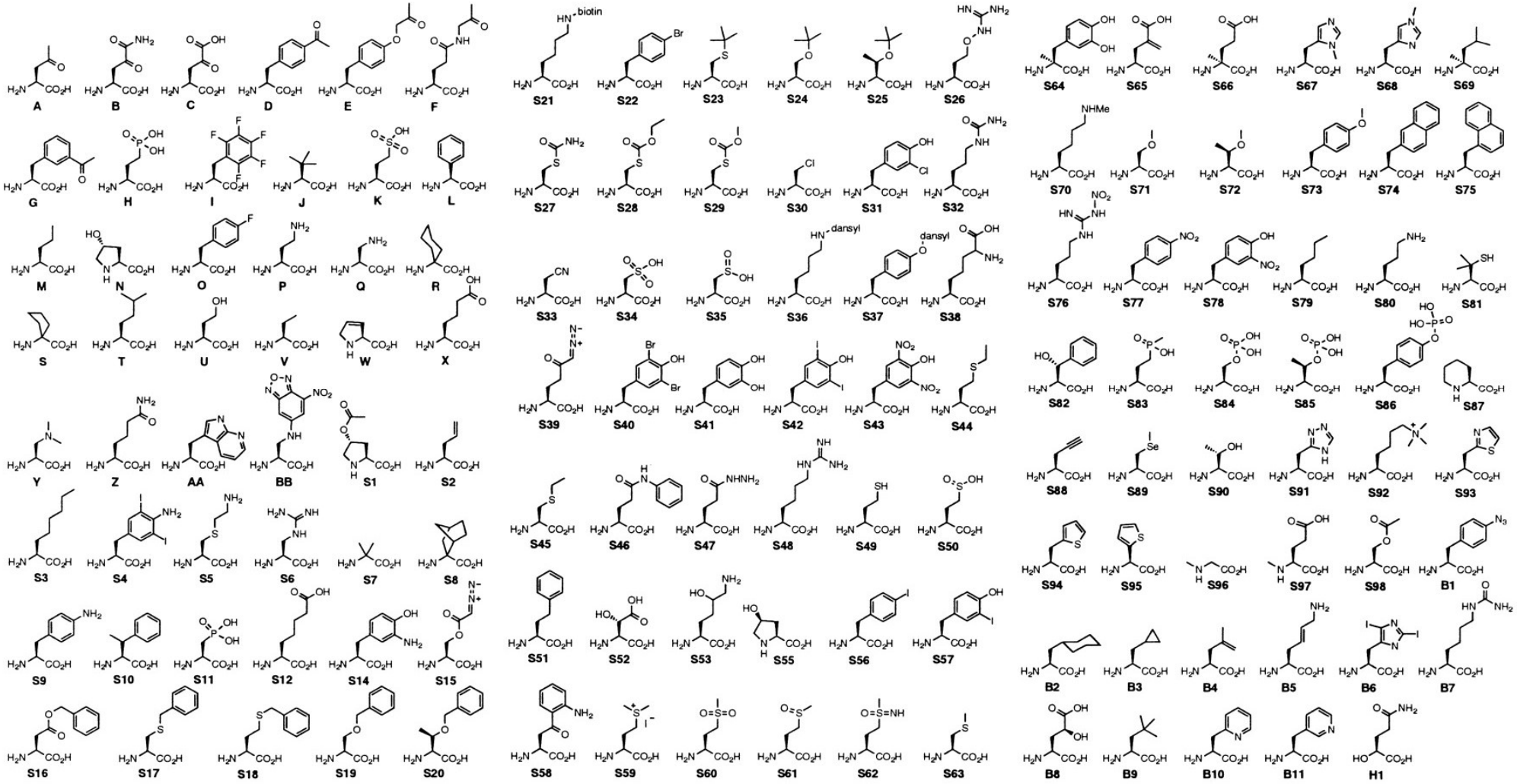
Altritol Nucleic Acid

Glycerol Nucleic Acid

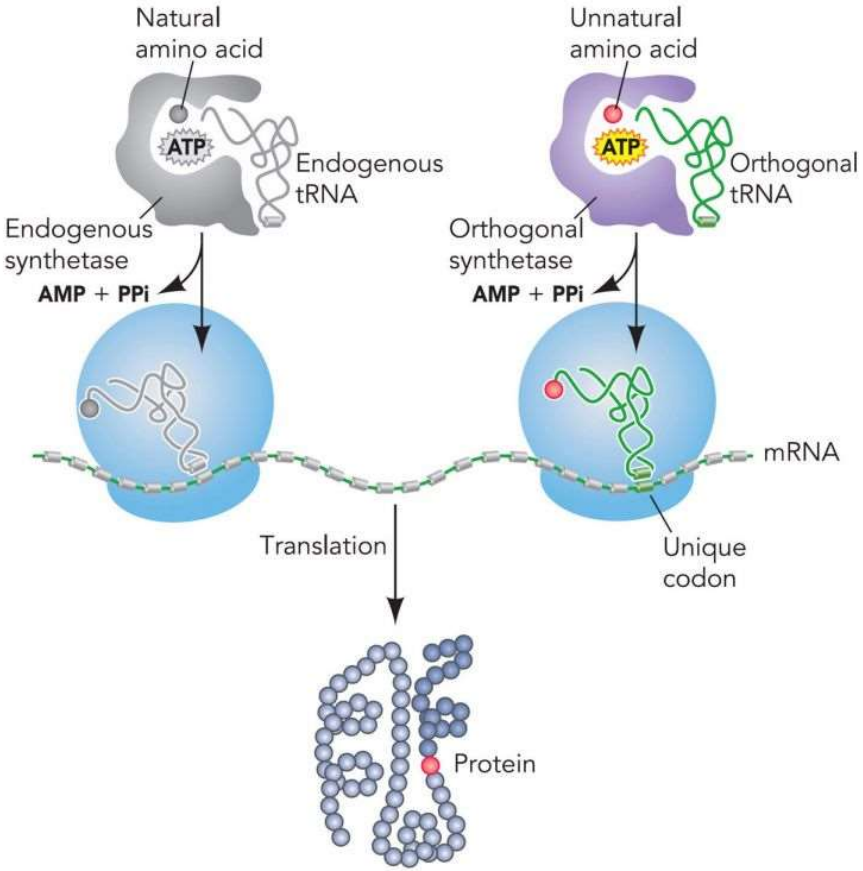


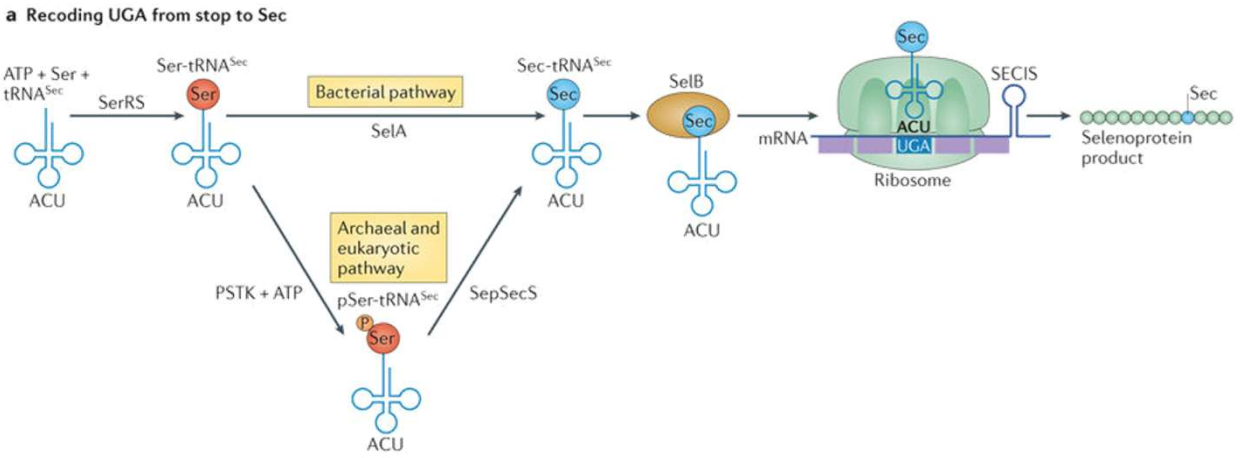


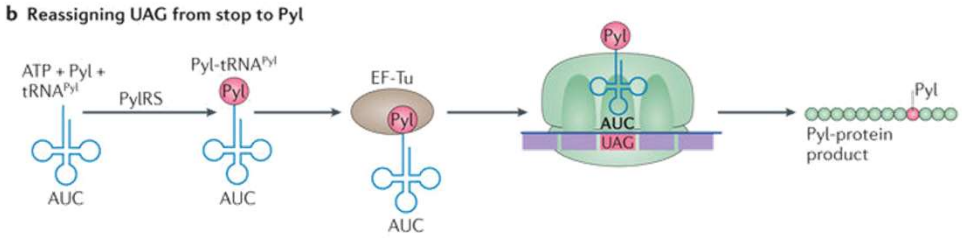
Expanded genetic code



Expanded genetic code

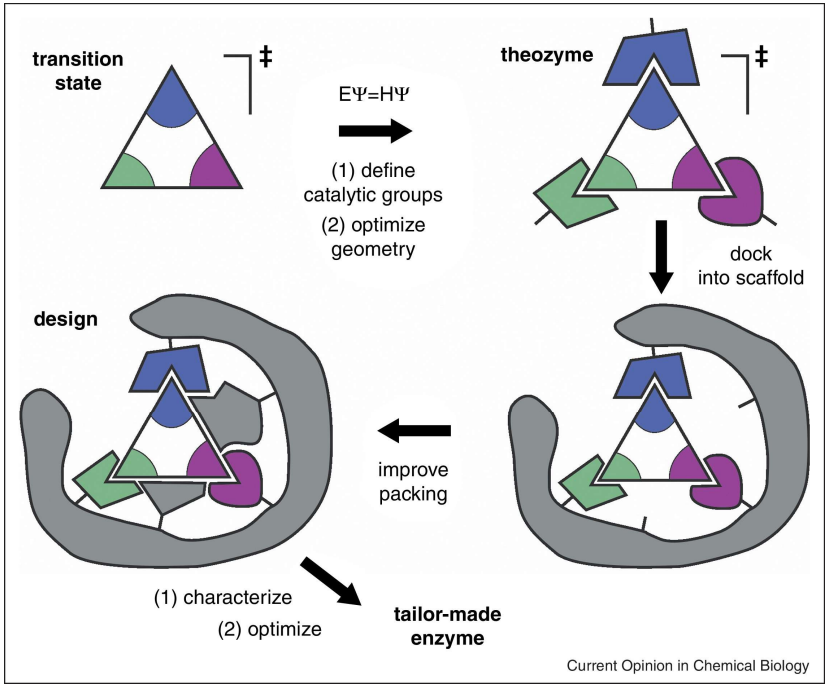


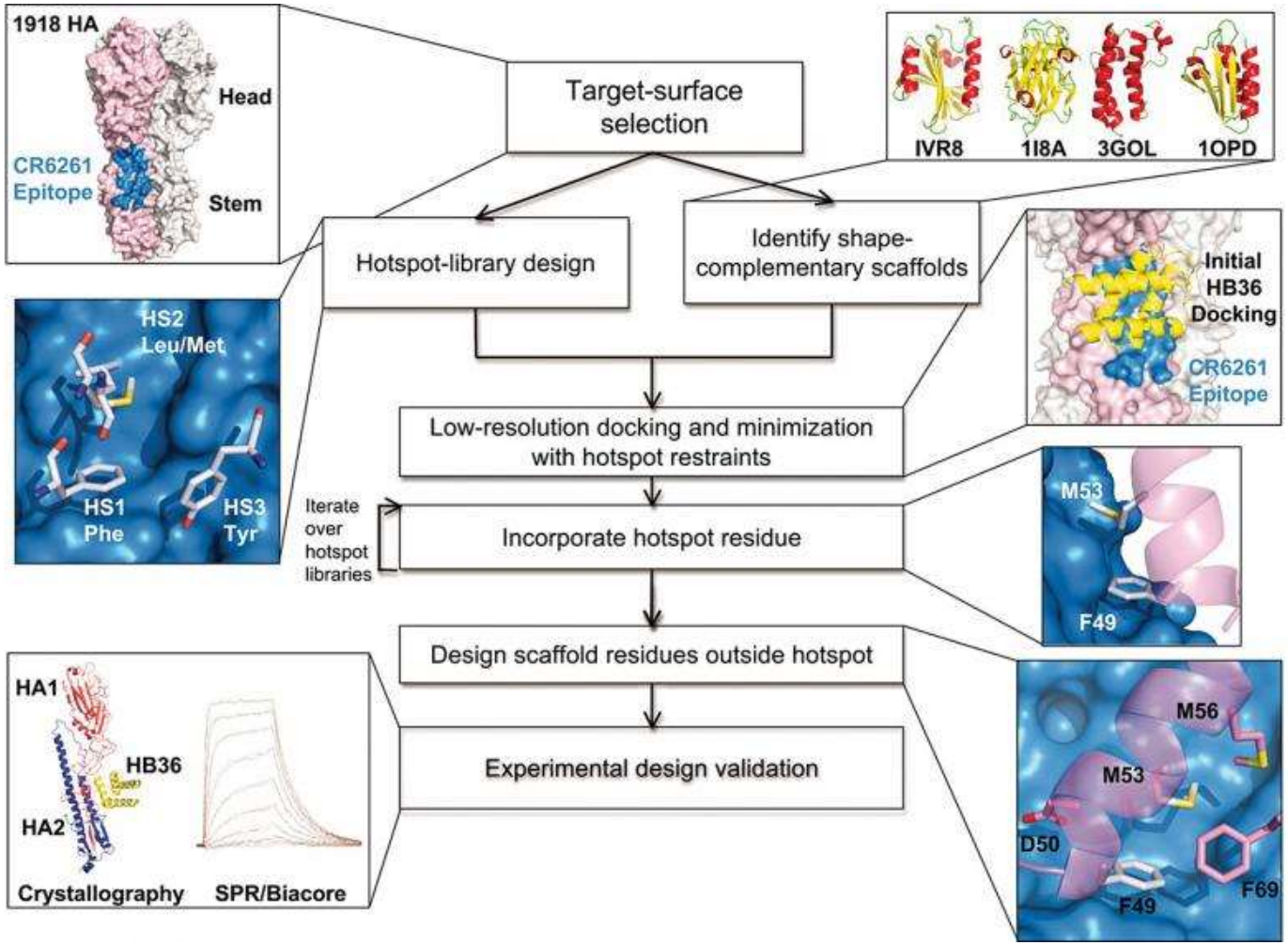




Nature Reviews | Microbiology

Protein engineering and de novo enzyme design





Introduction

How chemists and biologists are learning from each other?

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

Introduction

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

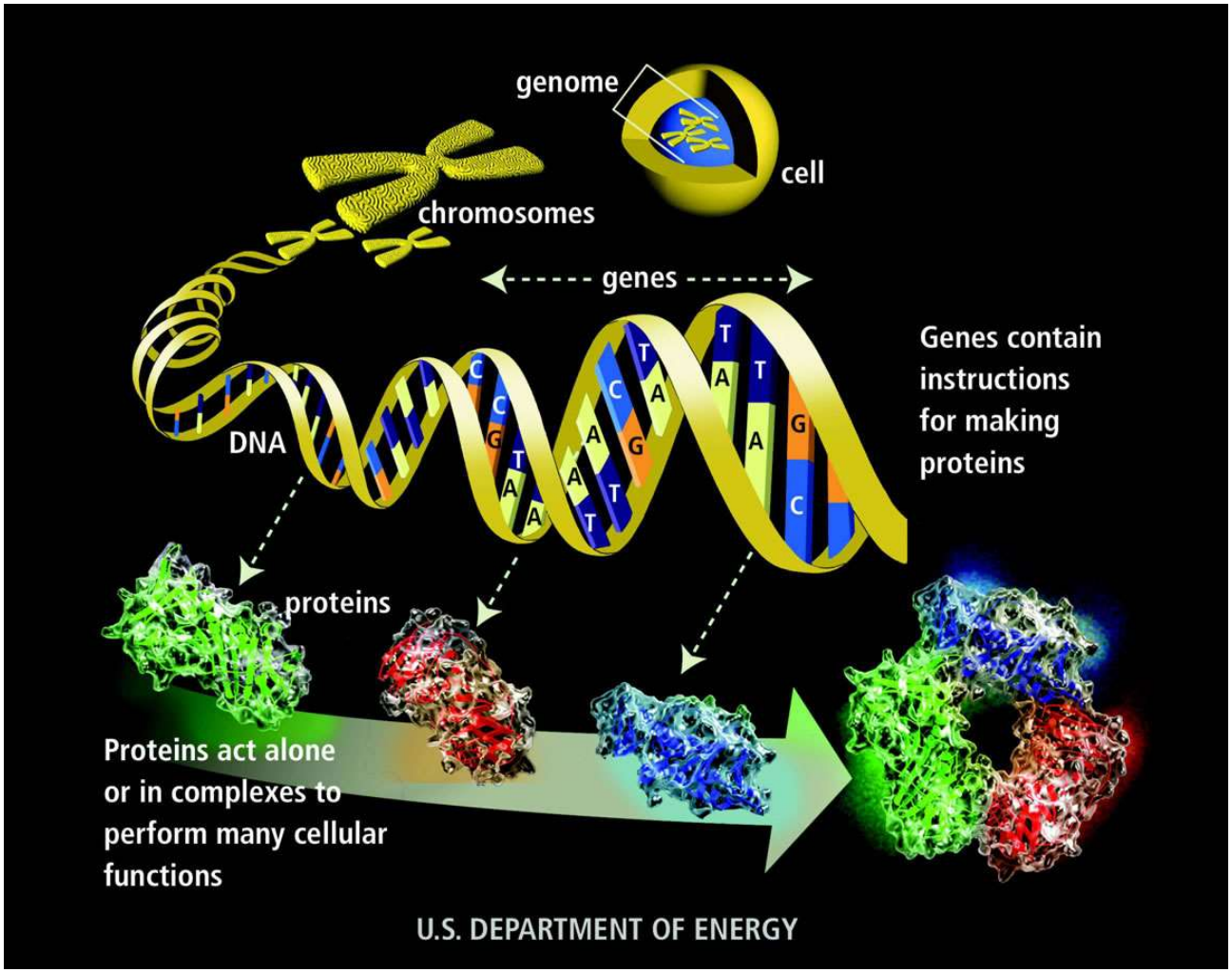
Polymerases and ribosomes - the molecular machines of life

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

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

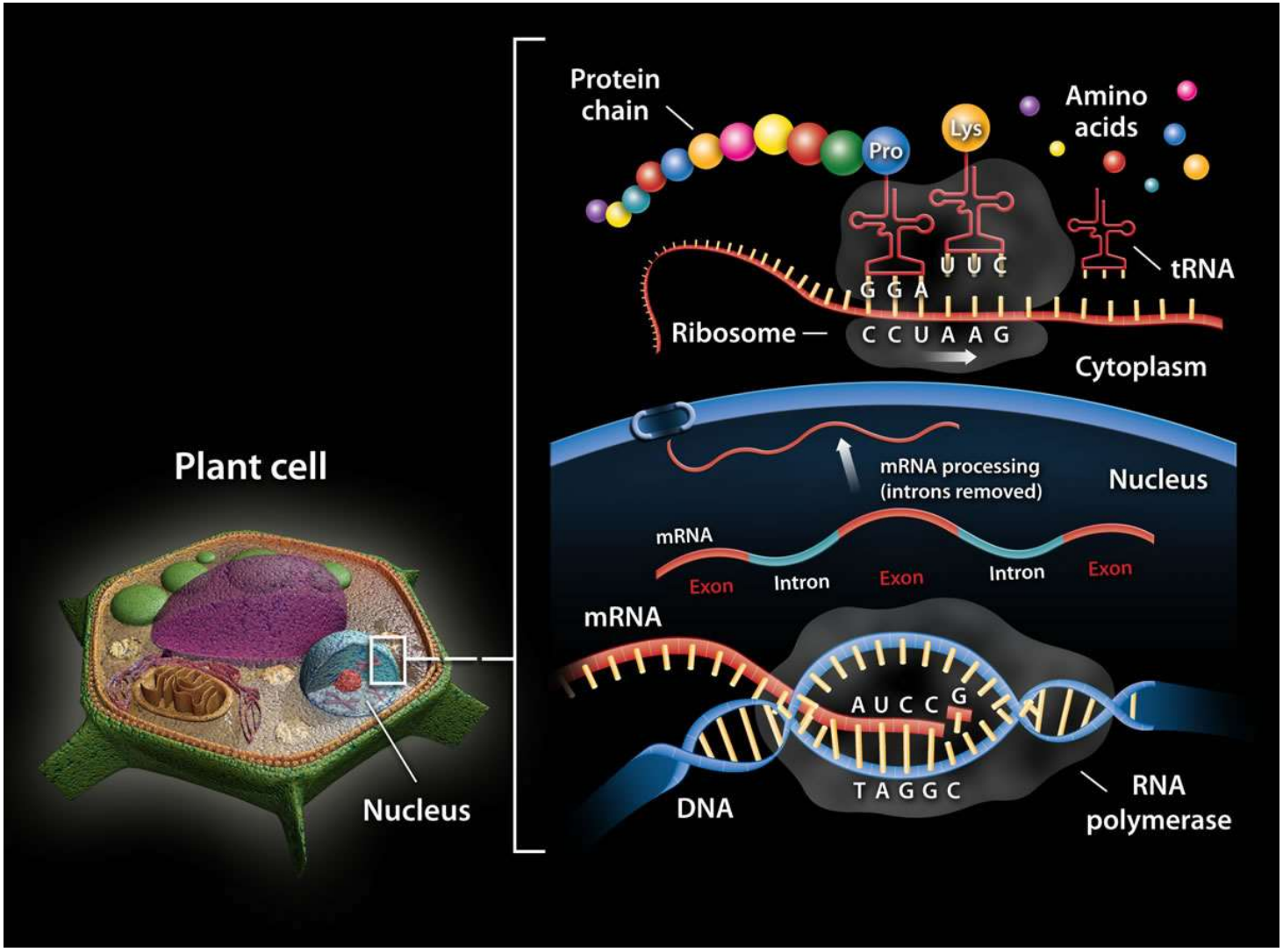
Protein engineering – how to make desired modifications in proteins

From DNA to proteins

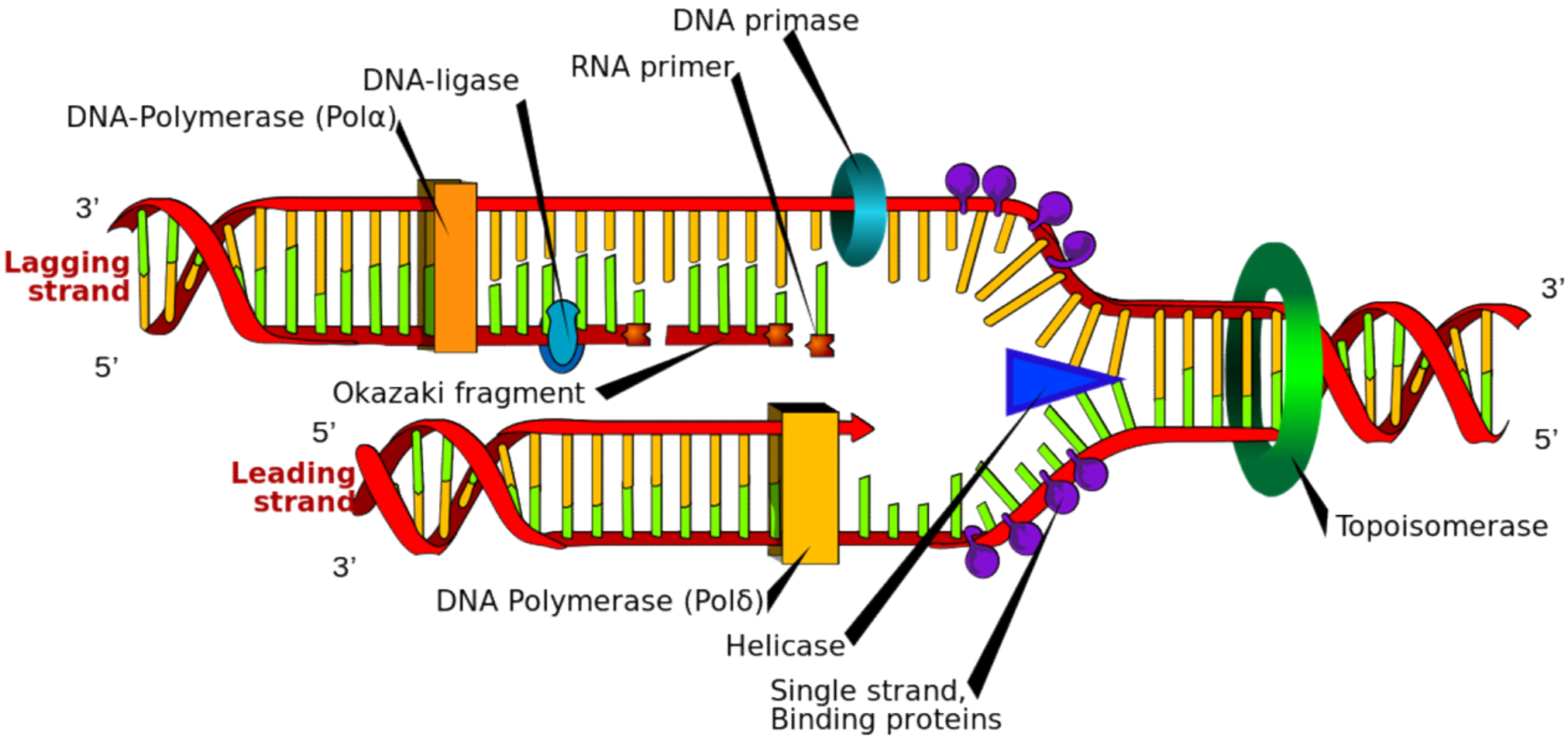


https://www.youtube.com/watch?v=gG7uCskUOrA

The Central Dogma: From DNA to proteins

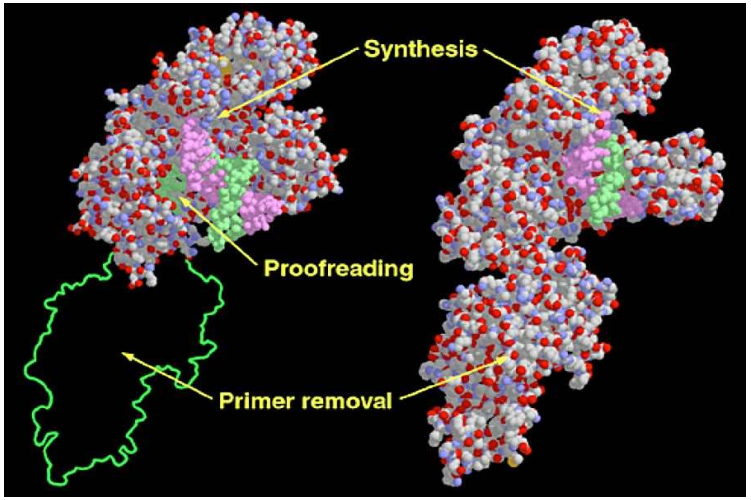


DNA replication

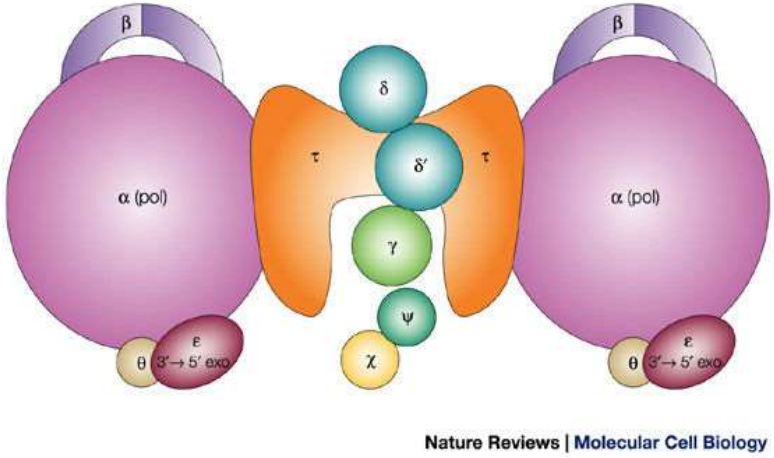


https://www.youtube.com/watch?v=TNKWgcFPHqw&ab_channel=yourgenome

DNA polymerases



Procaryotic DNA Polymerases



Nature Reviews | Molecular Cell Biology

Procaryotic DNA Polymerase III

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′

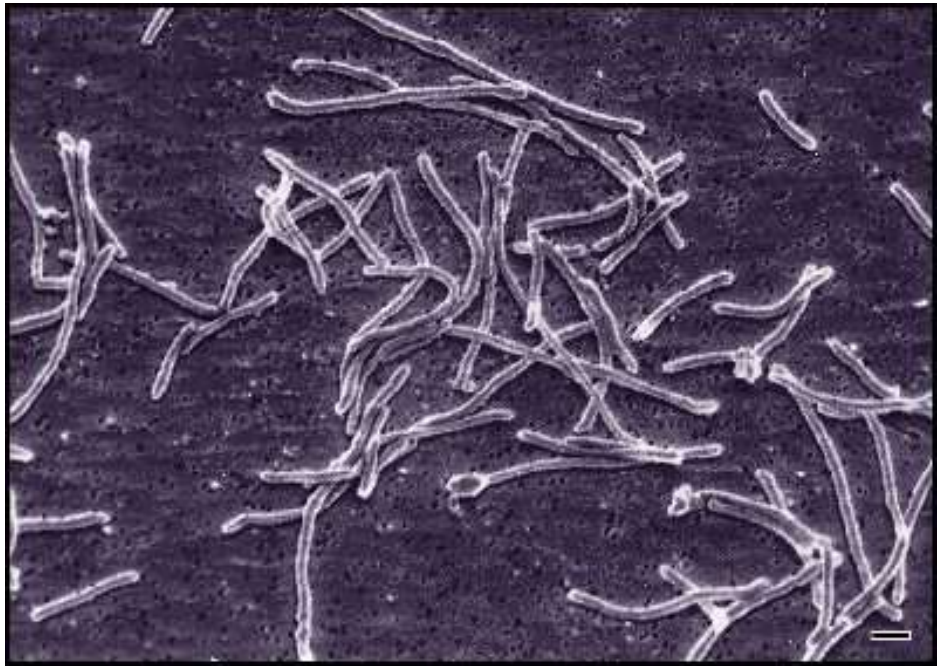
*The main enzyme of replication

70

SUBUNITS OF THE DNA POLYMERASE III HOLOENZYME

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

Extremophilic organisms



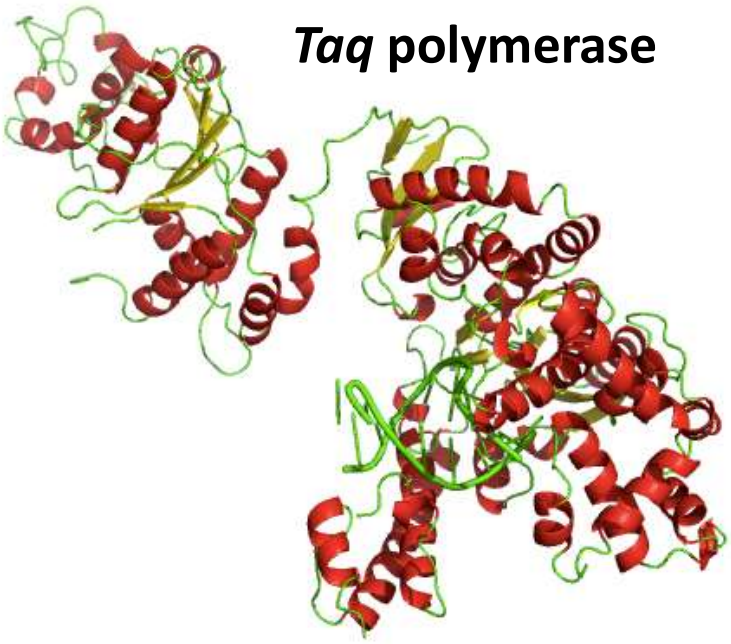
Brian W. Schaller, Yellowstone Park

Thermus aquaticus is a thermophilic bacteria from hot springs in Yellowstone Park
70^0C – optimum, living range: 50-80^0C

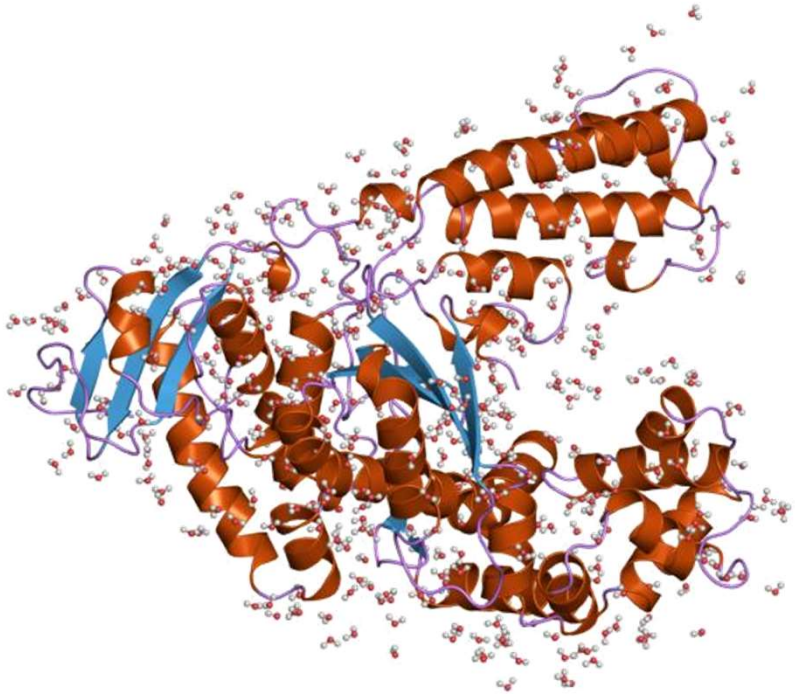
It is a source of thermostable enzymes

PCR – Polymerase Chain Reaction

Taq polymerase



„Adenosine"

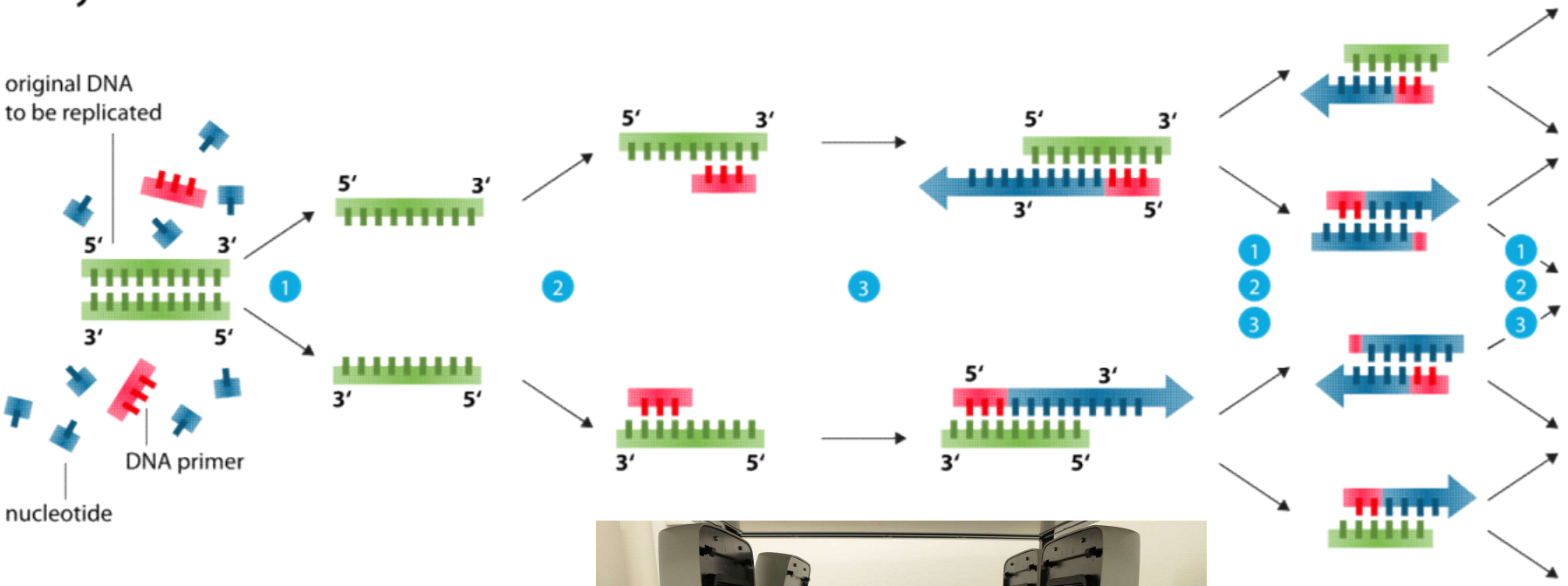


Taq polymerase withstands denaturing conditions (hot temperatures) detrimental for most enzymes. *Activity optimum: 75-80^{0}C, half-life at 95^{0}C > 2.5 h*

1990 – Kary Mullis optimized the PCR technique with *Taq* polymerase (1993 Nobel Prize)

https://www.dnalc.org/view/15475-The-cycles-of-the-polymerase-chain-reaction-PCR-3D-animation.html

Polymerase chain reaction - PCR



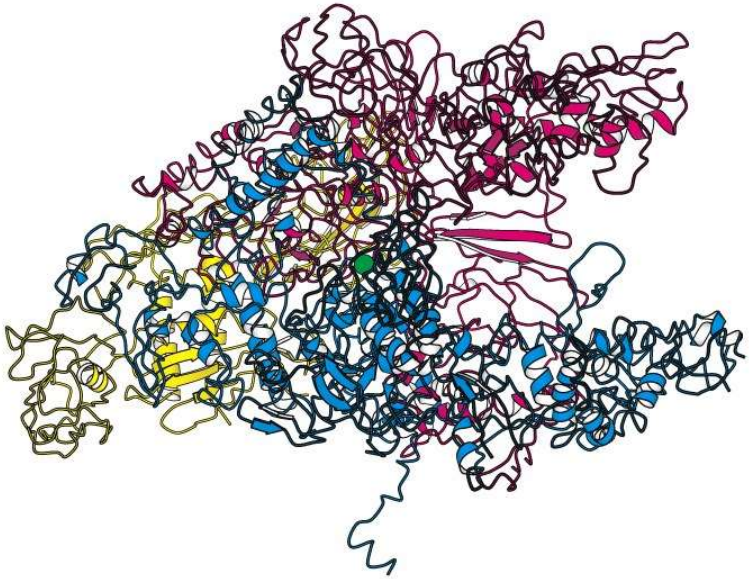
1. **Denaturation** at 94-96°C
2. **Annealing** at ~68°C
3. **Elongation** at ca. 72 °C

Rror

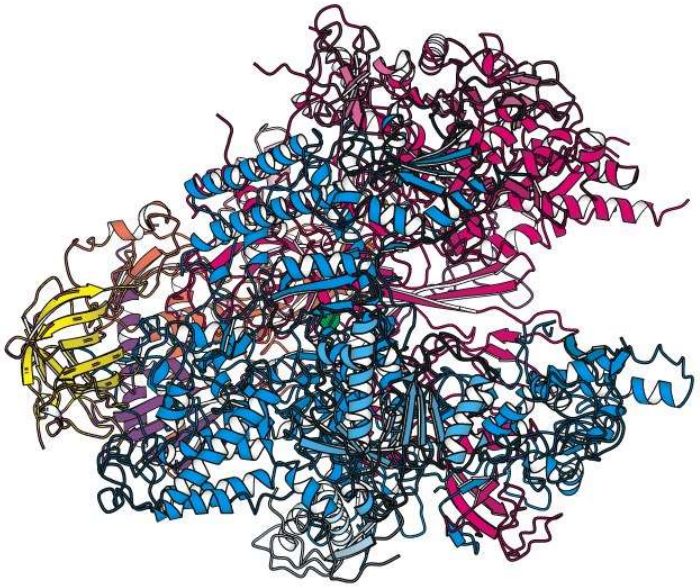
Transcription: DNA → RNA

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

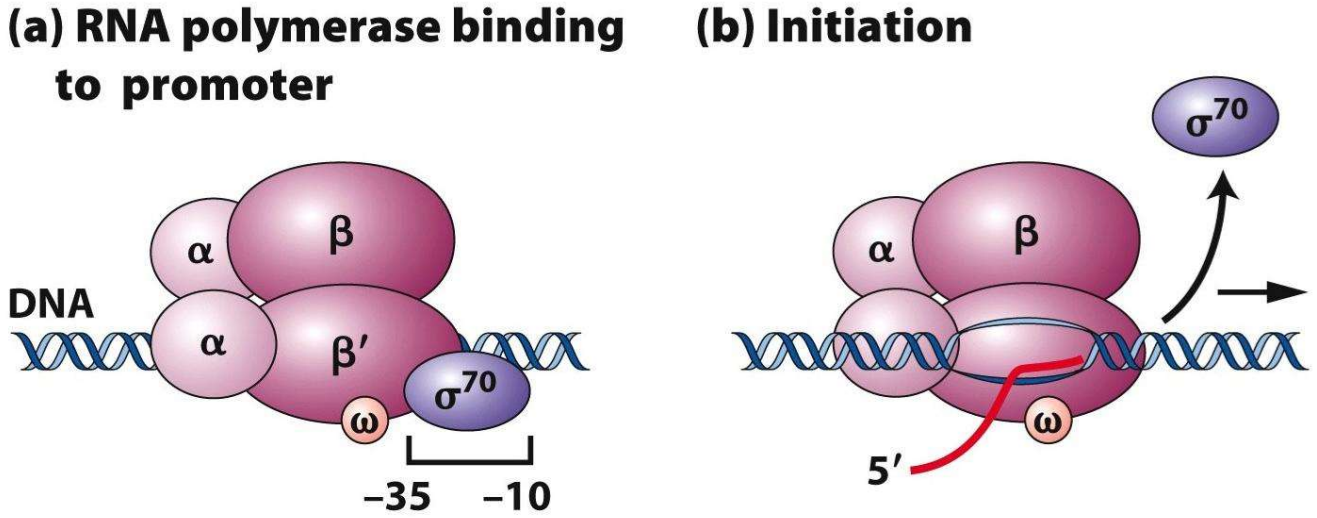
Figure 29.1
***Biochemistry*, Seventh Edition**

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

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

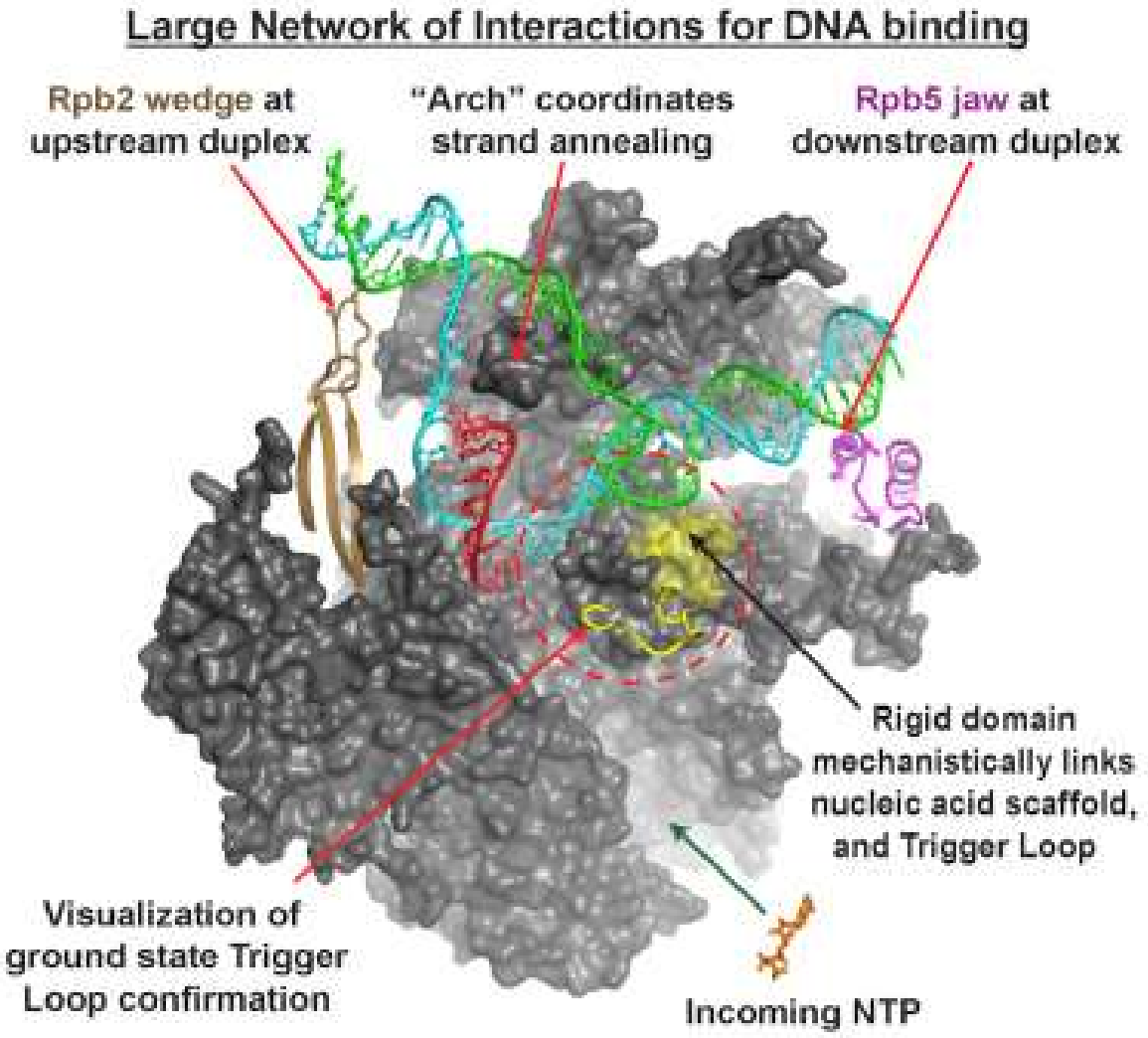


Transcription: DNA → RNA

Prokaryotic versus Eukaryotic Transcription

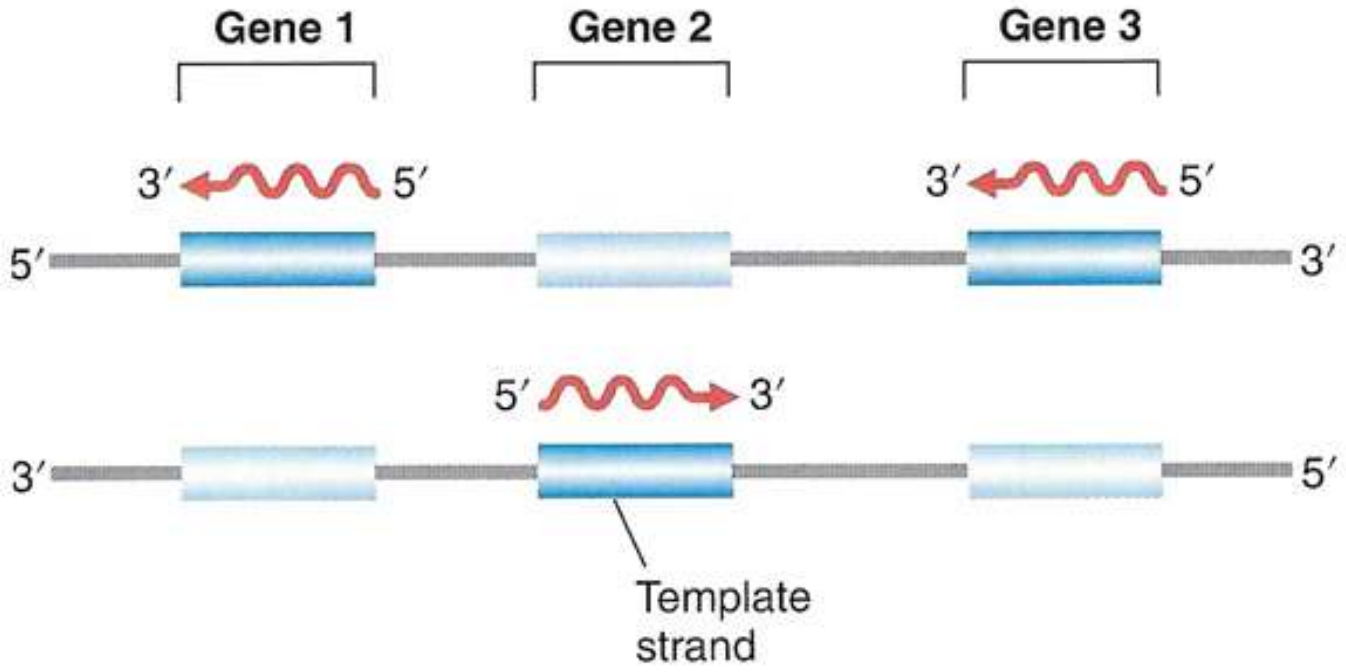
4) RNA polymerases

- 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.



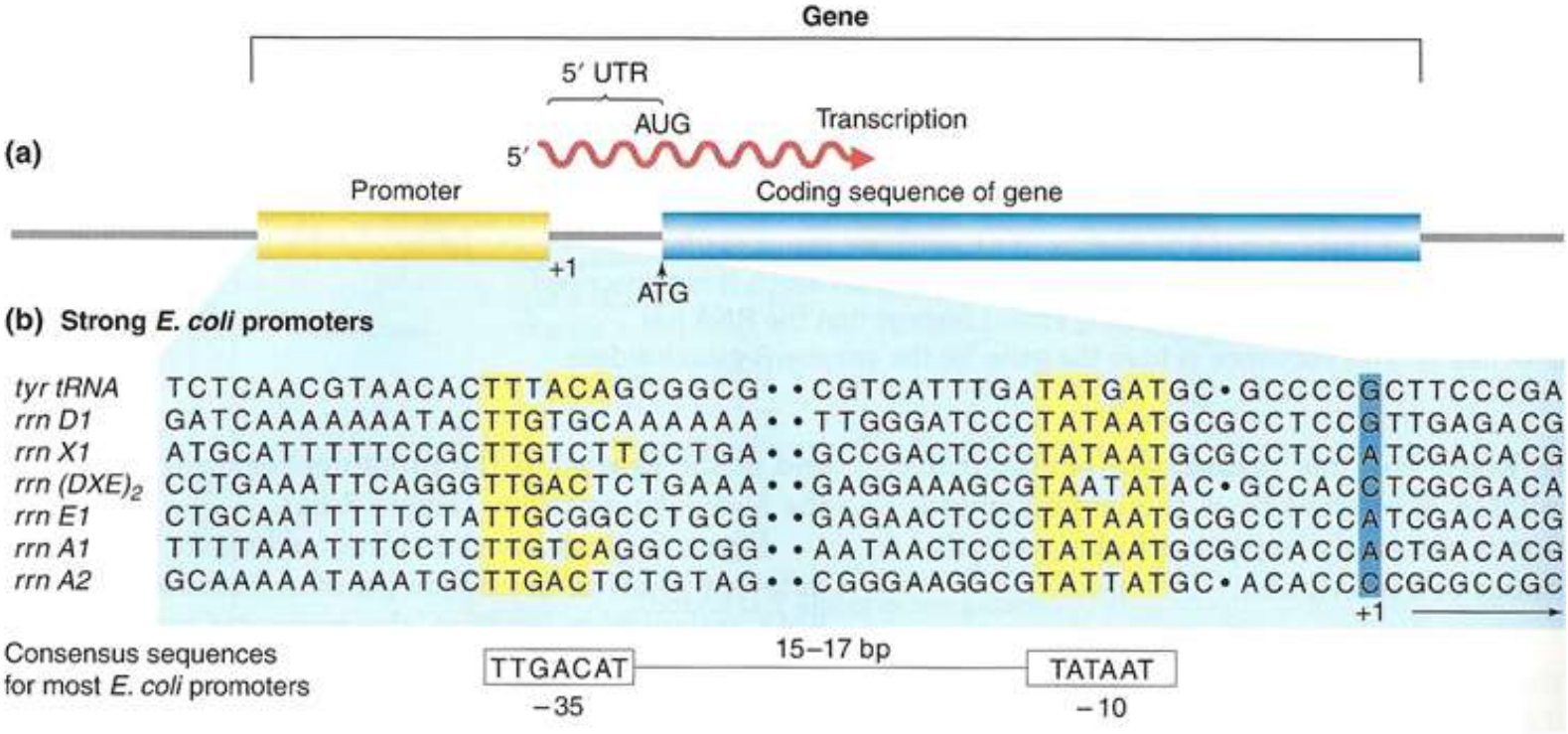
Transcription: DNA → RNA

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*):



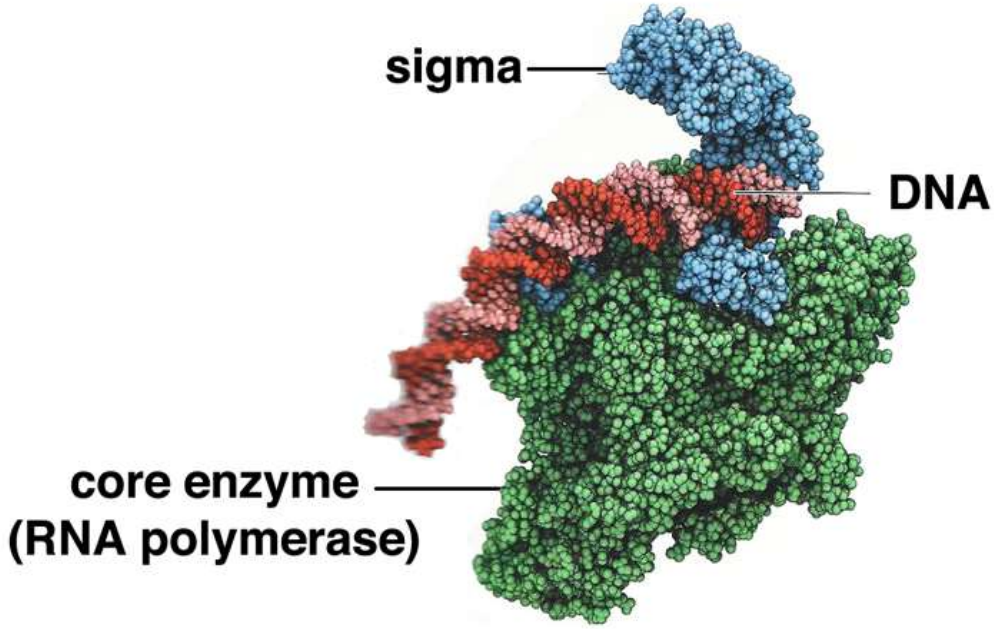
the transcription start site does not exactly correspond to the codon that is used to initiate translation – base 1 (+1). Transcription begins upstream of this sequence. The RNA between the „base 1" and the start of translation some distance downstream is called the 5' untranslated region or 5' UTR.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

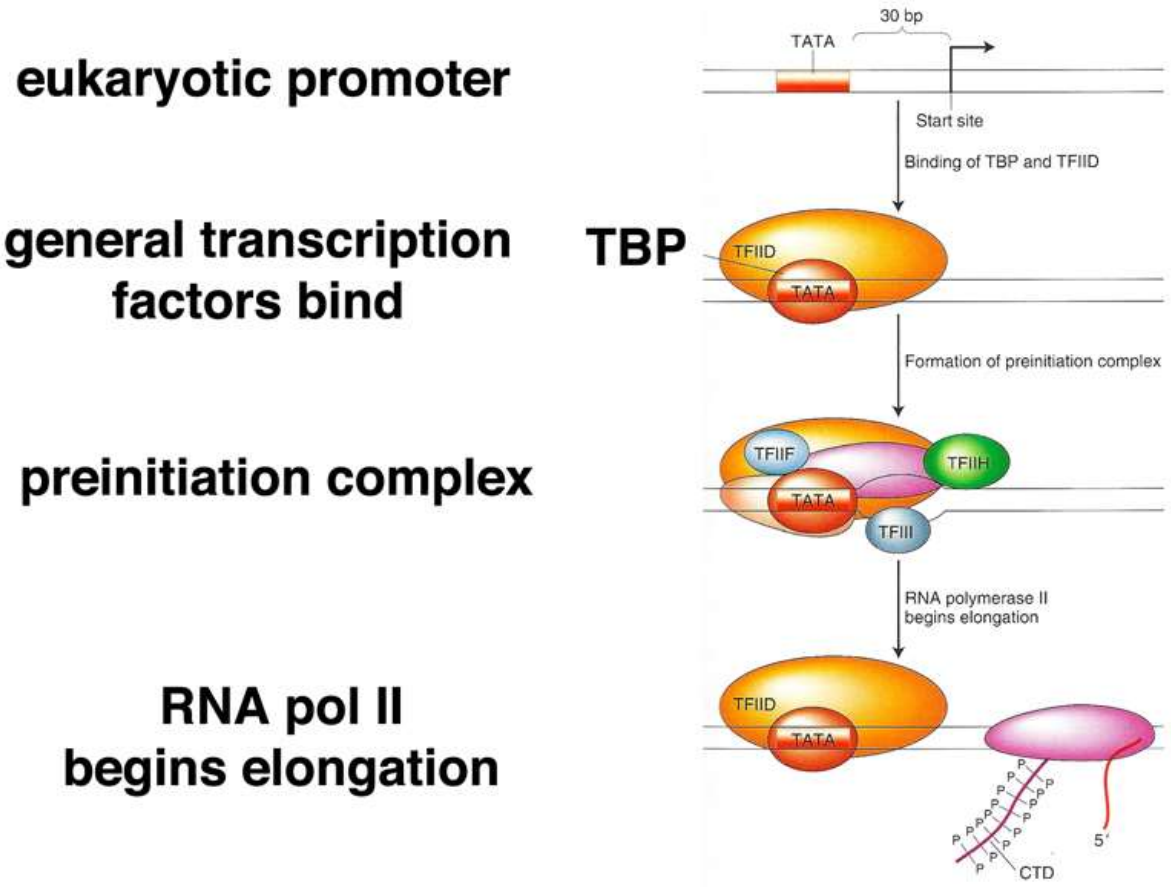
Transcription: DNA → RNA

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.

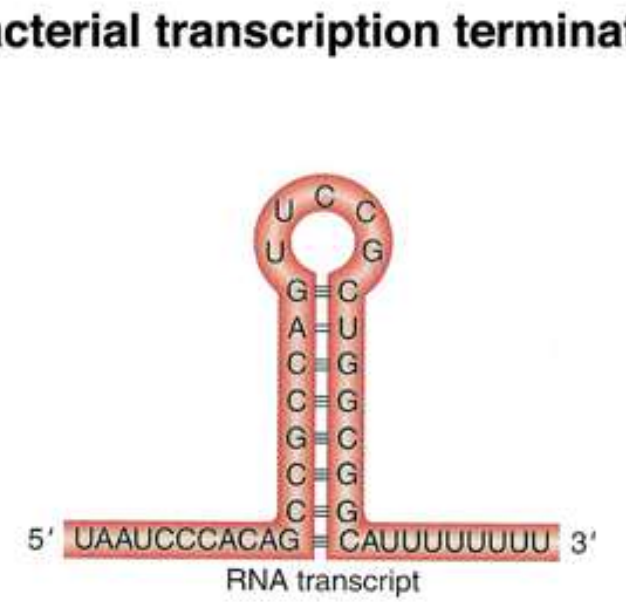
Eukaryotes use a set of general transcription factors that bind to the promoter region, then recruit other protein factors including RNA polymerase.

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



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.



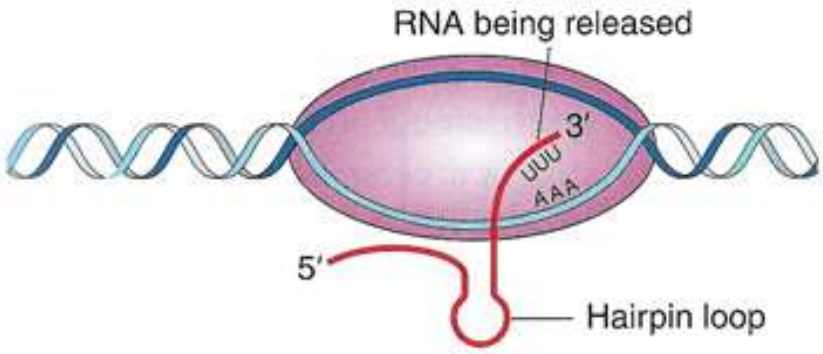
In the intrinsic mechanism, show above, there is a self-complementary sequence past the end of the coding sequence that forms a hairpin loop once it is transcribed. The base-paired part of the loop is very GC-rich, so the hairpin is stable. The presence of this structure interferes with RNA eleongation,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

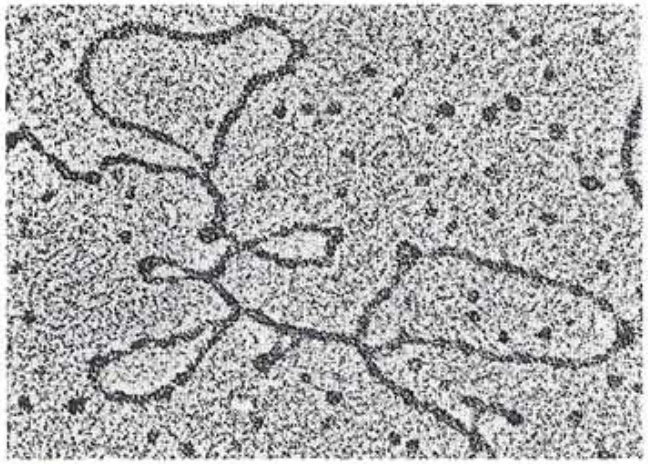
Bacterial termination: intrinsic mechanism



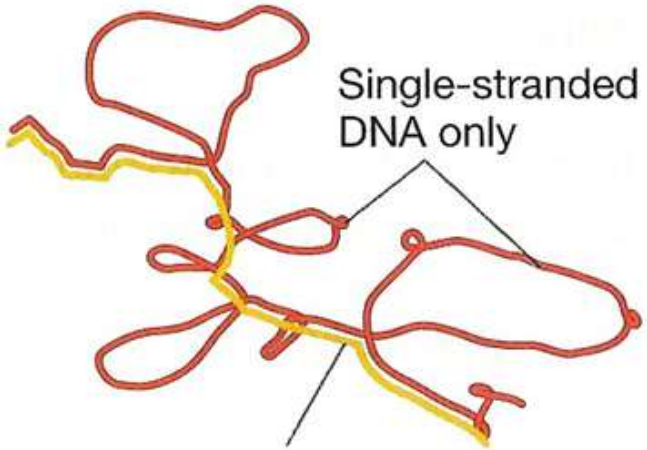
mRNA processing

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.

(a) Micrograph of DNA-RNA hybrid



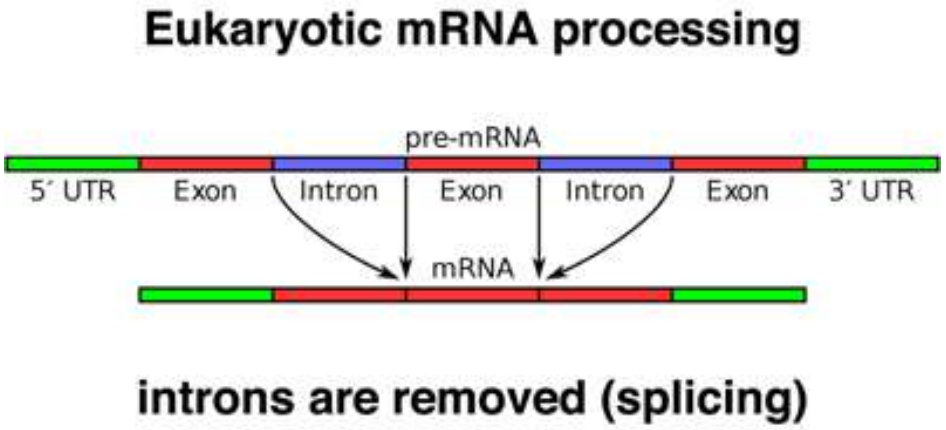
(b) Interpretation of micrograph



Eucaryotic genes are discontinuous!!!

mRNA processing

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

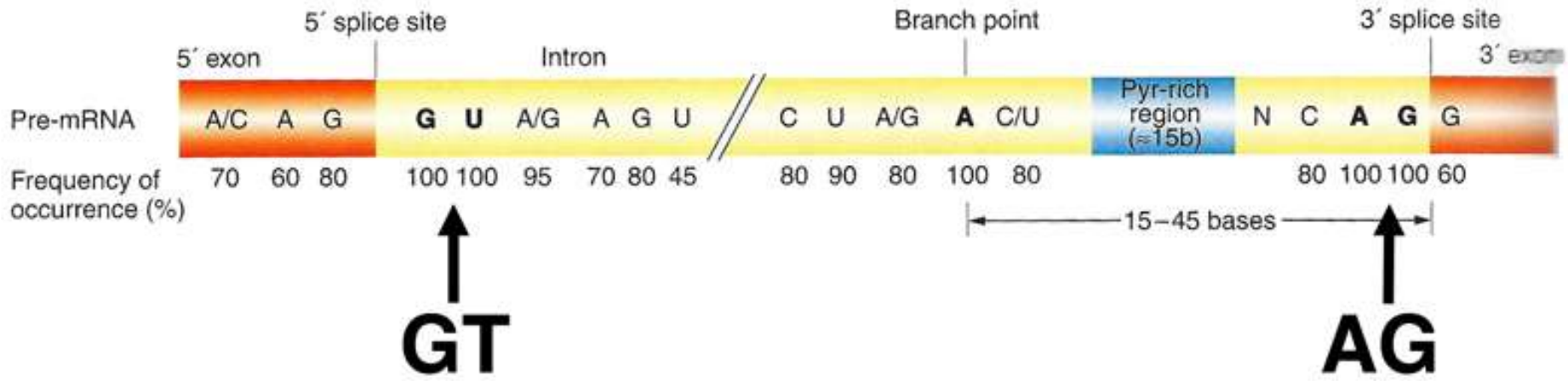


When a eukaryotic gene is transcribed, the primary transcript is processed in the nucleus in several ways. The most striking modification is splicing. Parts of the primary transcript, called introns, are spliced out of the mRNA. The remaining segments of mRNA are called exons.

https://www.youtube.com/watch?v=aVgwr0QpYNE

mRNA processing

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

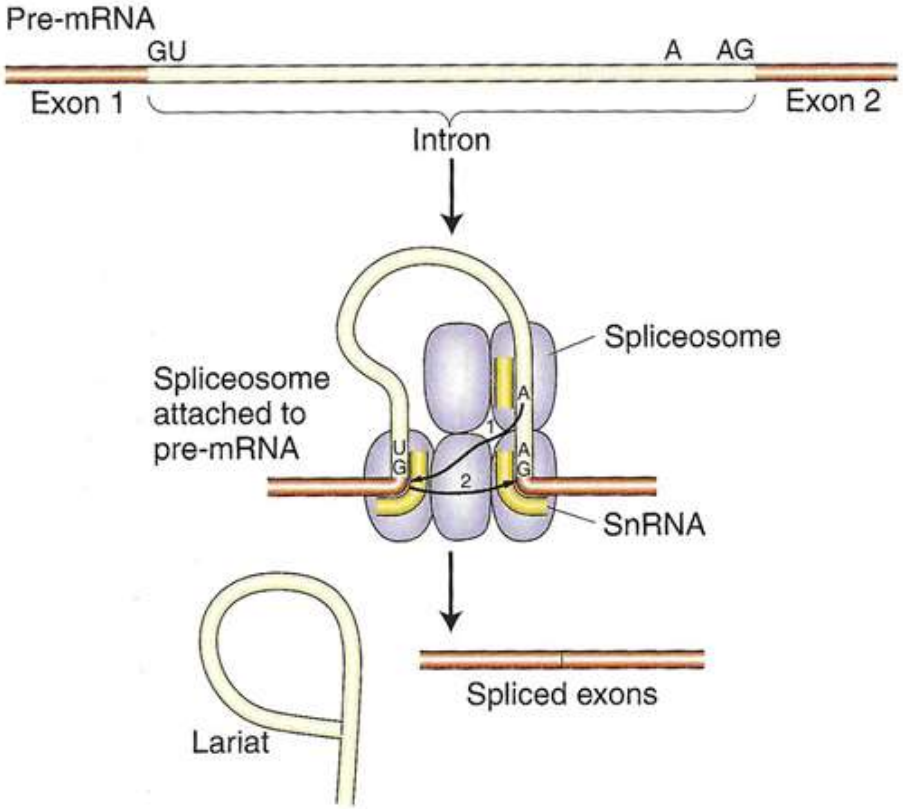


The 5' end of the intron begins with a splice donor site that almost always inclues GT as the first two bases of the intron (very rarely, it's GC).

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.

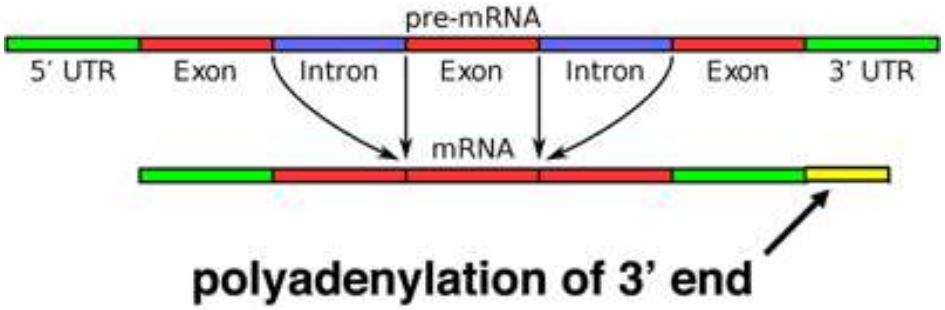
mRNA processing

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.

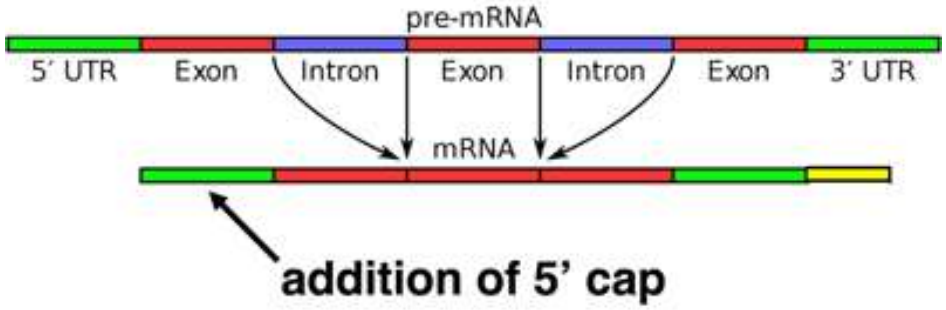


mRNA processing

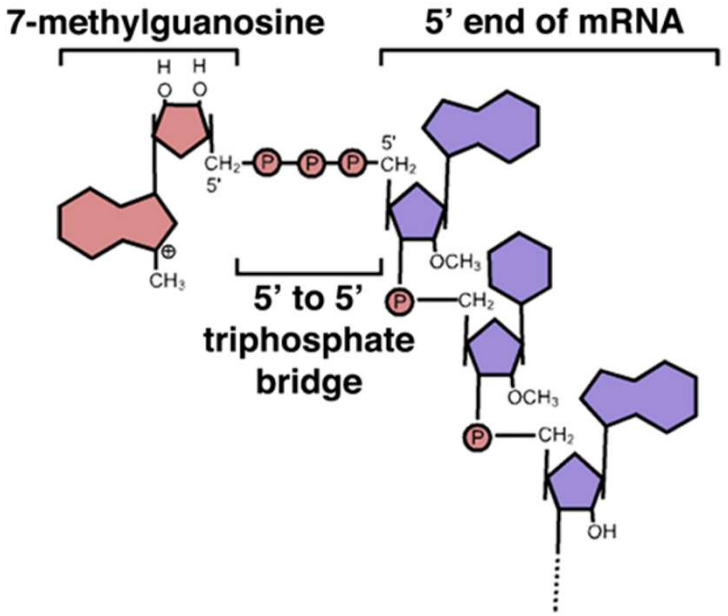
Eukaryotic mRNAs are also modified at the 3' end by the addition of a poly-A tail.



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

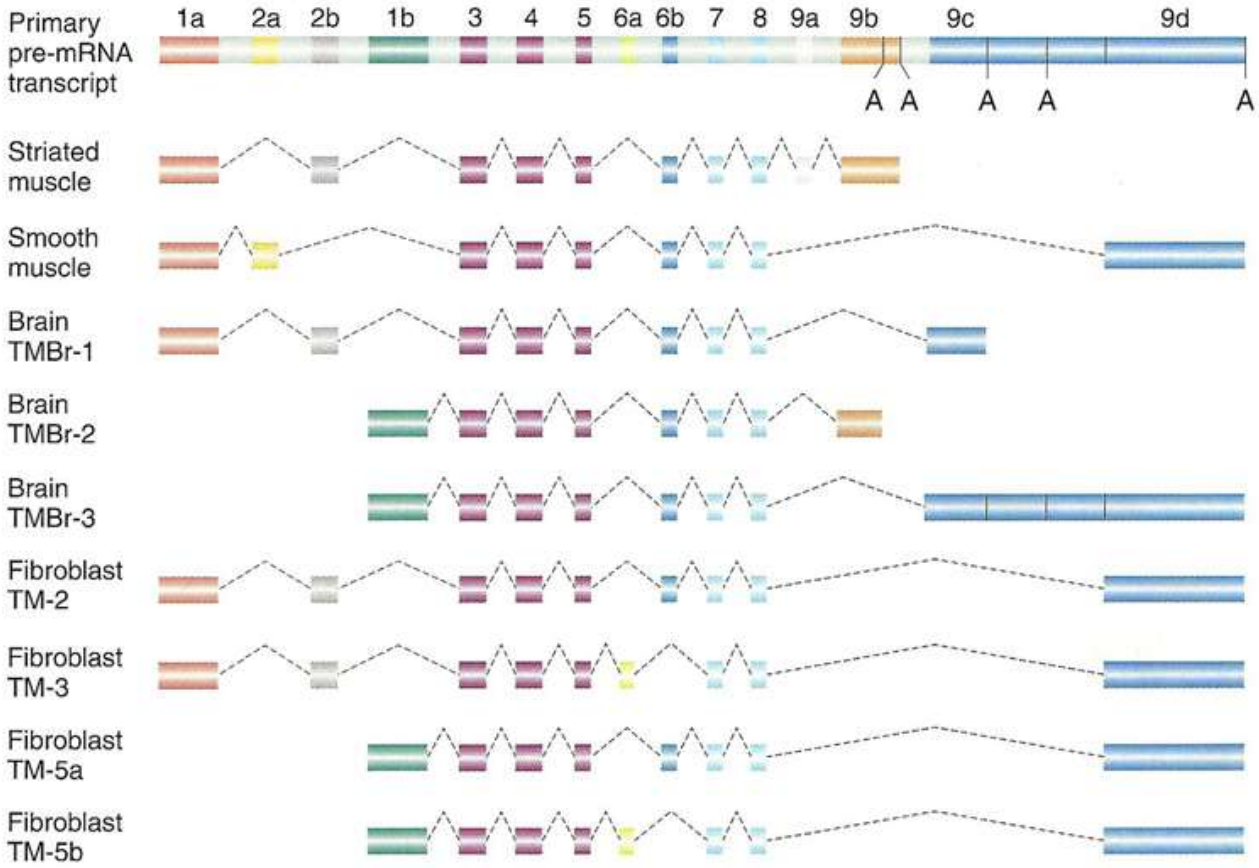


Eukaryotic 5' cap is a modified GTP

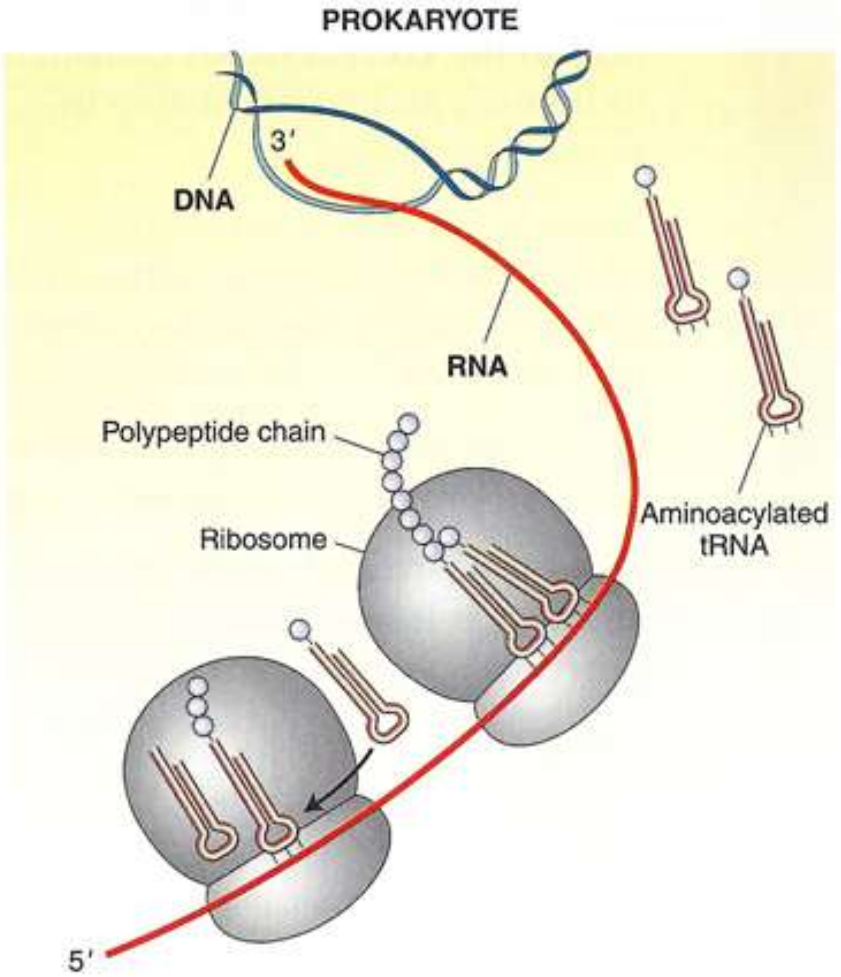


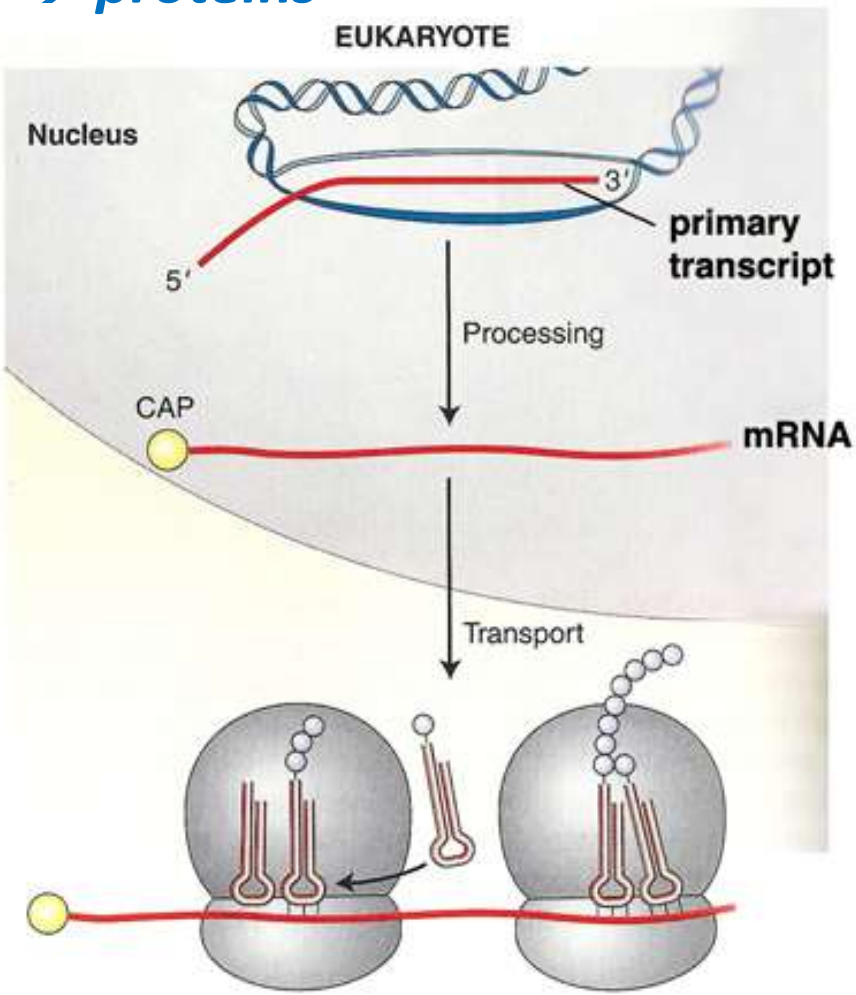
Alternative splicing - tropomyosin

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



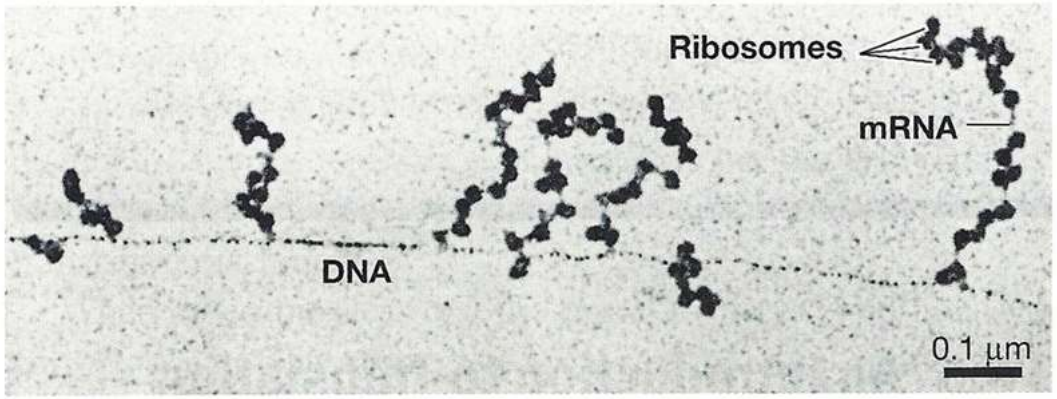


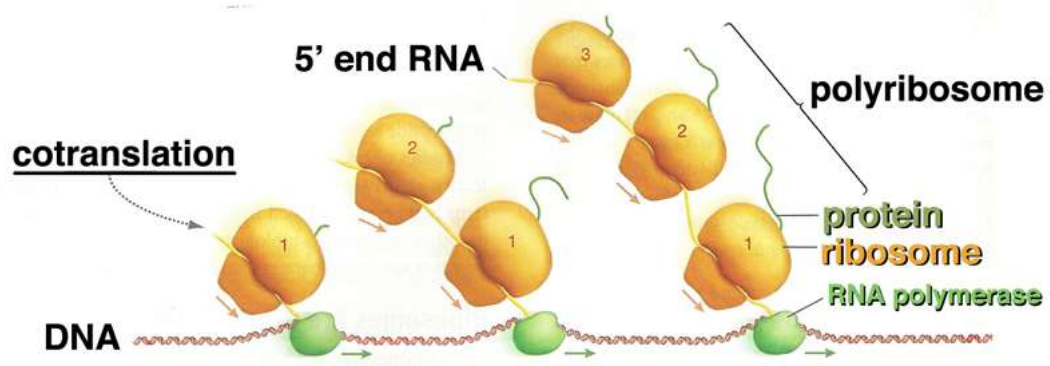
https://www.youtube.com/watch?v=TfYf_rPWUdY

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

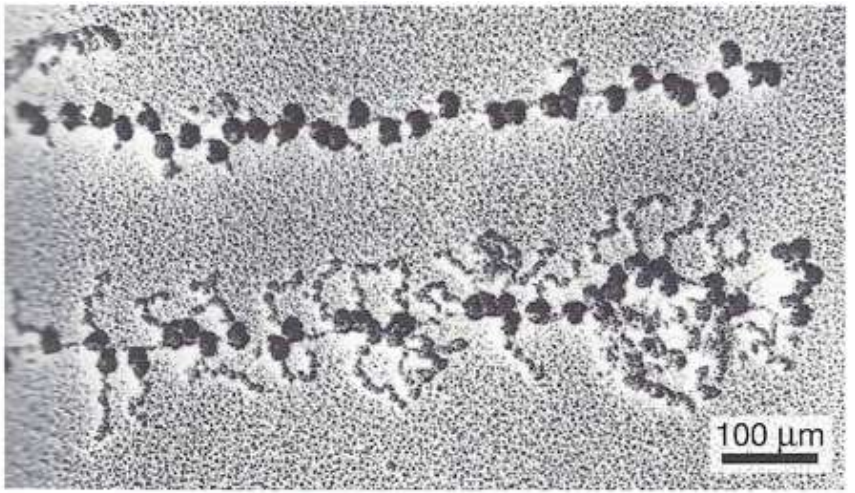
Translation: RNA → proteins

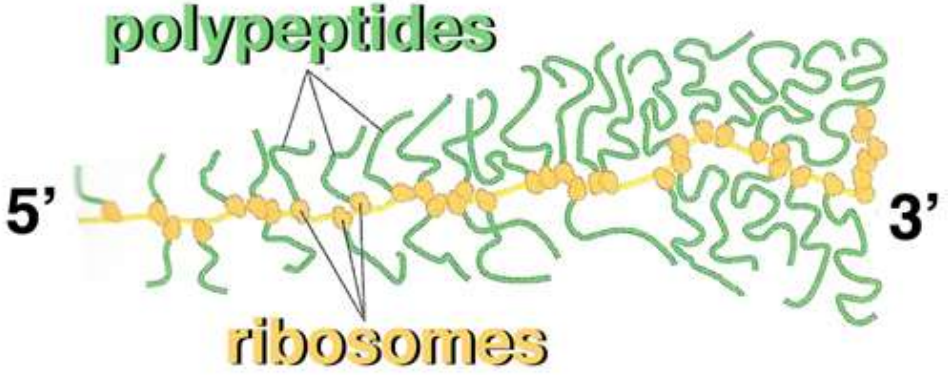
The coupling of transcription and translation in bacteria





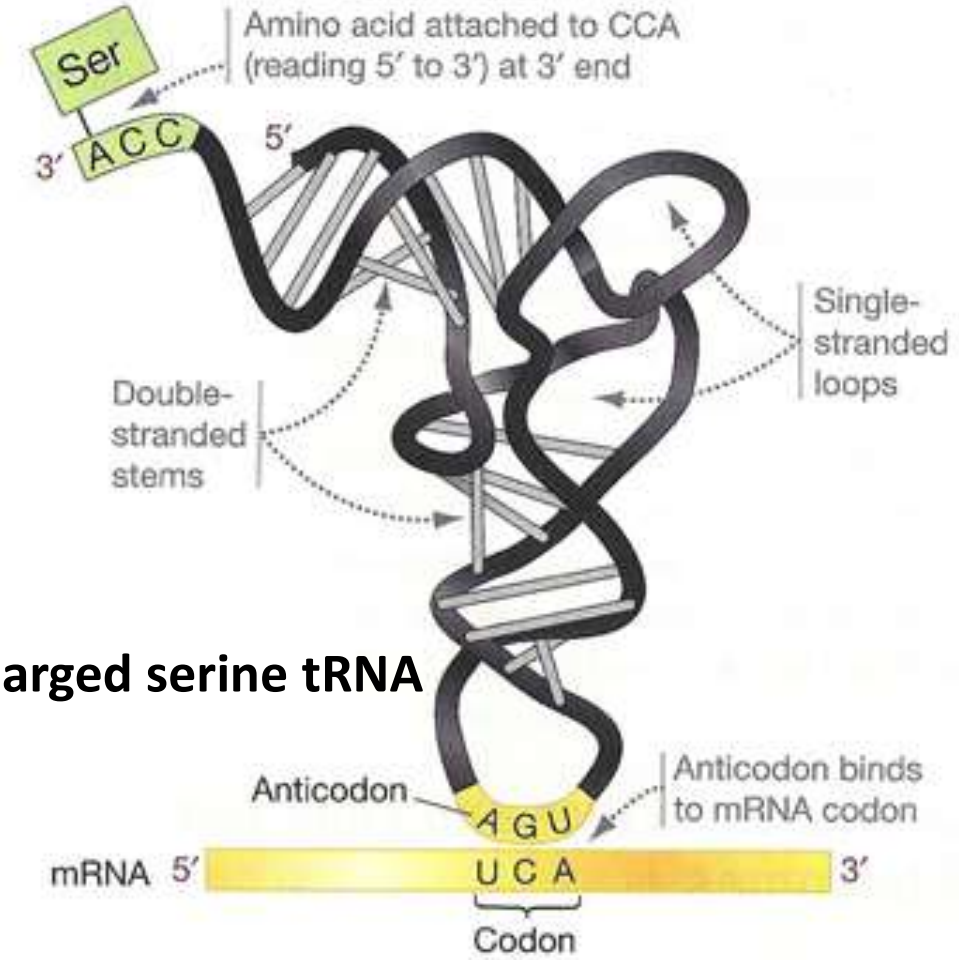
Eukaryotic polyribosomes





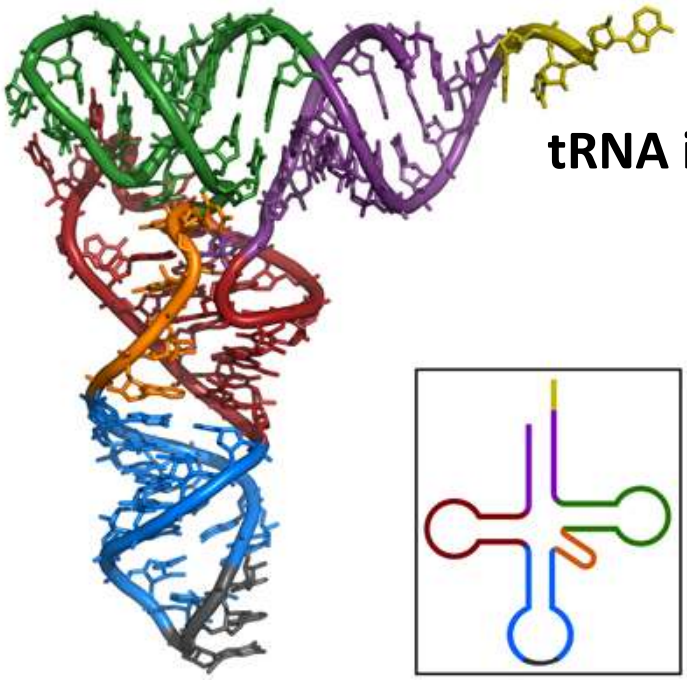
5' end of the mRNA, because there are shorter protein tails on the ribosomes at that end

Translation: RNA → proteins



Charged serine tRNA

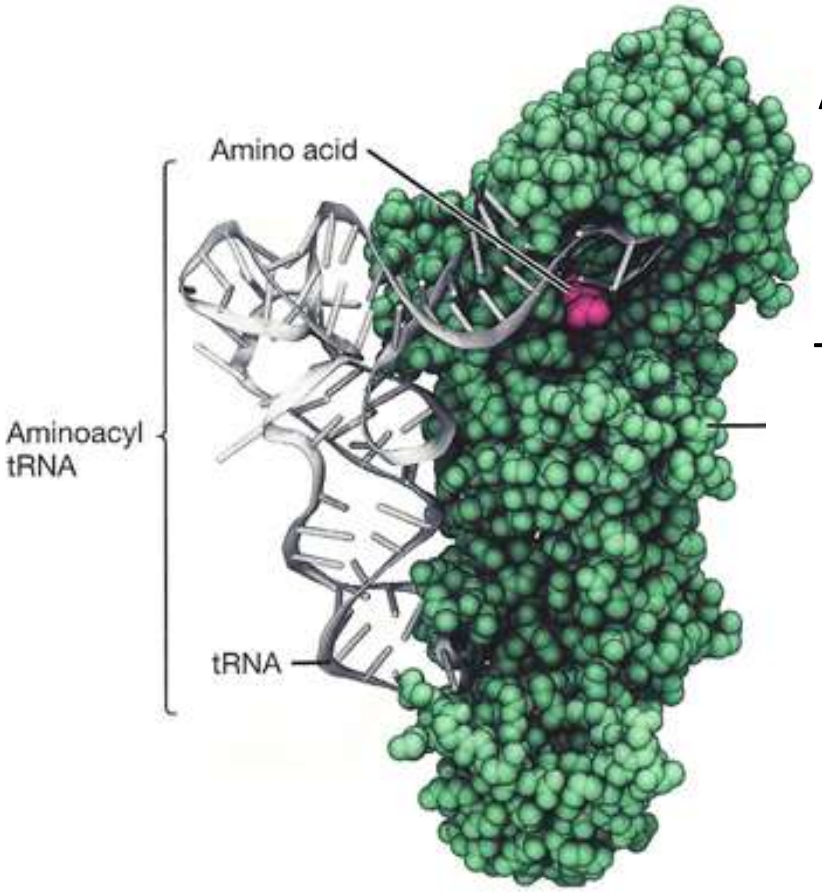
This shows a "charged" serine tRNA, covalently attached to the amino acid serine at its 3' end, with the anticodon paired to a serine codon



tRNA in 3D

Translation: RNA → proteins

Aminoacyl tRNA synthethase



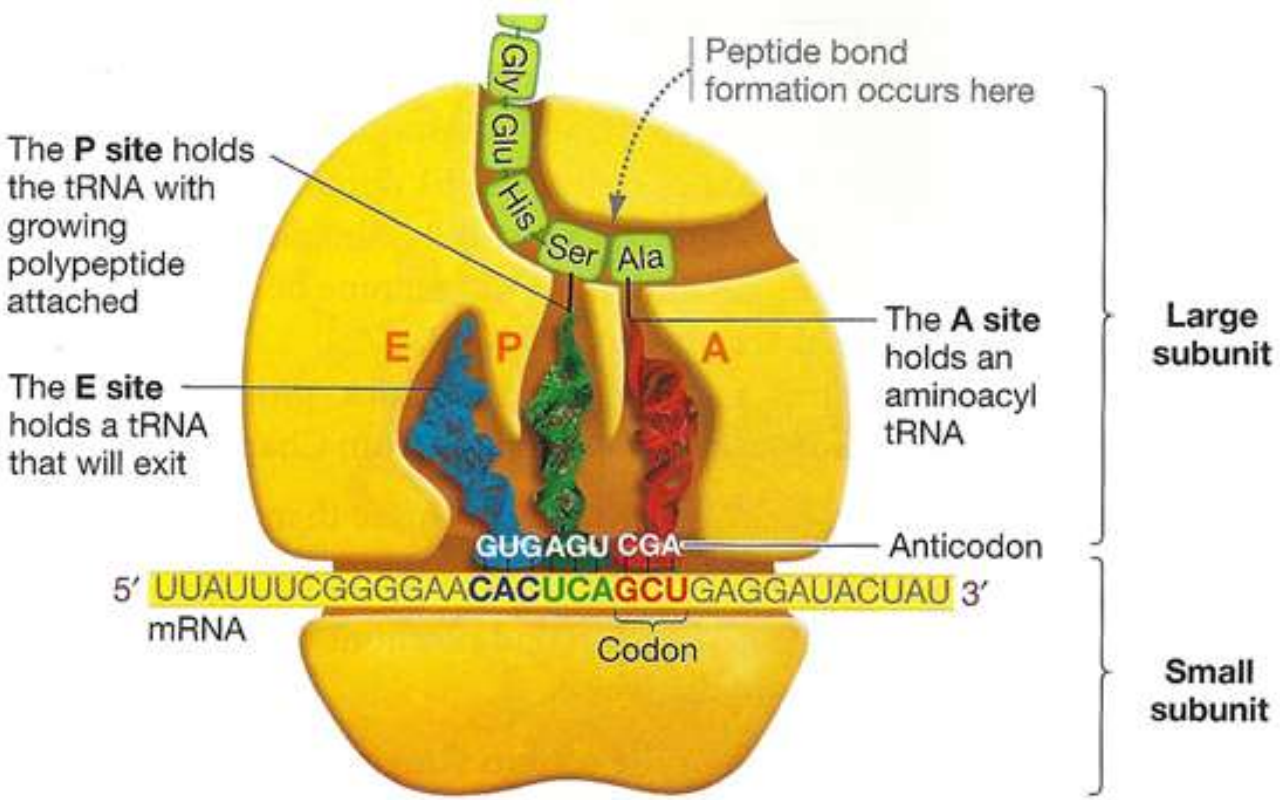
A special set of enzymes "charges" tRNAs, attaching the correct amino acid to particular tRNAs.

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

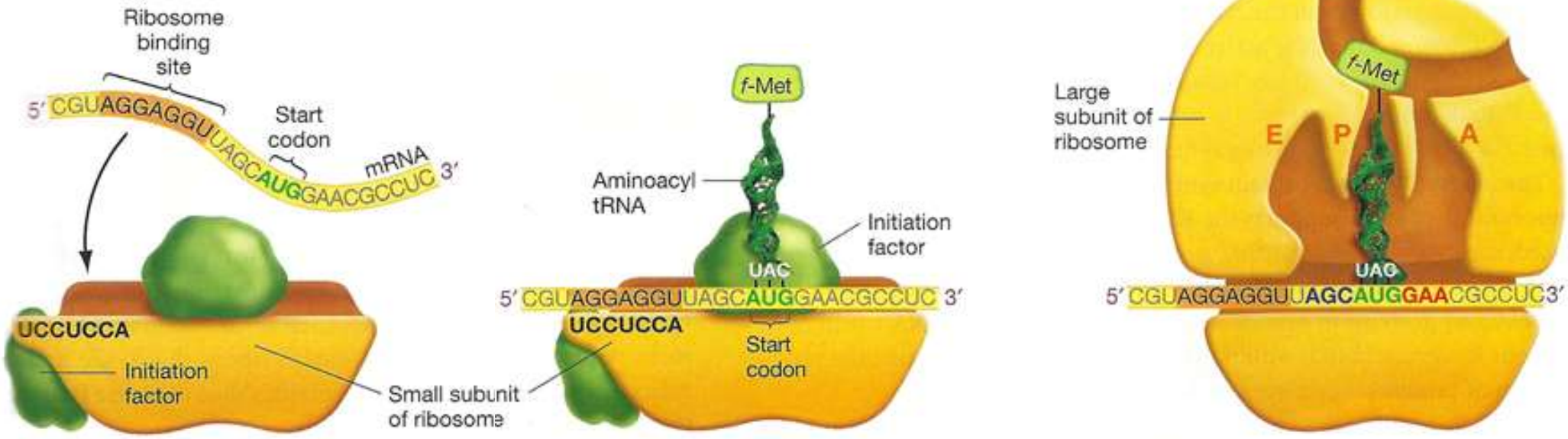
Translation: RNA → proteins



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



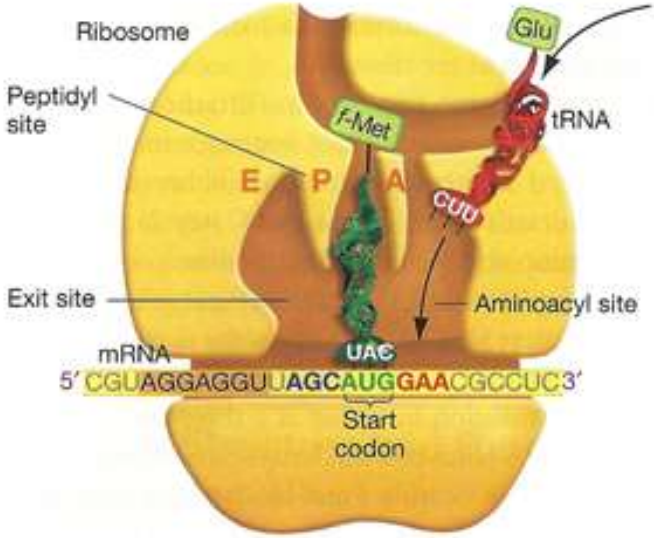
1. mRNA binds to small subunit. Ribosome binding site sequence binds to a complementary sequence in an RNA molecule in the small subunit of the ribosome, with the help of protein initiation factors.

2. Initiator aminoacyl tRNA binds to start codon.

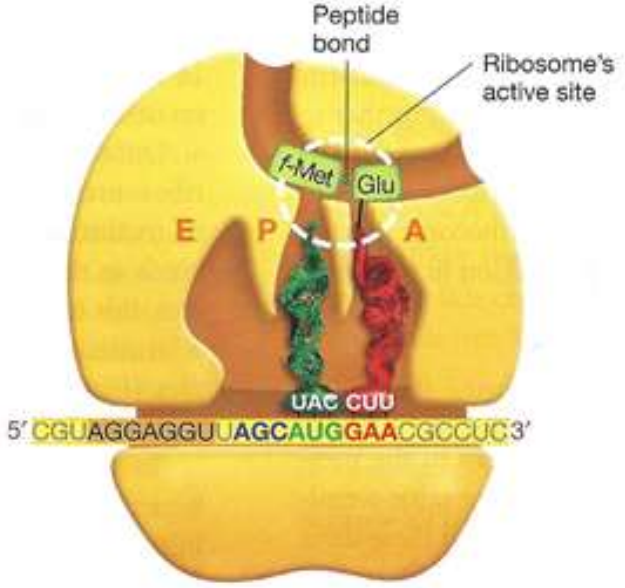
3. Large subunit of ribosome binds, completing ribosome assembly. Translation begins.

Translation: RNA → proteins

Translation elongation



1. Incoming aminoacyl tRNA
New tRNA moves into A site, where its anticodon base pairs with the mRNA codon.

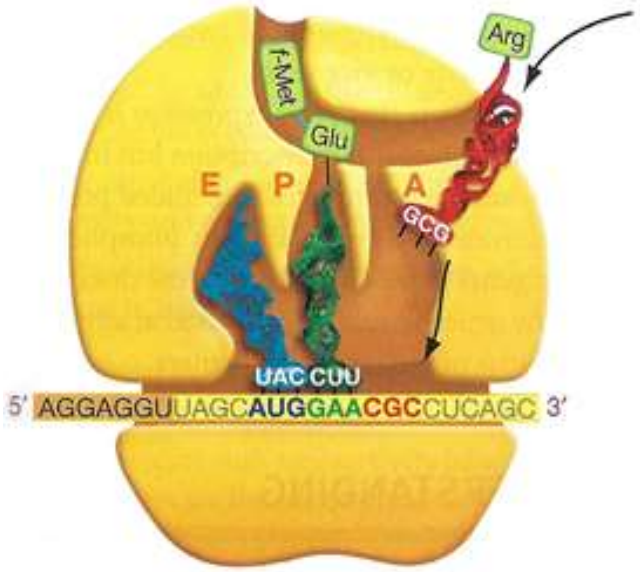


2. Peptide bond formation
The amino acid attached to the tRNA in the P site is transferred to the tRNA in the A site.

3. Translocation
mRNA is ratcheted through the ribosome by elongation factors (not shown). The tRNA attached to the polypeptide chain moves into the P site. The A site is empty.

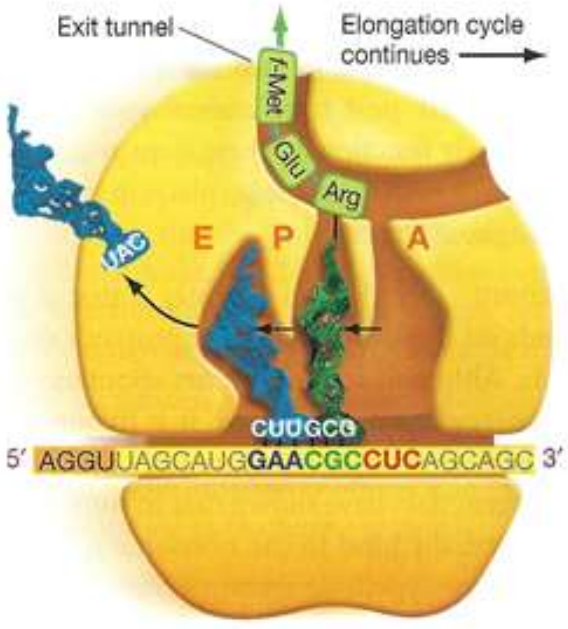
Translation: RNA → proteins

Translation elongation



4. Incoming aminoacyl tRNA
New tRNA moves into A site, where its anticodon base pairs with the mRNA codon.

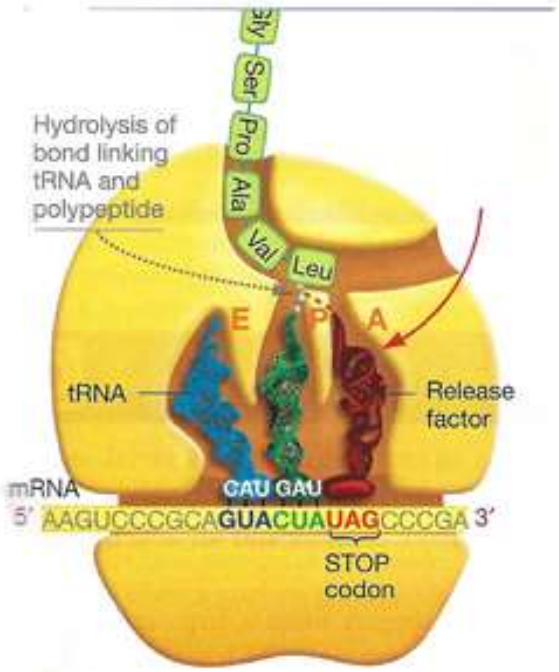
5. Peptide bond formation
The polypeptide chain attached to the tRNA in the P site is transferred to the aminoacyl tRNA in the A site.



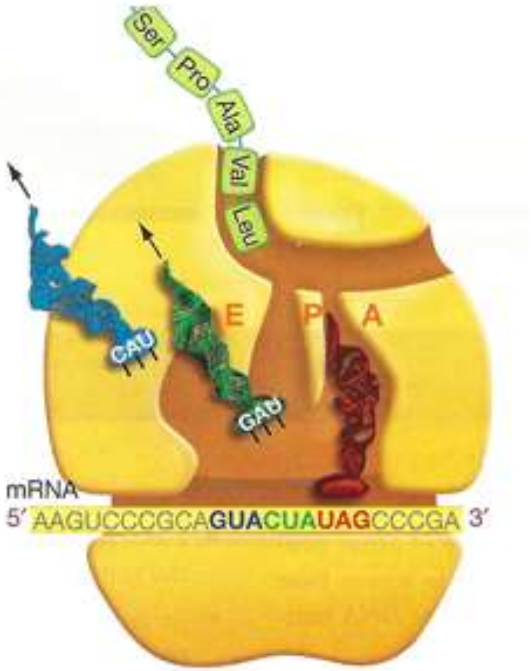
6. Translocation
mRNA is ratcheted through the ribosome again. The tRNA attached to polypeptide chain moves into P site. Empty tRNA from P site moves to E site, where tRNA is ejected. The A site is empty again.

Translation: RNA → proteins

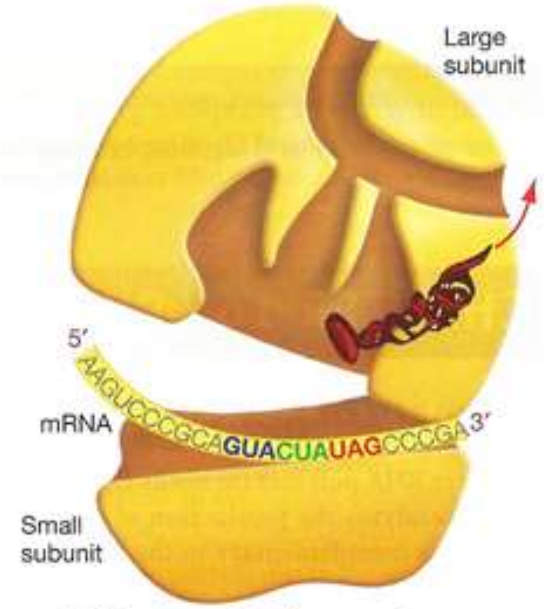
Translation termination



1. Release factor binds to stop codon. When translocation exposes a stop codon, a release factor fills the A site. The release factor breaks the bond linking the tRNA in the P site to the polypeptide chain.



2. Polypeptide is released. The hydrolysis reaction frees the polypeptide, which is released from the ribosome. The empty tRNAs are released either along with the polypeptide or...



3. Ribosome subunits separate. ...when the ribosome separates from the mRNA, and the two ribosomal subunits dissociate. The subunits are ready to attach to the start codon of another message and start translation anew.

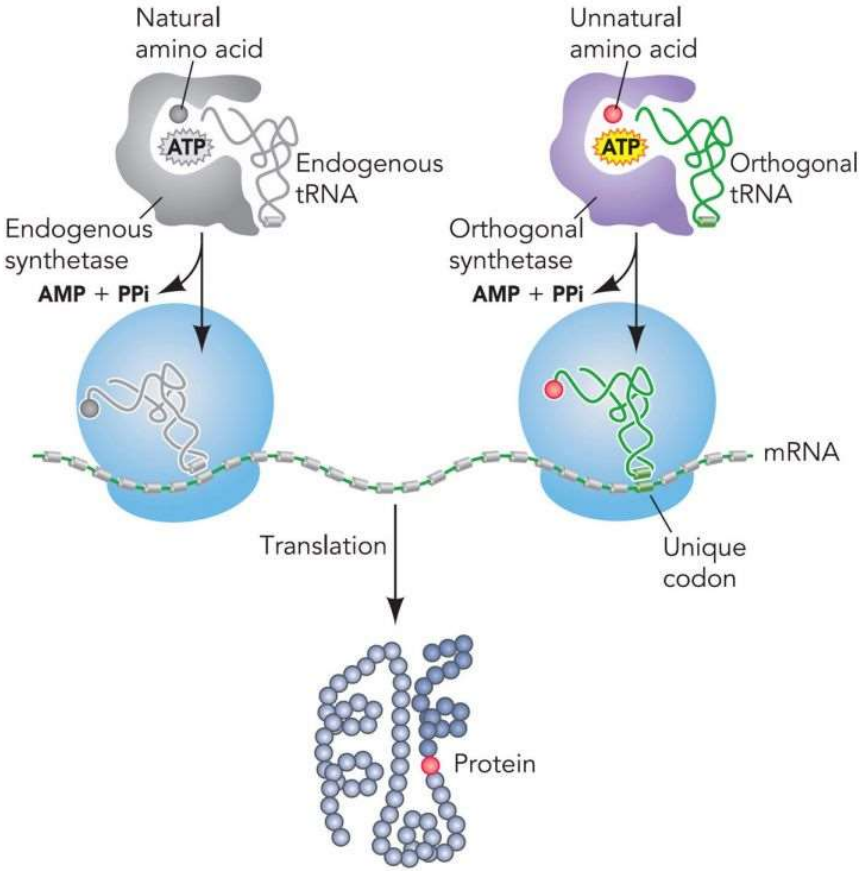
Translation: RNA → proteins – the genetic code

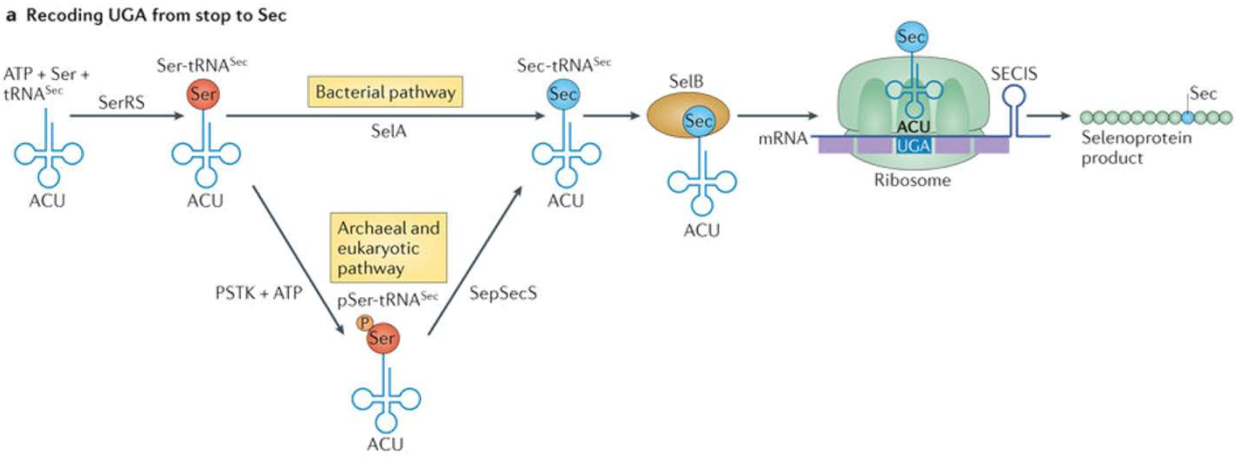
nonpolar polar basic acidic (stop codon)

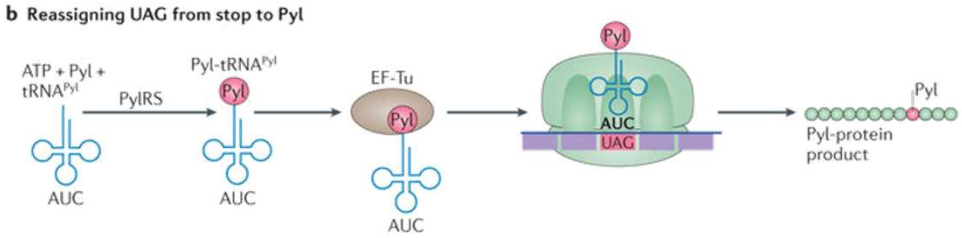
Standard genetic code

1st base	2nd base								3rd base
	U		C		A		G		
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		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

Expanded genetic code

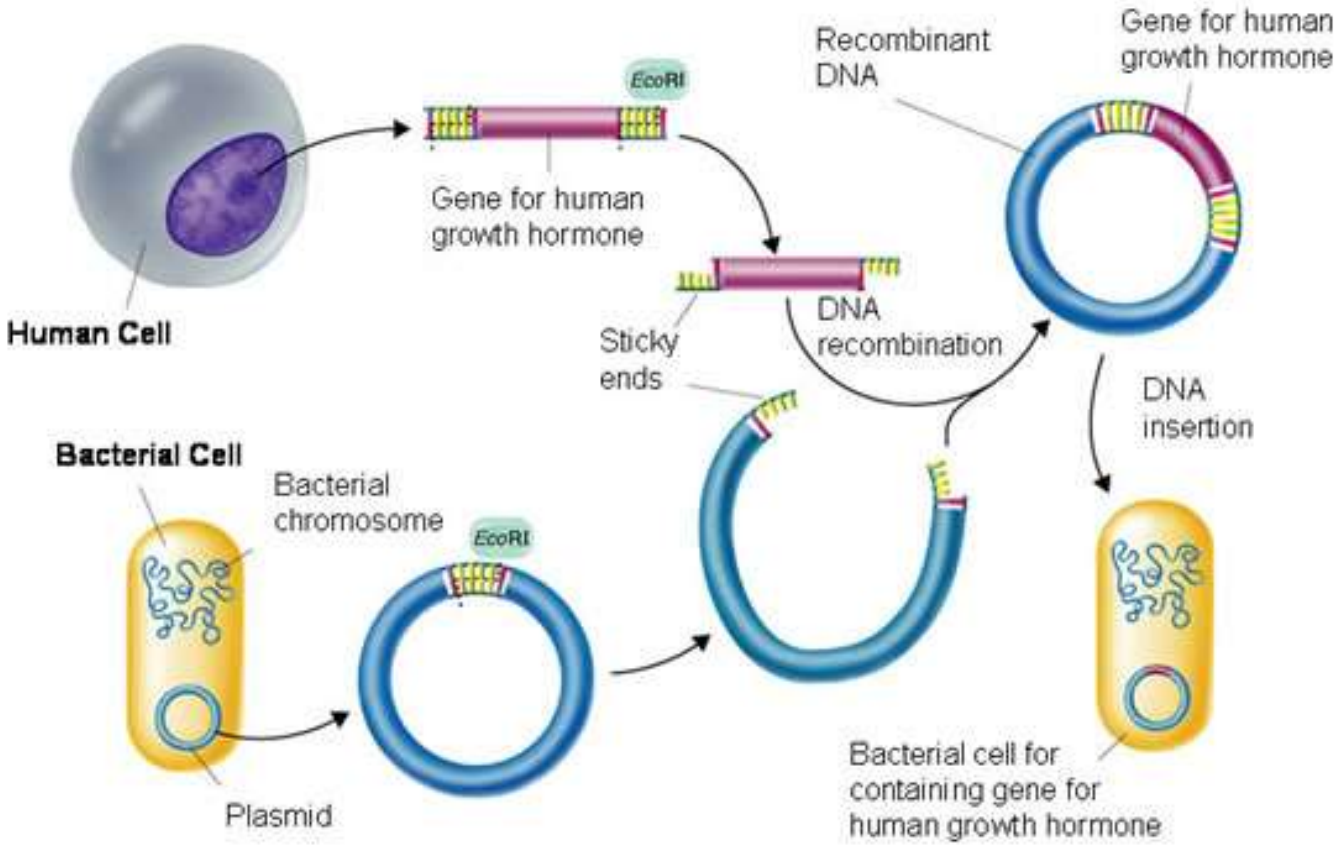






Nature Reviews | Microbiology

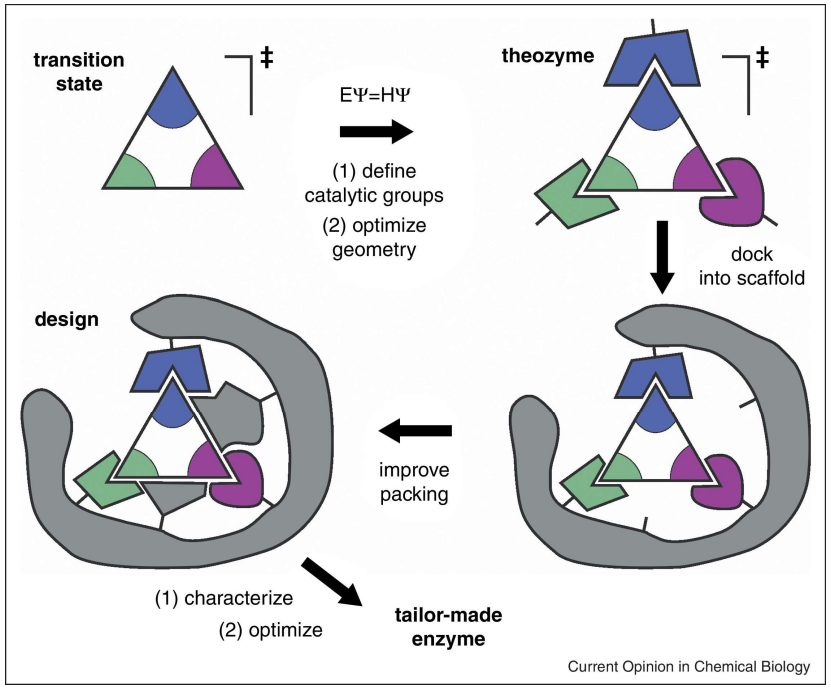
Recombinant proteins

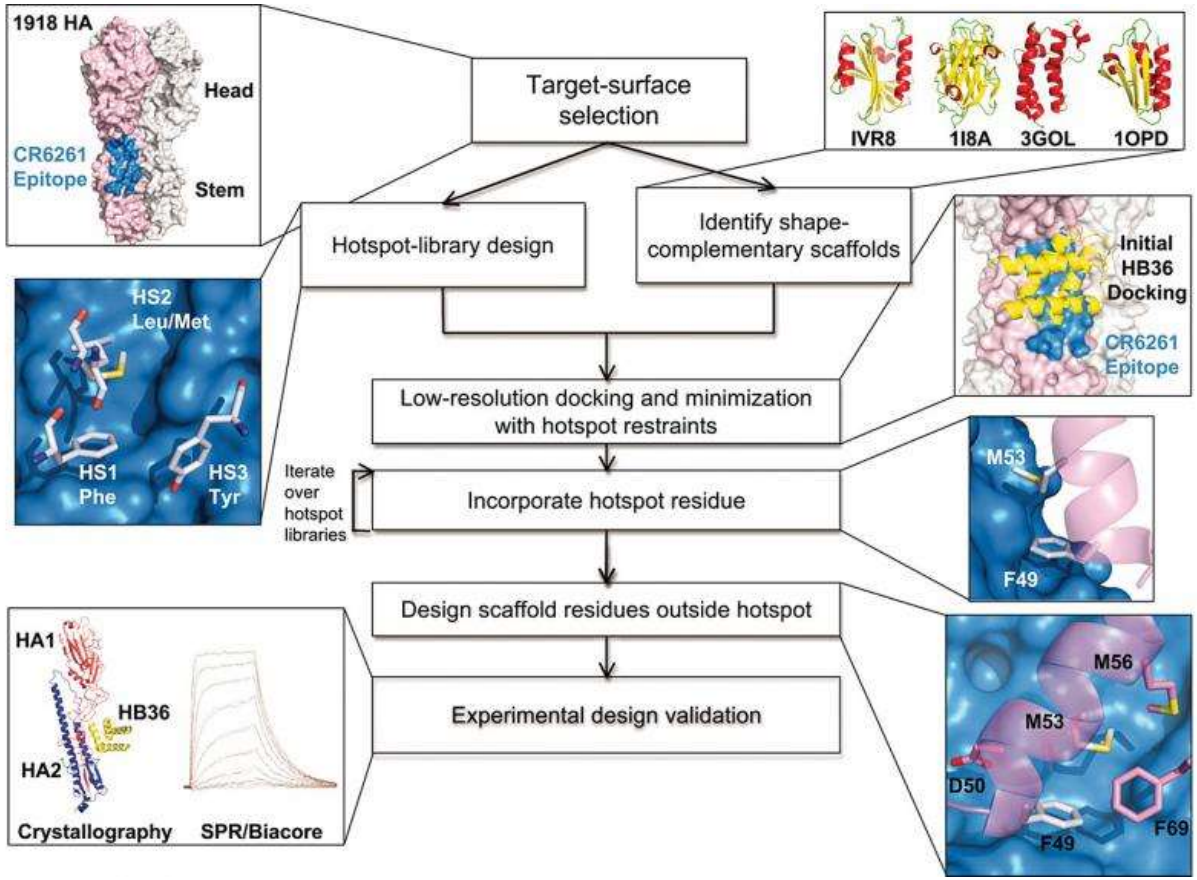


Recombinant insulin:

https://www.youtube.com/watch?v=glt8iAqK8NQ&ab_channel=ScienceForStudent

de novo enzyme design

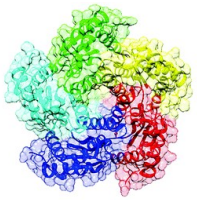




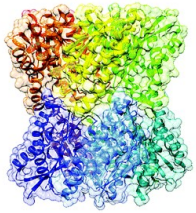
For interested in details:

Introduction to protein design – Part1: https://www.youtube.com/watch?v=0LetJMbu7uY&ab_channel=iBiology

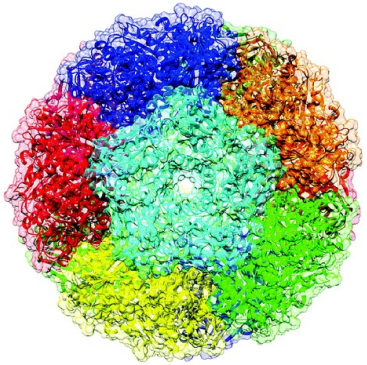
Introduction to protein design – Part2: https://www.youtube.com/watch?v=ZrAwWx7meTk&ab_channel=iBiology



pentamer
Schizosaccharomyces pombe

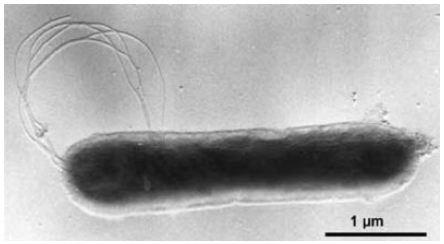


dimer of pentamers
Brucella abortus

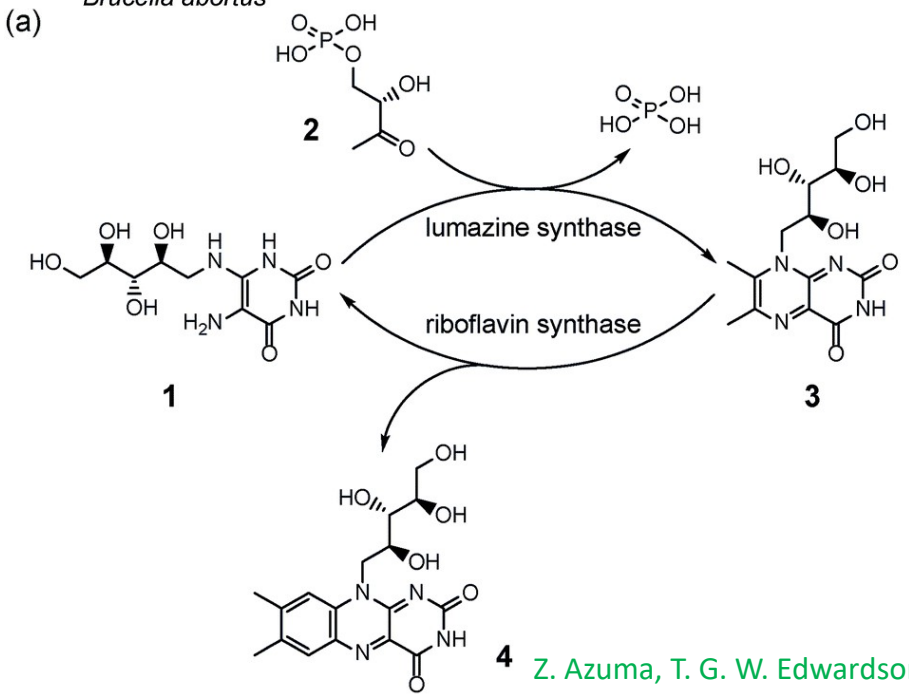


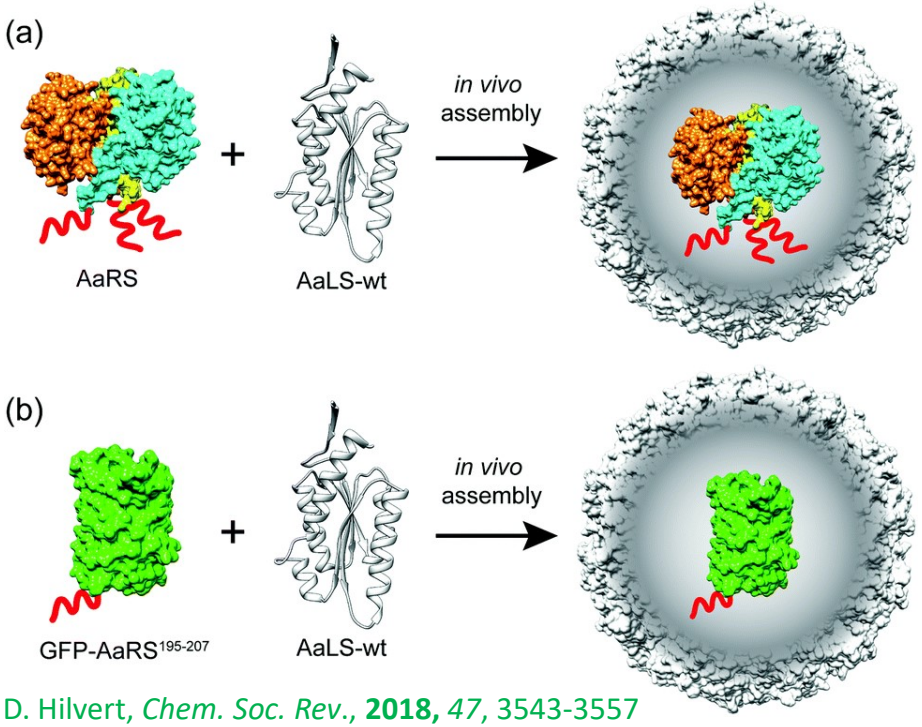
dodecahedron
formed from 12 pentamers
Aquifex aeolicus

Evolvability of proteins from thermophiles

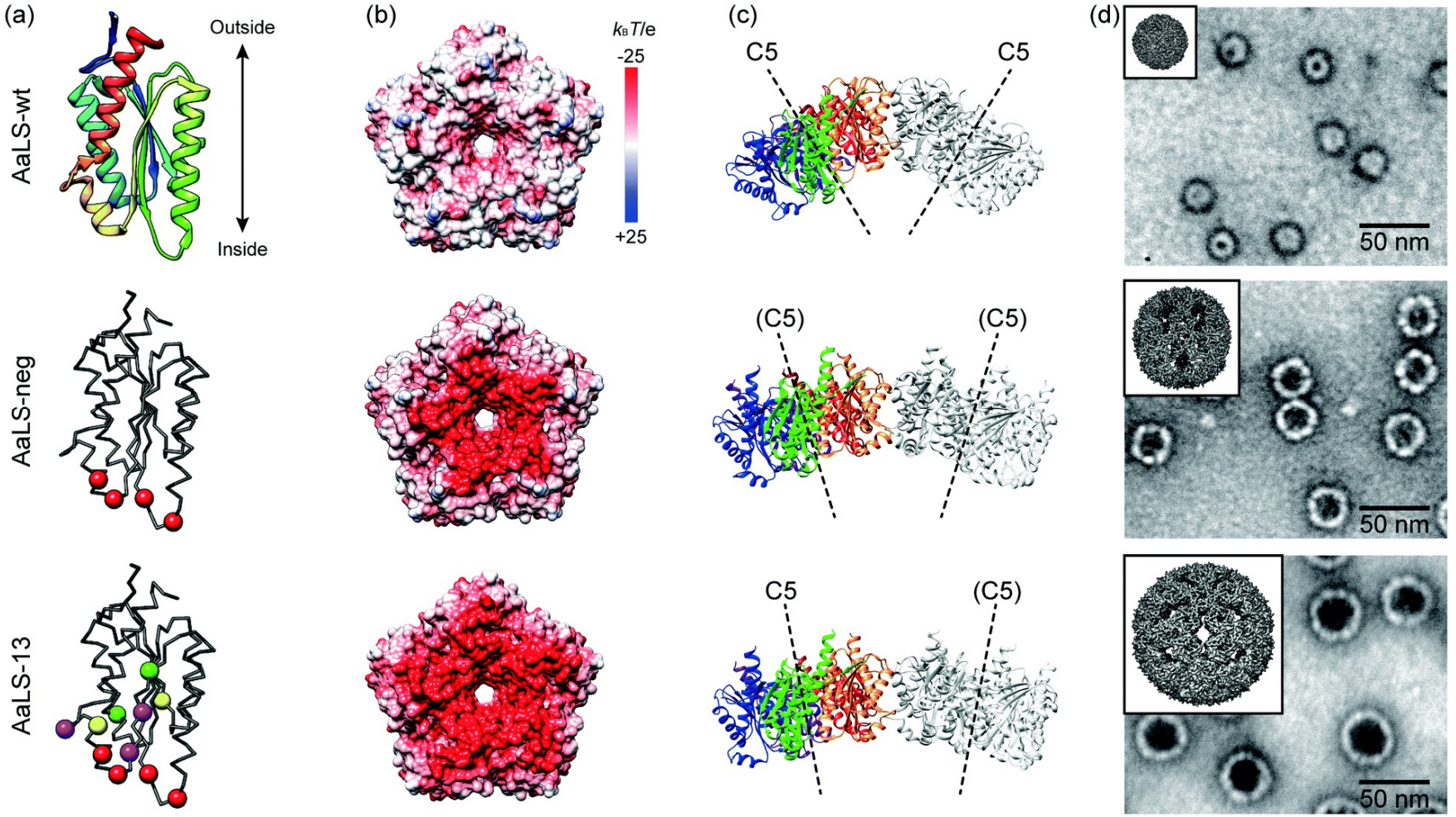


Aquifex aeolicus (Aa) – thermophilic bacteria

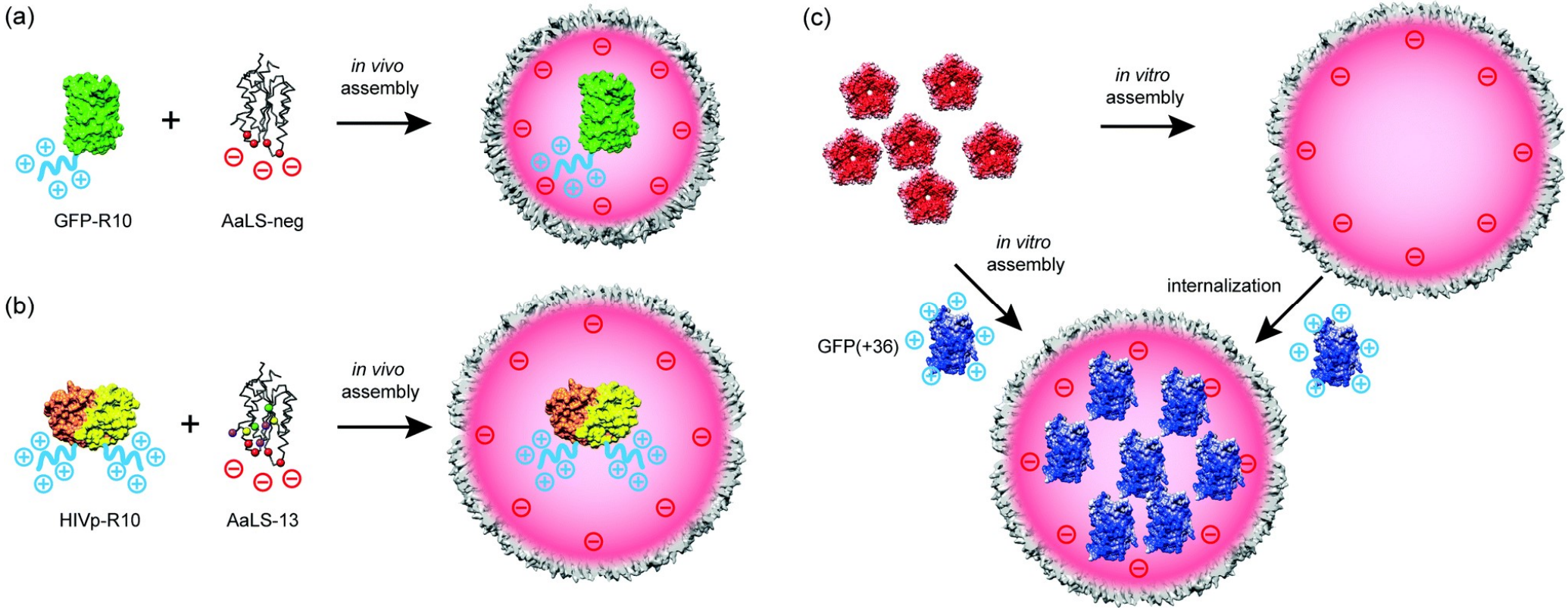




Z. Azuma, T. G. W. Edwardson and D. Hilvert, *Chem. Soc. Rev.*, **2018,** *47*, 3543-3557



B. Wörsdörfer, K. J. Woycechowsky and D. Hilvert, *Science*, **2011**, *331*, 589–592



B. Wörsdörfer, Z. Pianowski and D. Hilvert, *J. Am. Chem. Soc.*, **2012**, *134*, 909–911

Overview of the course

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

unnatural base pairing – expansion of the genetic alphabet;

artificial genetic polymers and oligonucleotide analogues (XNA);

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;

design of artificial organisms