Synthetic life

(continuation of „The molecular origins of life” SoSe 2022)

NaturalNews.com

WiSe 2022/23

Zbigniew Pianowski

7 lectures (90 min. each) in English

1st lecture: 26th Oct. 2022 (Wed.)
Zoom:
https://us02web.zoom.us/j/84205962071?pwd=Y2VSL0RaOC90bllqeEZrM3N3YmExdz09

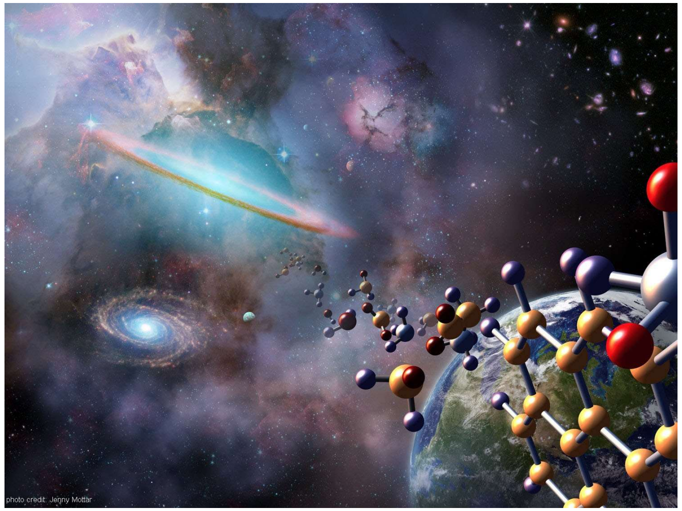
Following lecture terms 14:00-15:30, Wed.:
2.11, 9.11., 16.11., 23.11., 30.11. and 7.12.2022 (backup date 14.12.2022).

The most current dates, handouts – on the website:
https://www.ioc.kit.edu/pianowski/99.php

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?

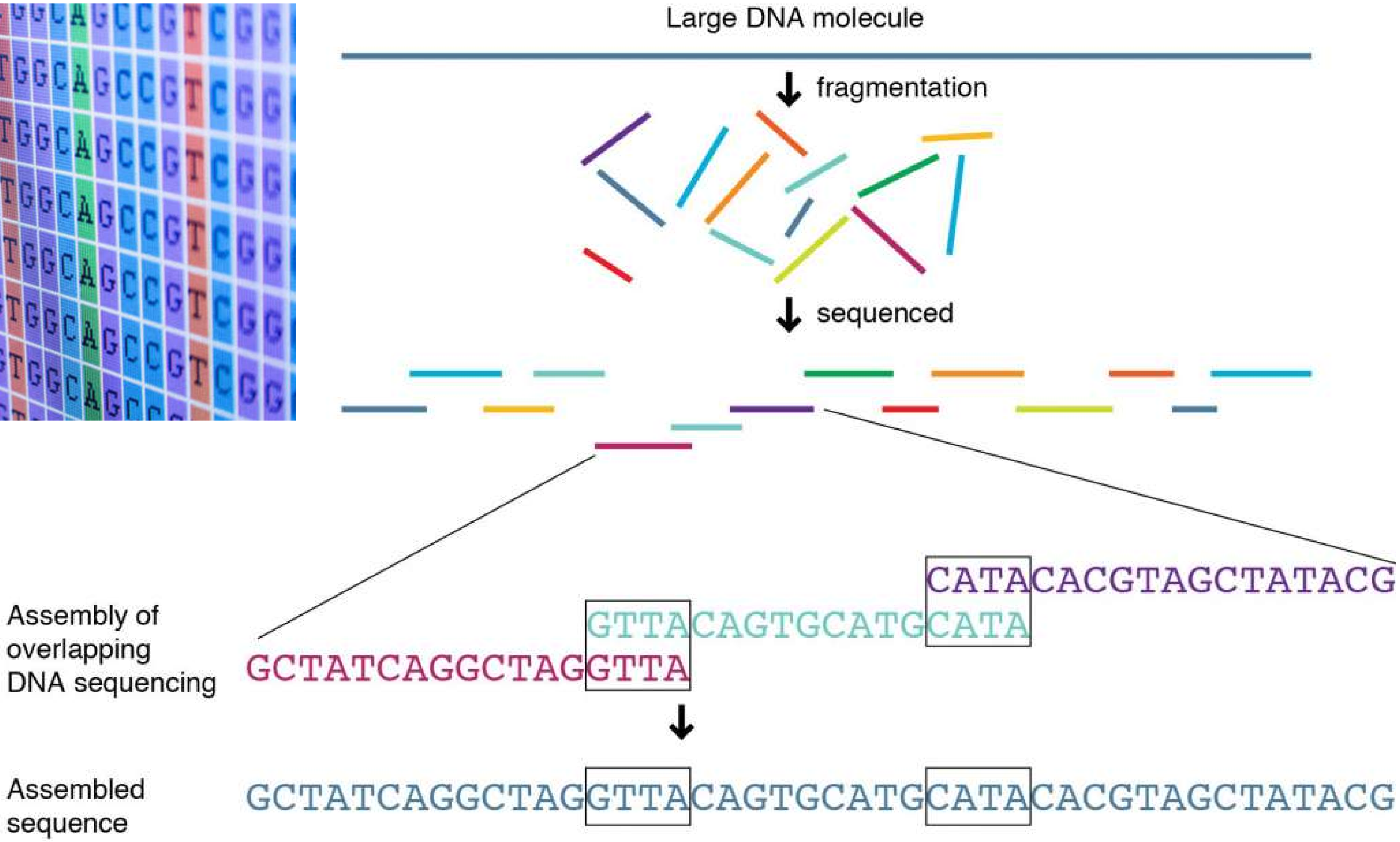
Everything – living or not – is just chemicals made of atoms.

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



Large DNA molecule

fragmentation

sequenced

Assembly of
overlapping
DNA sequencing

CATACACGTAGCTATACG

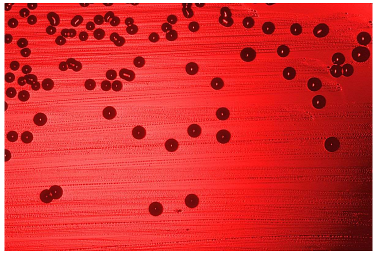
GTTACAGTGCATGCATA

GCTATCAGGCTAGGTTA

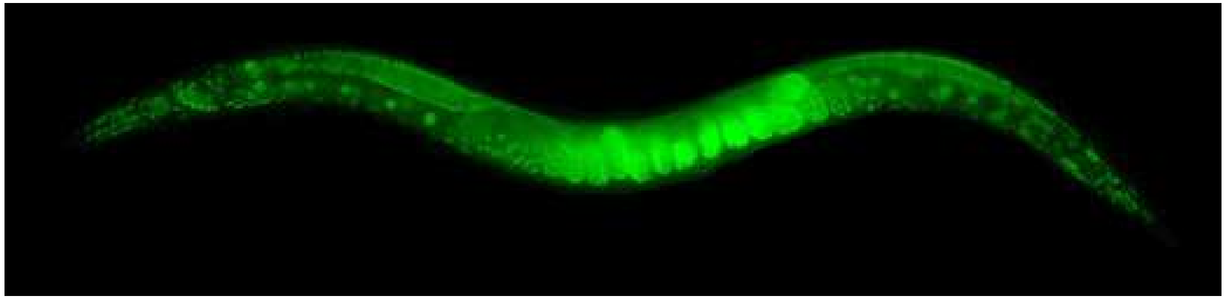
Assembled
sequence

GCTATCAGGCTAGGTTACAGTGCATGCATACACGTAGCTATACG

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



Haemophilius influenzae
1995



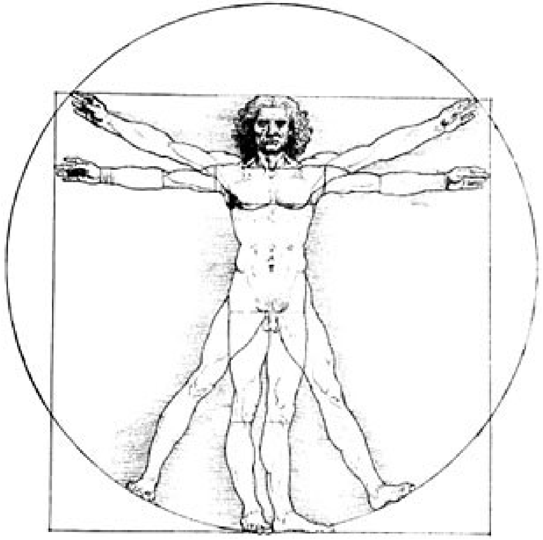
Caenorhabdis elegans
1998

Drosophila melanogaster
2000

Arabidopsis thaliana
2000

... and way up to mammals and human

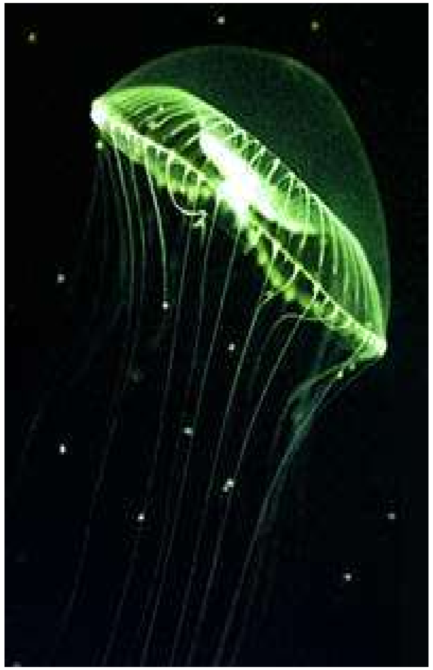
Mus musculus
2002



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

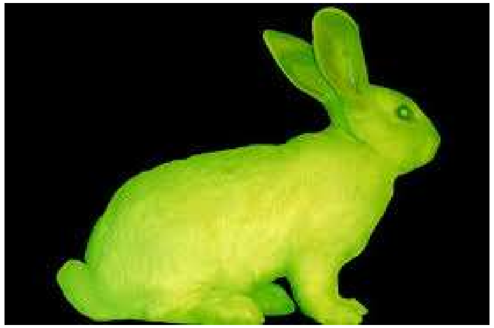
With that knowledge, scientists begun to tinker...

... 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?

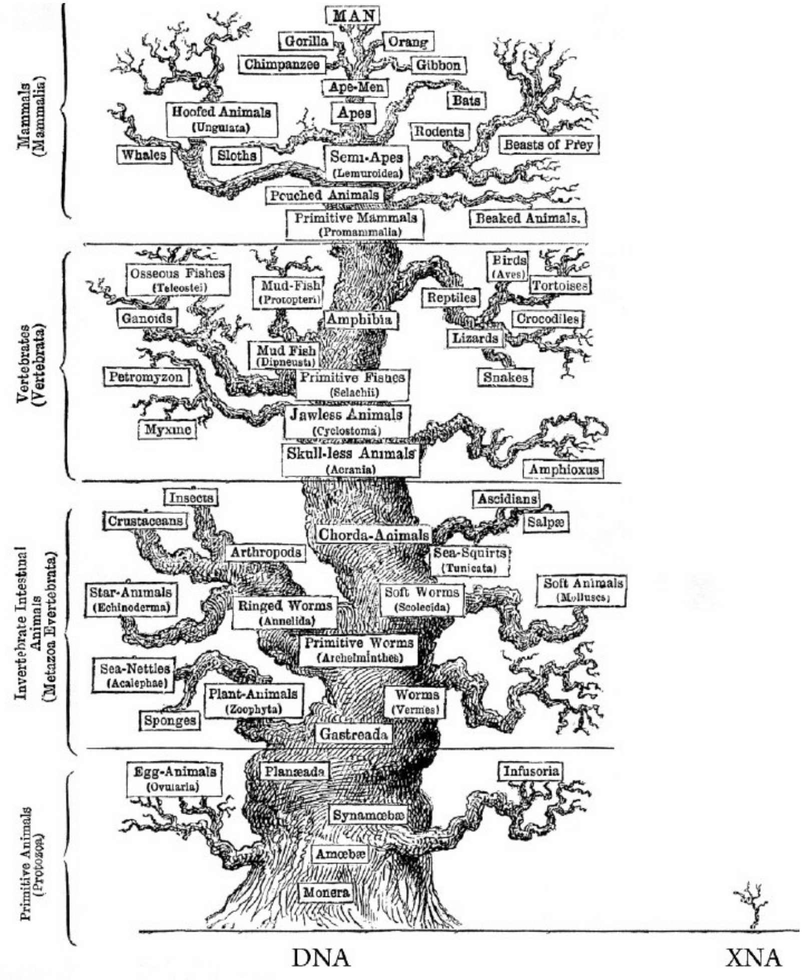
We can improve what was spelled out
for the 3,5 Billion years of evolution.

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);

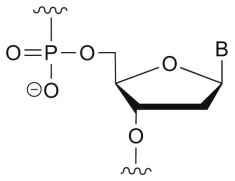
biosynthetic incorporation of **unnatural aminoacids (UAAs)** into proteins;

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

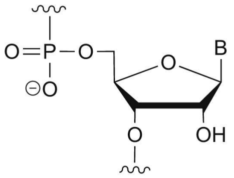
Artificial lipid vesicles as models for protocell multiplication;

design of artificial organisms

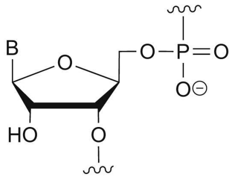
Artificial genetic polymers



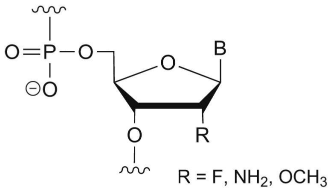
DNA



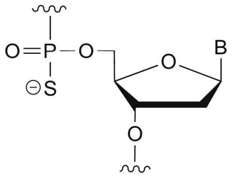
D-RNA
(natural)



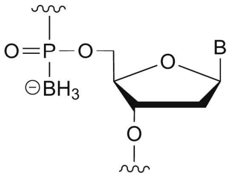
L-RNA
(unnatural)



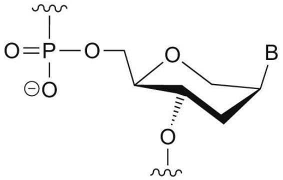
2'-modified RNA



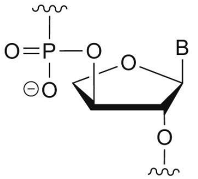
Phosphorothioate



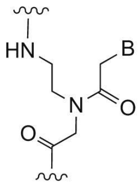
Boranophosphate



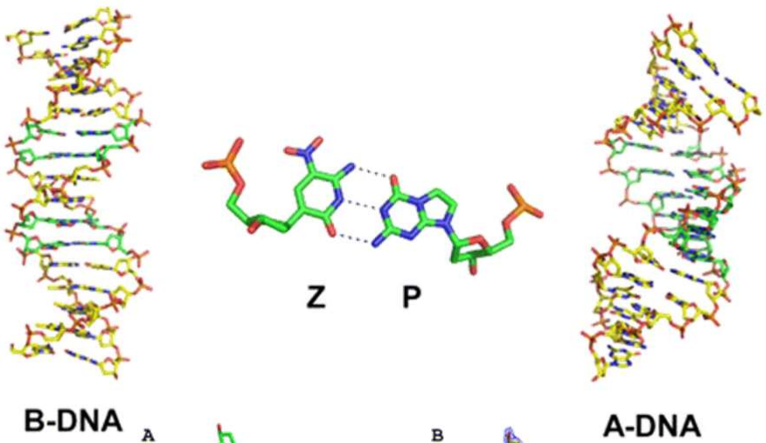
Hexitol Nucleic Acid
(HNA)

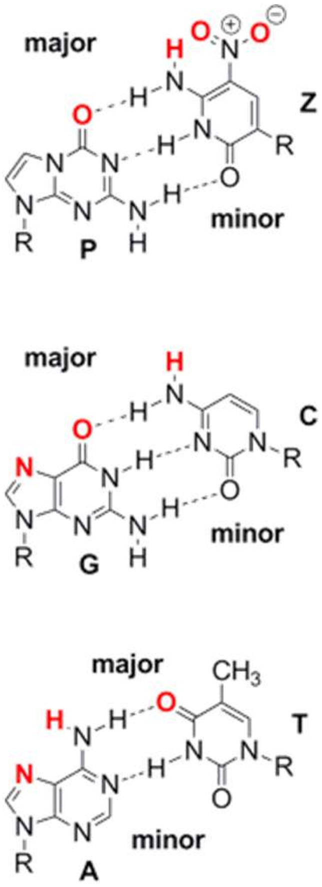


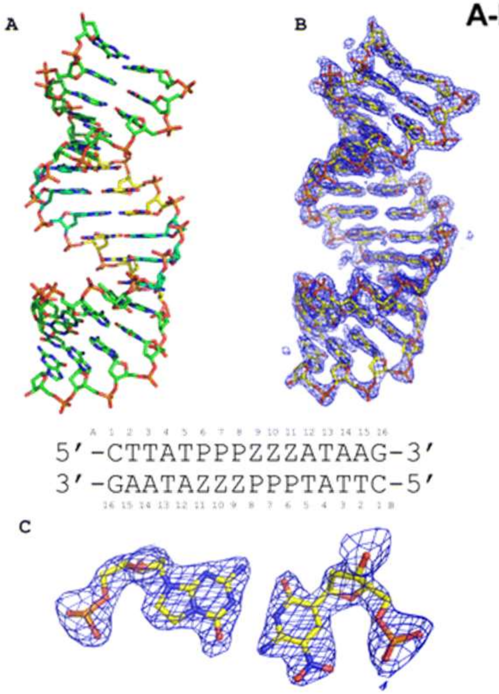
Threose Nucleic Acid
(TNA)



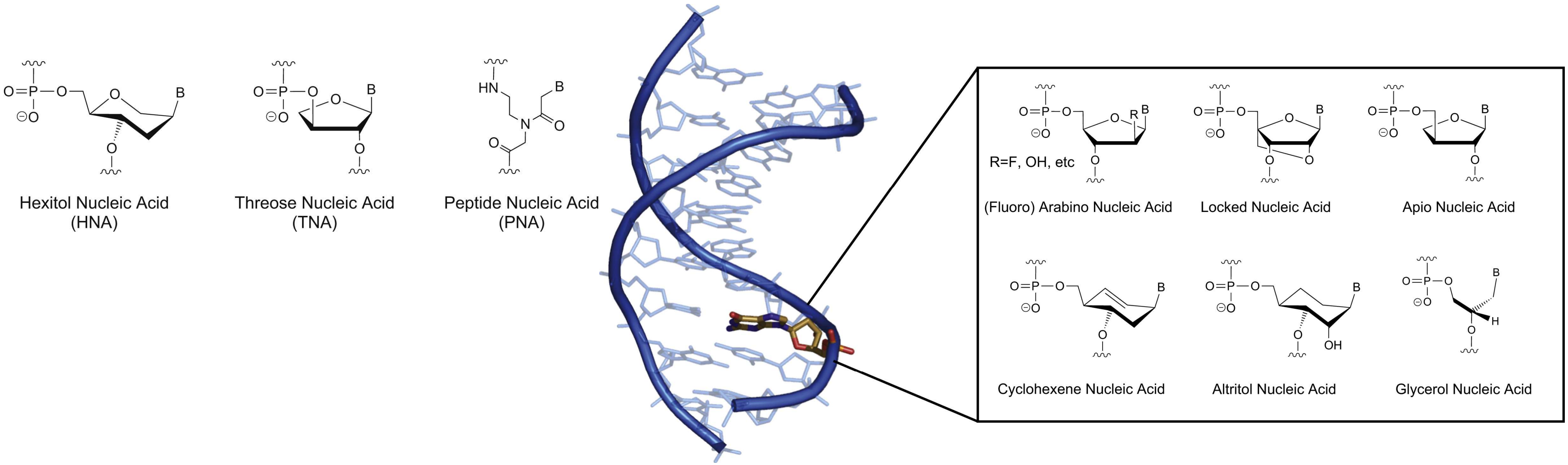
Peptide Nucleic Acid
(PNA)







Artificial genetic polymers



Hexitol Nucleic Acid
(HNA)

Threose Nucleic Acid
(TNA)

Peptide Nucleic Acid
(PNA)

R=F, OH, etc

(Fluoro) Arabino Nucleic Acid

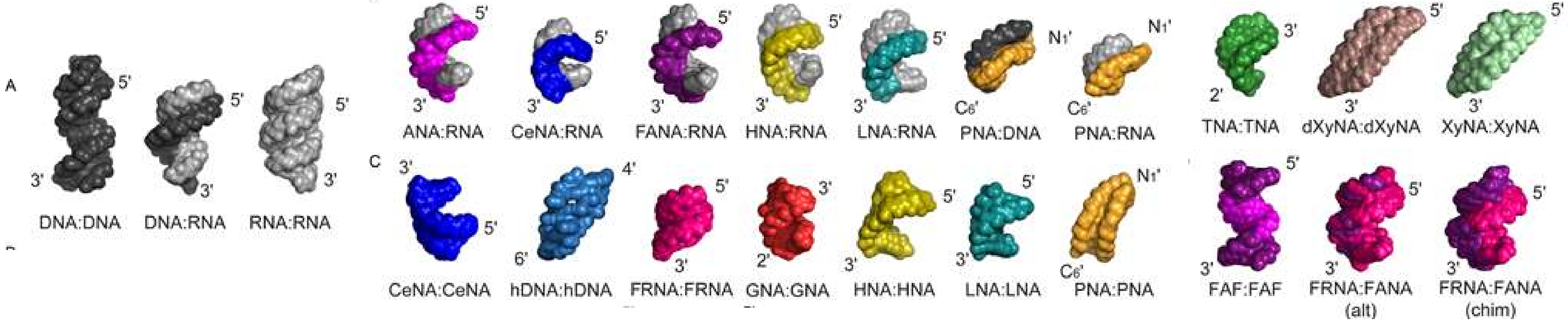
Locked Nucleic Acid

Apio Nucleic Acid

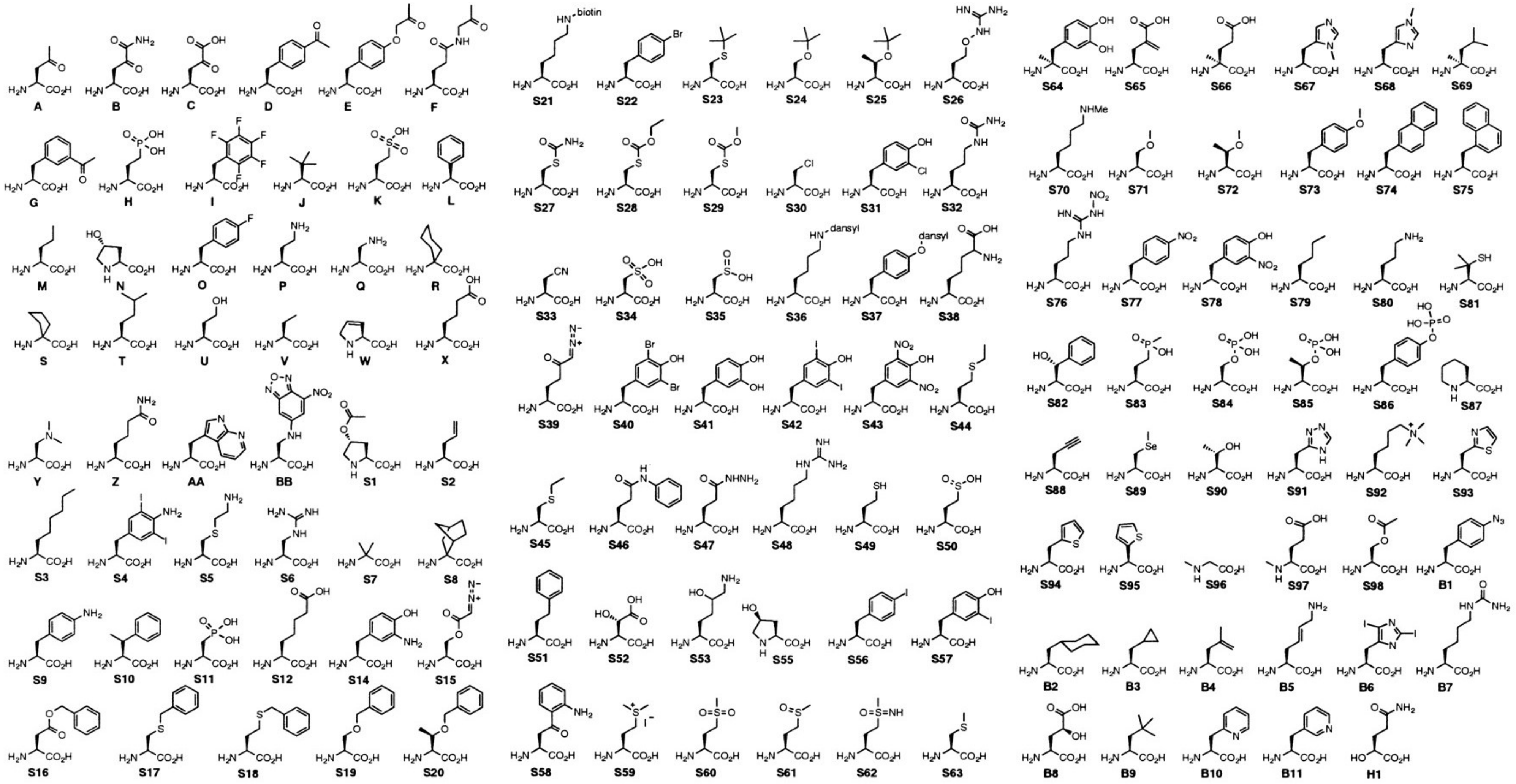
Cyclohexene Nucleic Acid

Altritol Nucleic Acid

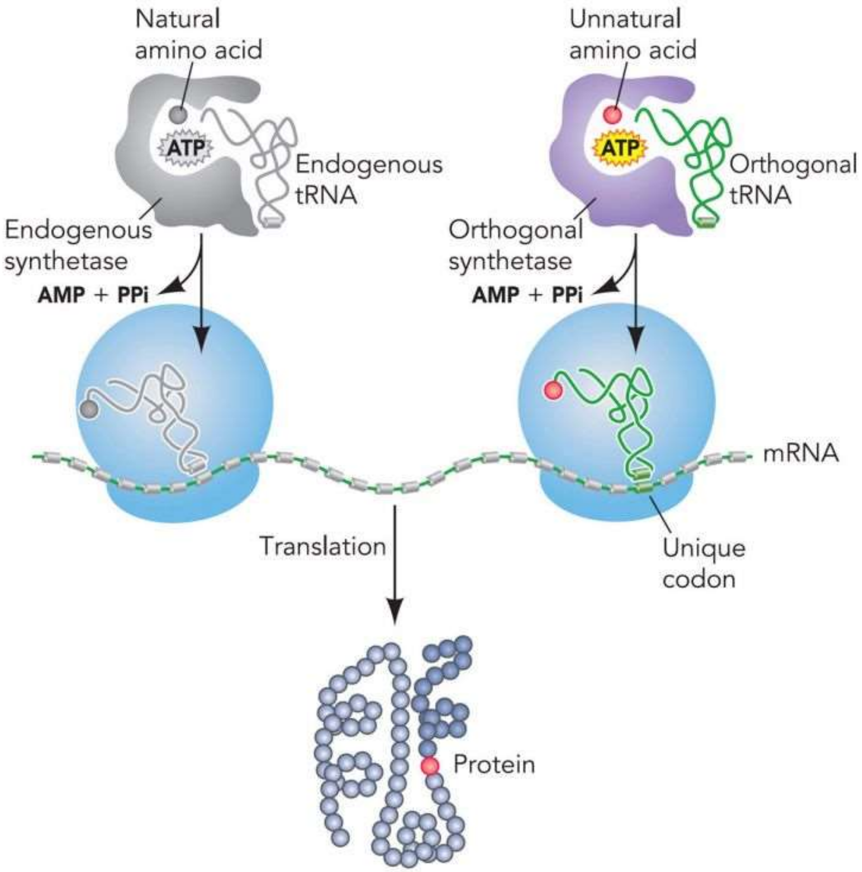
Glycerol Nucleic Acid

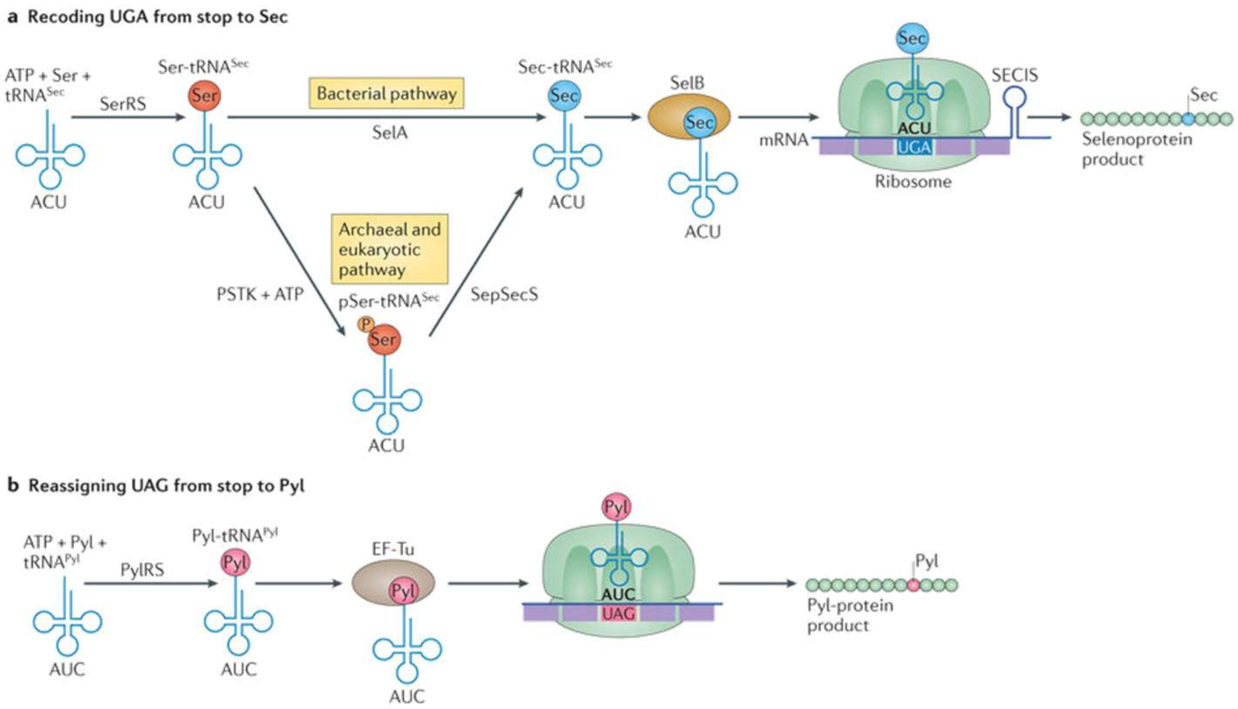


Expanded genetic code



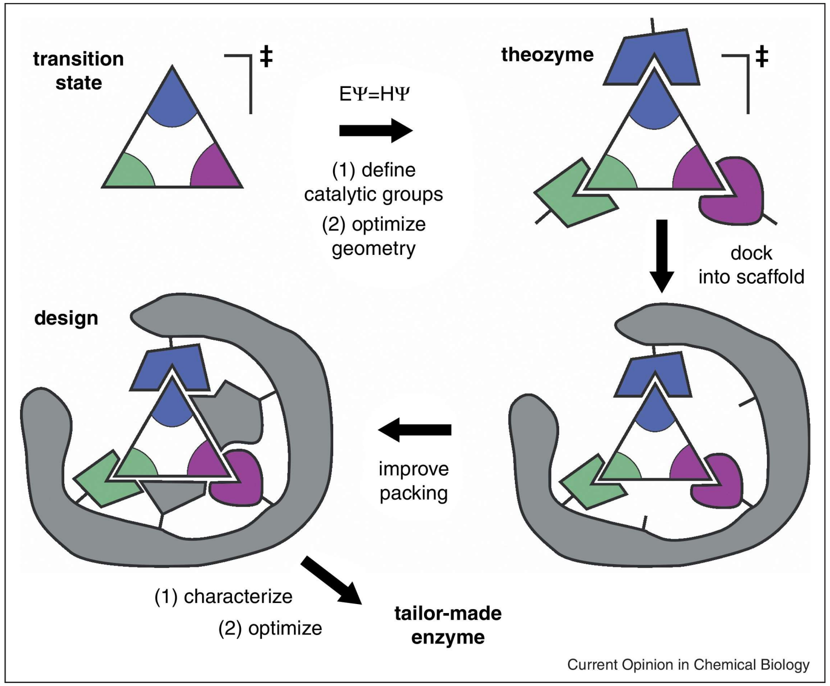
Expanded genetic code



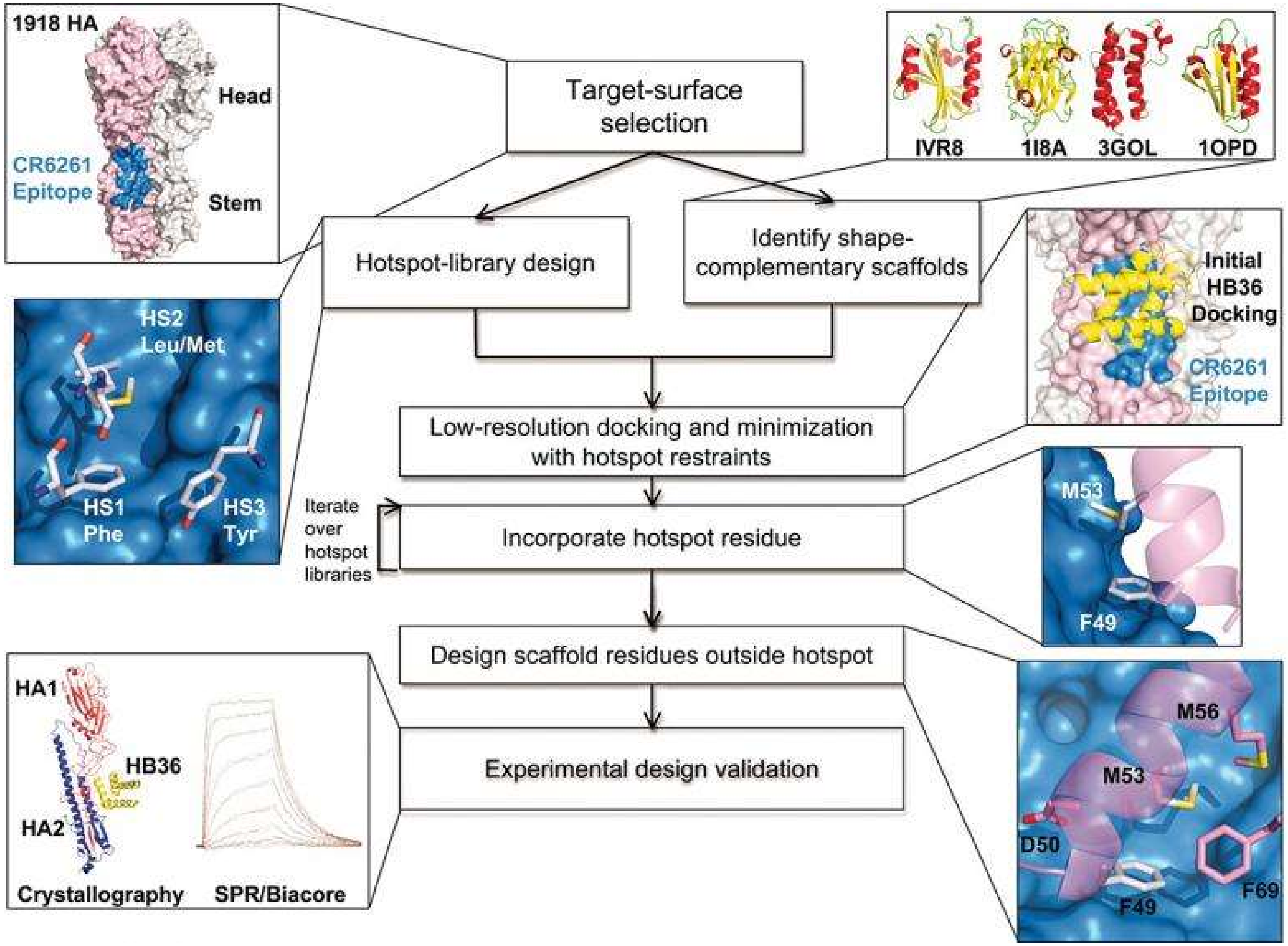


Nature Reviews | Microbiology

Protein engineering and de novo enzyme design



Current Opinion in Chemical Biology



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

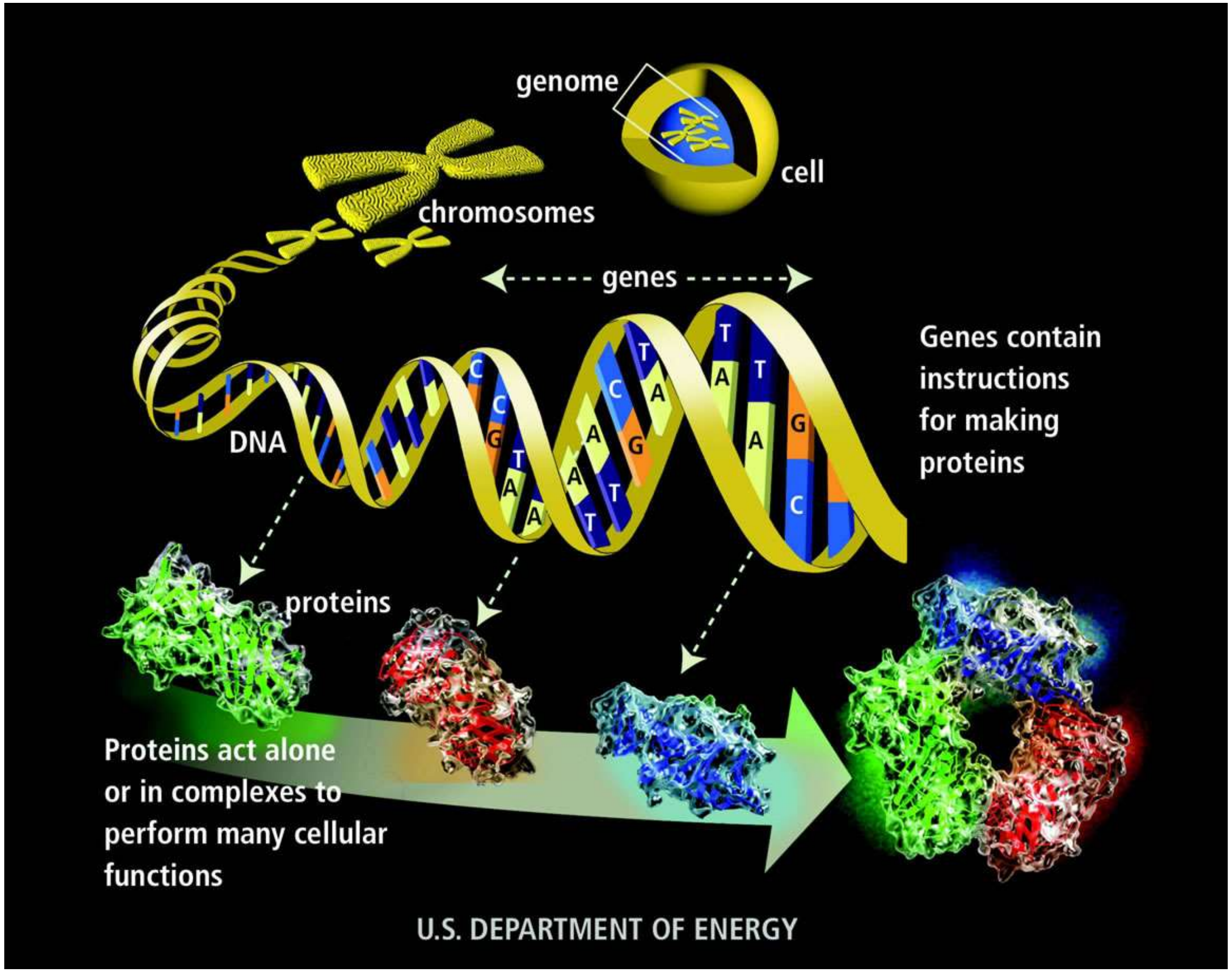
Polymerases and ribosomes - the molecular machines of life

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

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

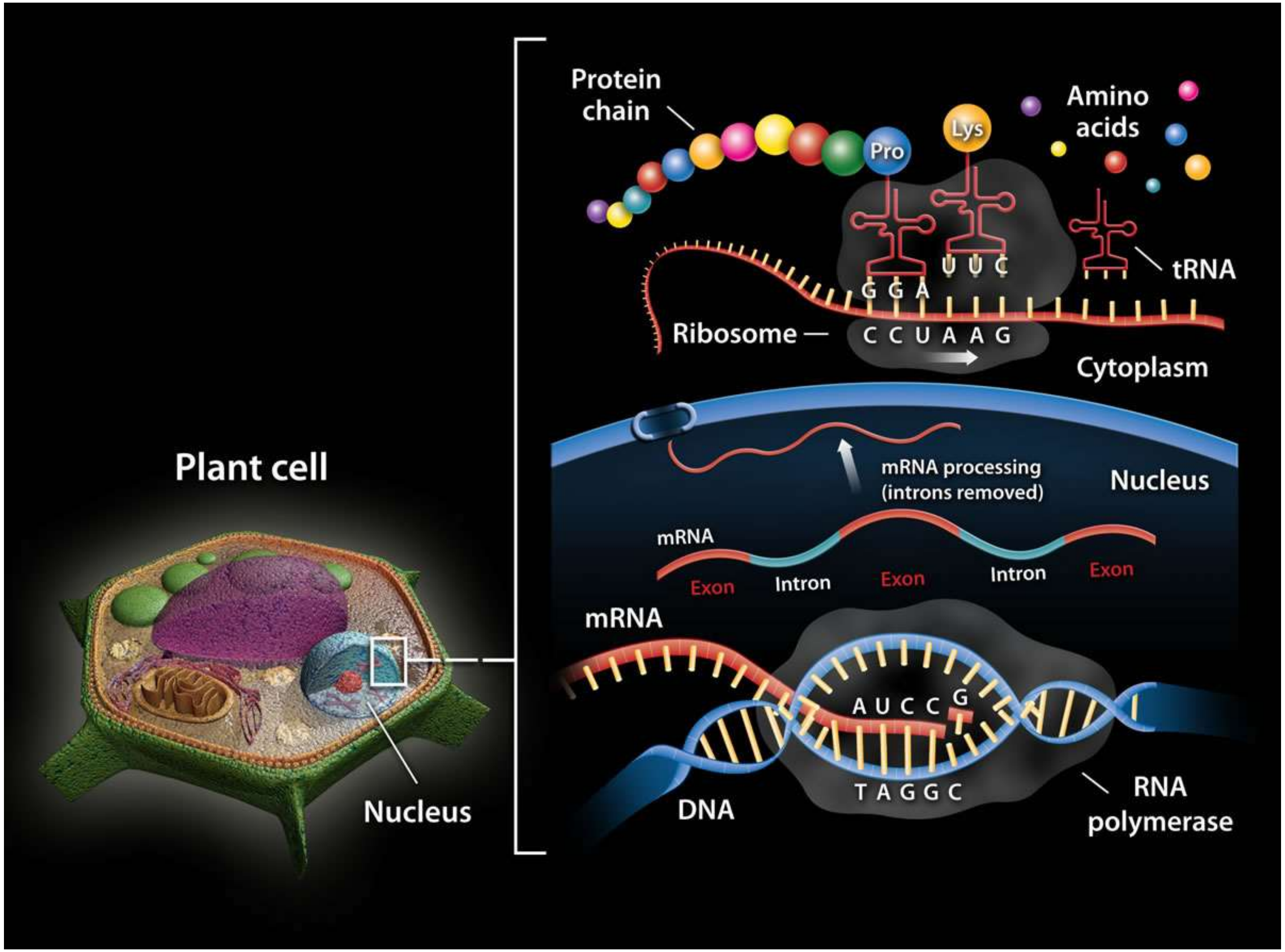
Protein engineering – how to make desired modifications in proteins

From DNA to proteins

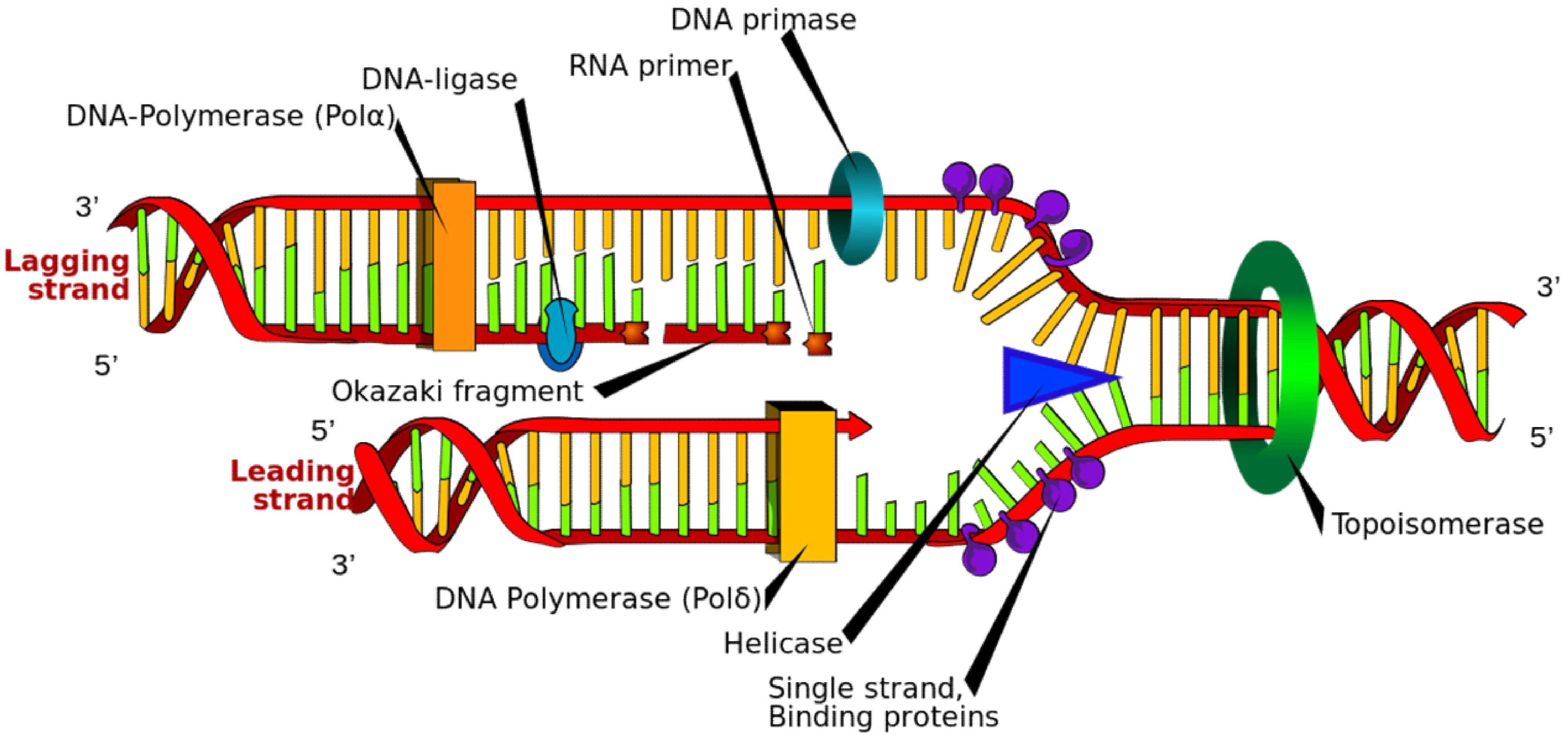


https://www.youtube.com/watch?v=gG7uCskUOrA

The Central Dogma: From DNA to proteins

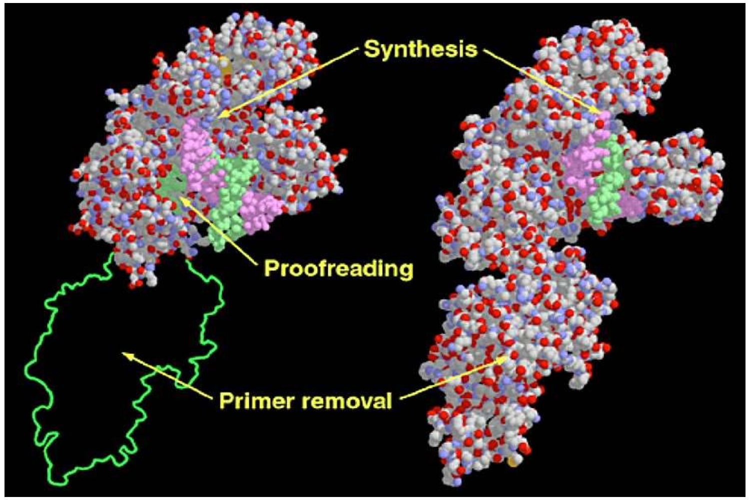


DNA replication

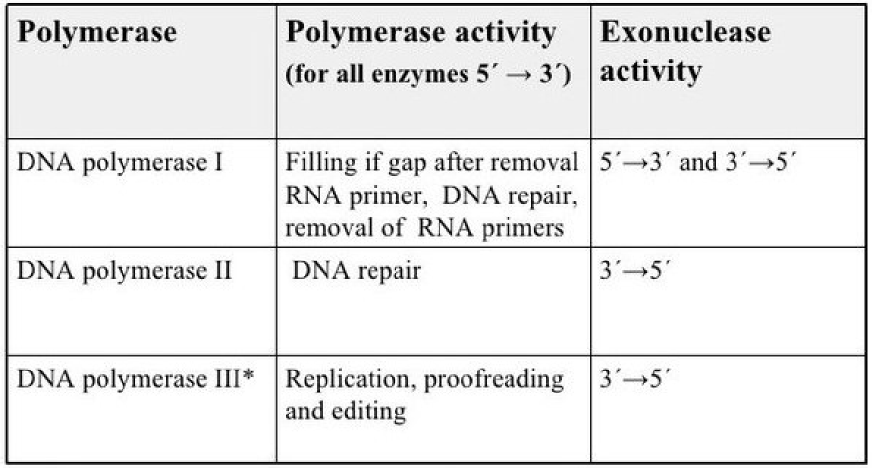


https://www.youtube.com/watch?v=TNKWgcFPHqw&ab_channel=yourgenome

DNA polymerases

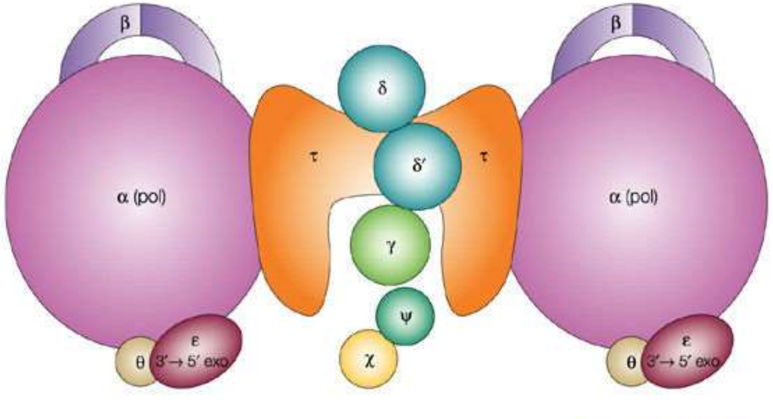


Procaryotic DNA Polymerases



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′

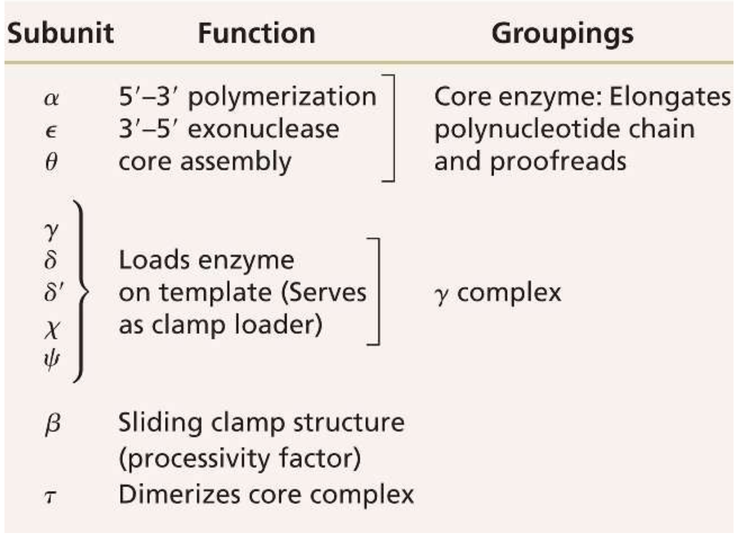
*The main enzyme of replication



Nature Reviews | Molecular Cell Biology

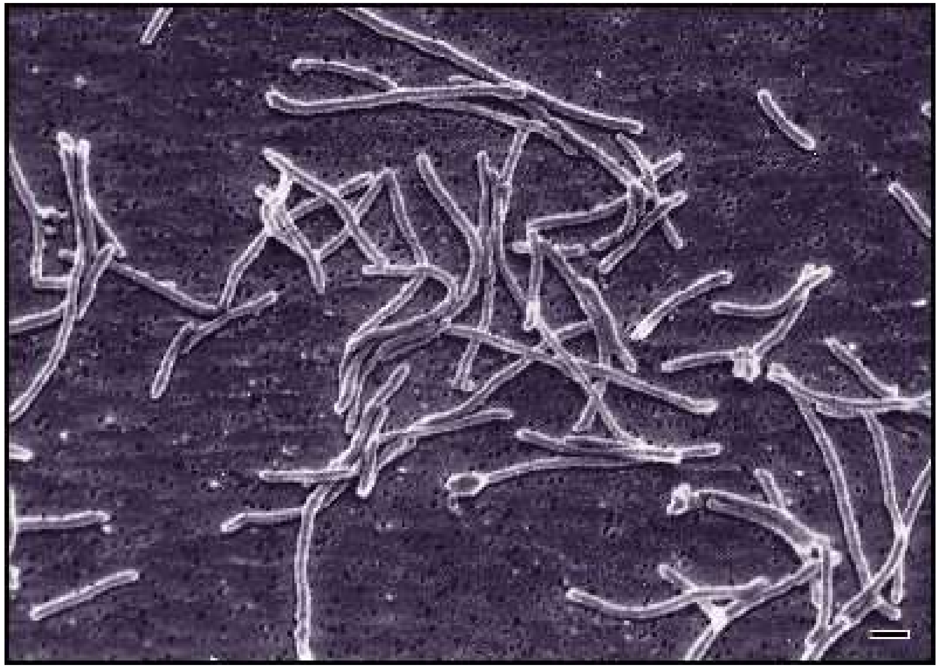
Procaryotic DNA Polymerase III

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 δ δ' χ ψ	Loads enzyme on template (Serves as clamp loader)	γ complex
β	Sliding clamp structure (processivity factor)	
τ	Dimerizes core complex	

Extremophilic organisms



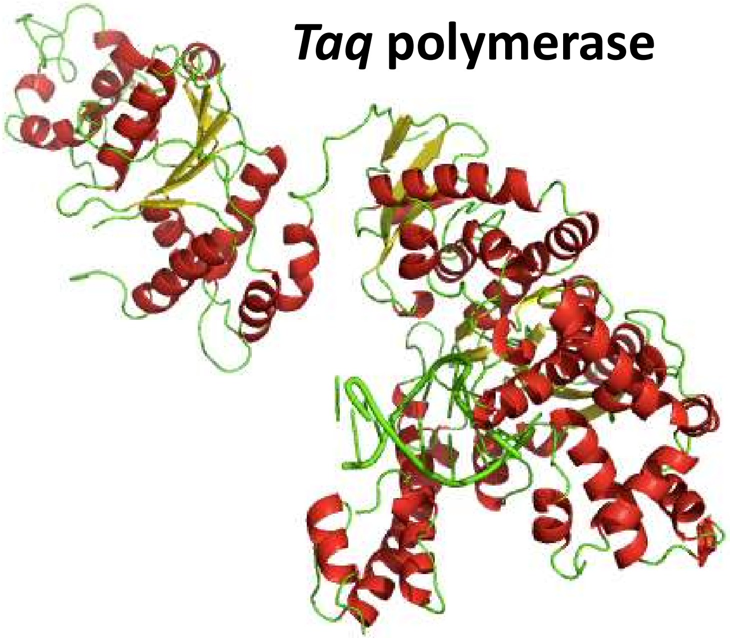
Brian W. Schaller, Yellowstone Park

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

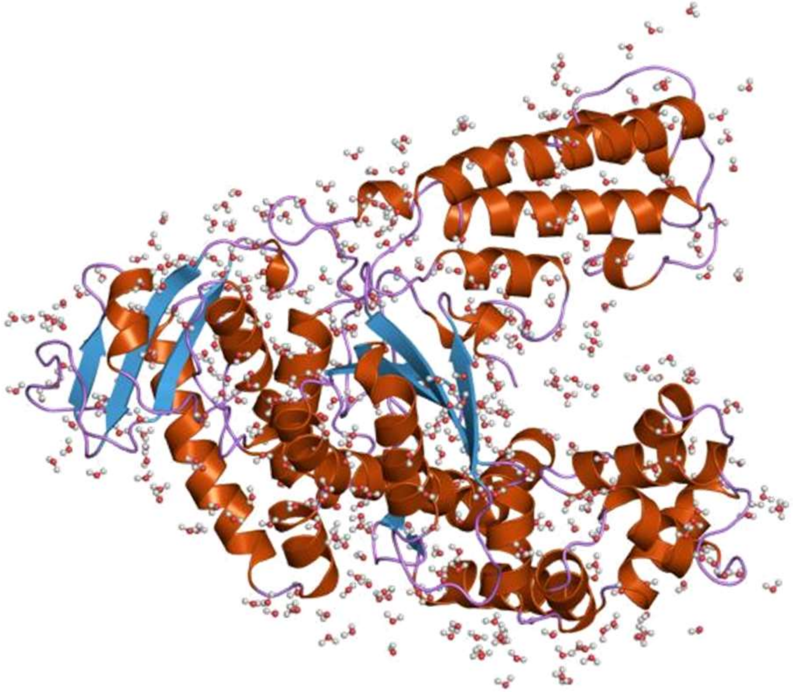
It is a source of thermostable enzymes

PCR – Polymerase Chain Reaction

***Taq* polymerase**



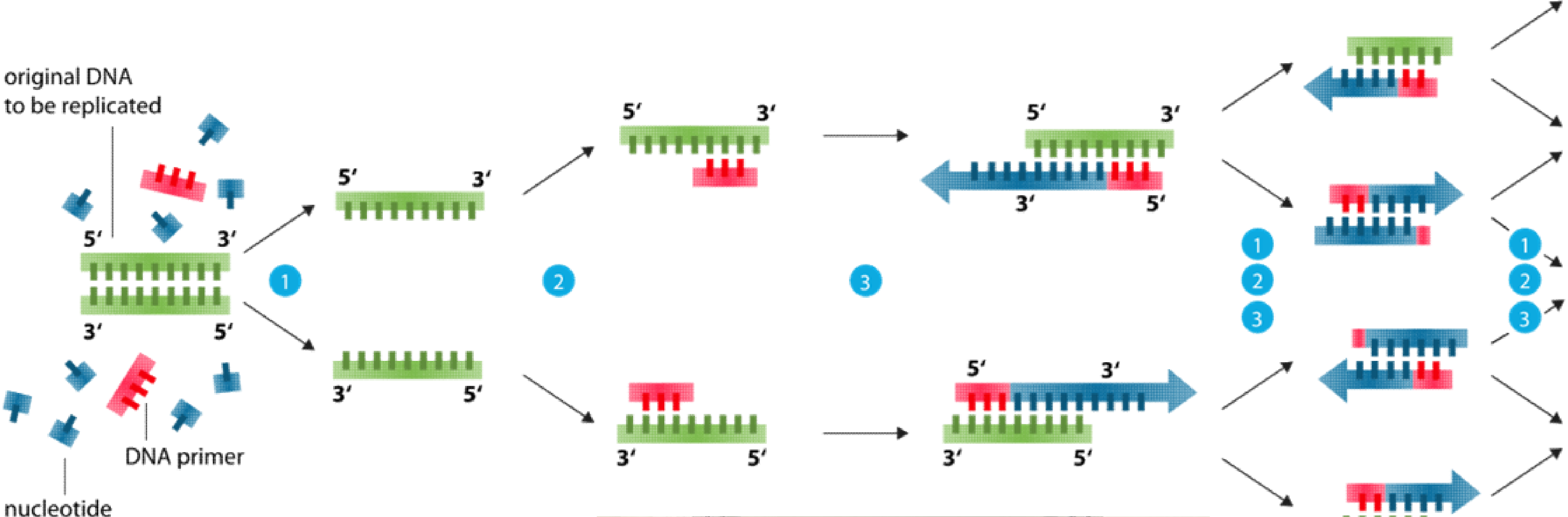
„Adenosine”



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

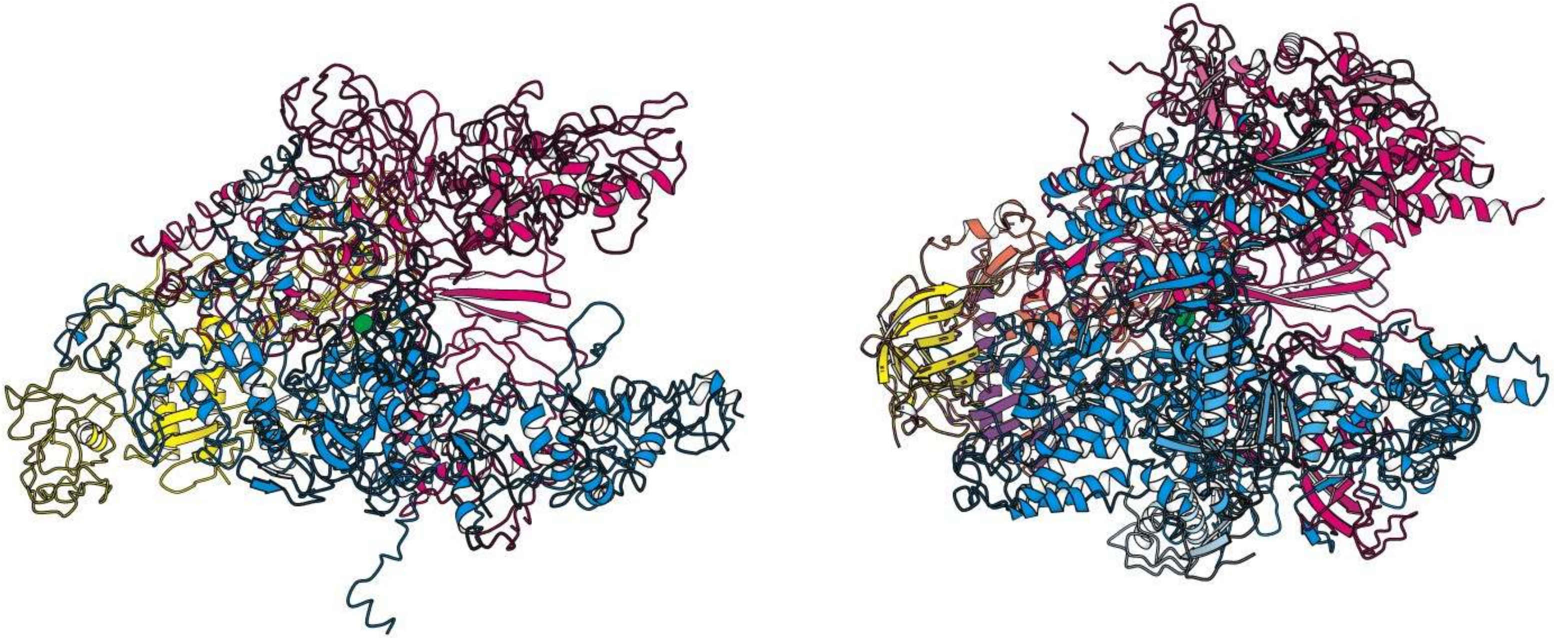
③ **Elongation** at ca. 72 °C

Rror

Transcription: DNA → RNA

RNA POLYMERASES IN EUKARYOTES

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

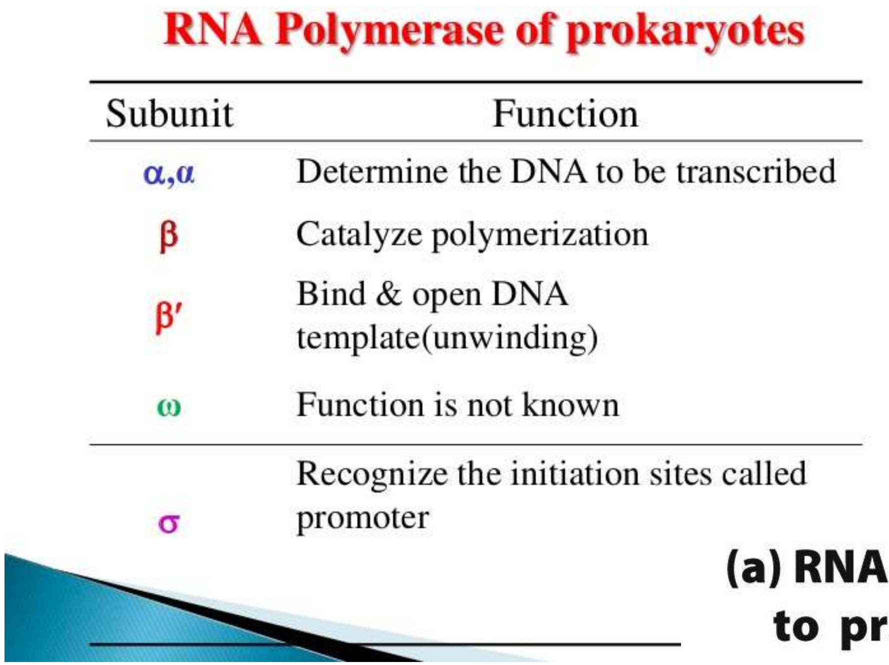


Prokaryotic RNA polymerase

Eukaryotic RNA polymerase

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

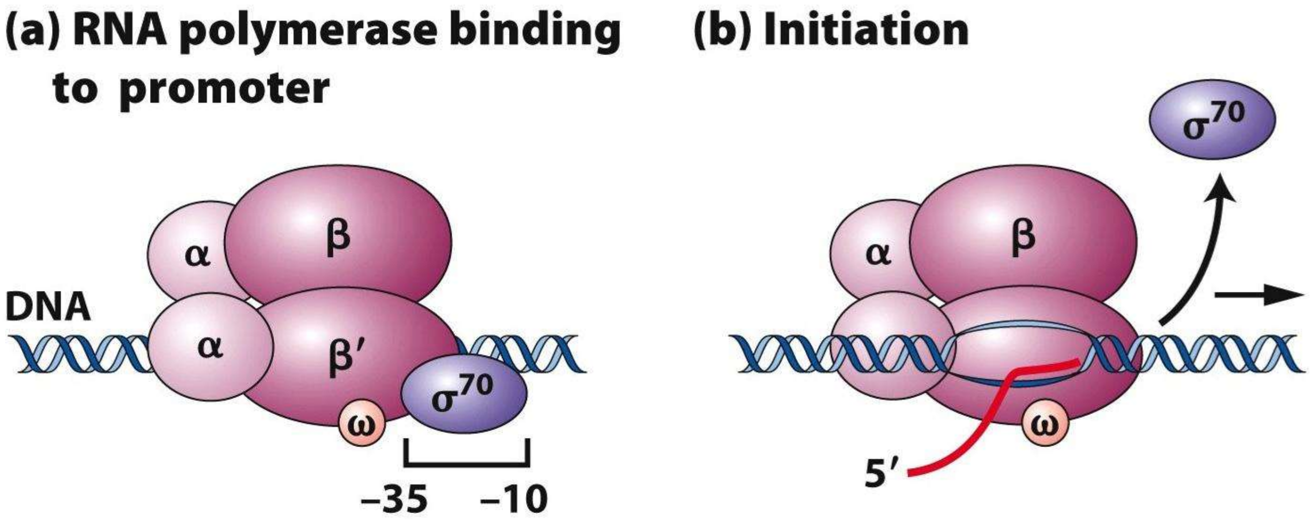
Transcription: DNA → RNA



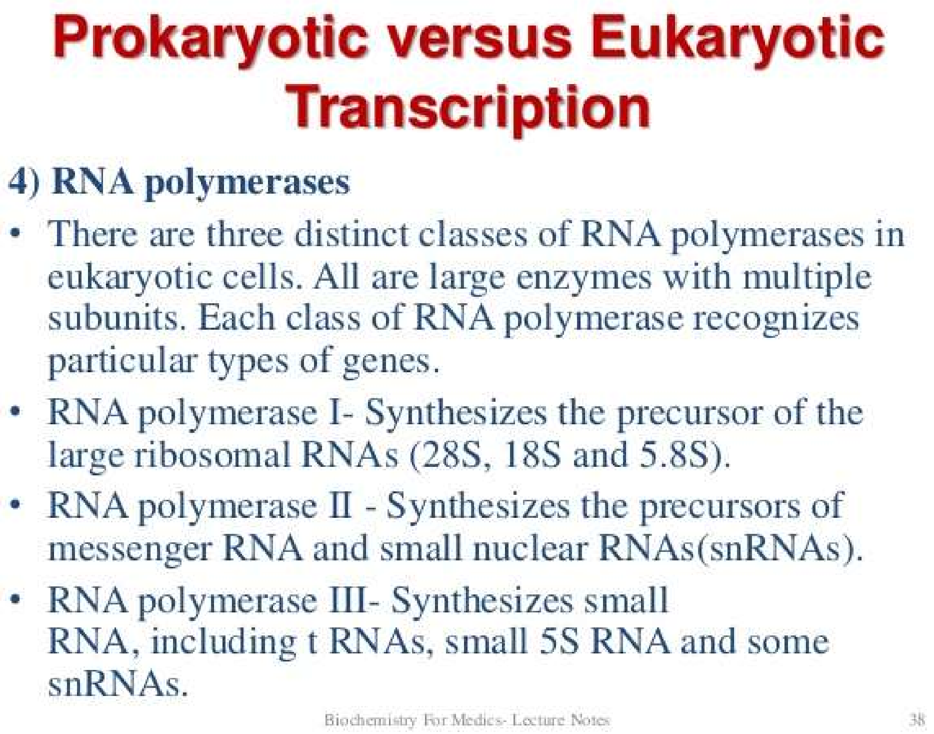
RNA Polymerase of prokaryotes

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



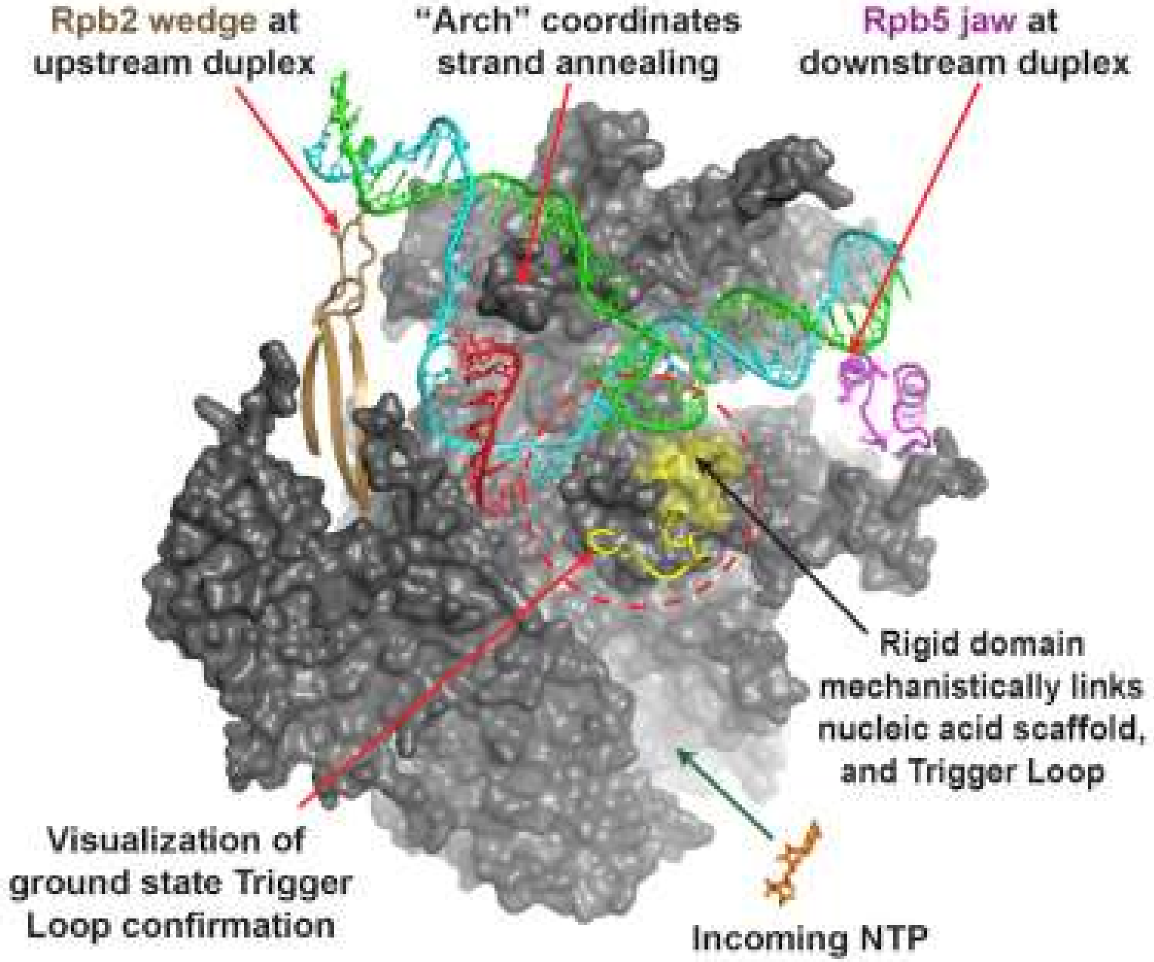
Transcription: DNA → RNA



Prokaryotic versus Eukaryotic Transcription

4) RNA polymerases

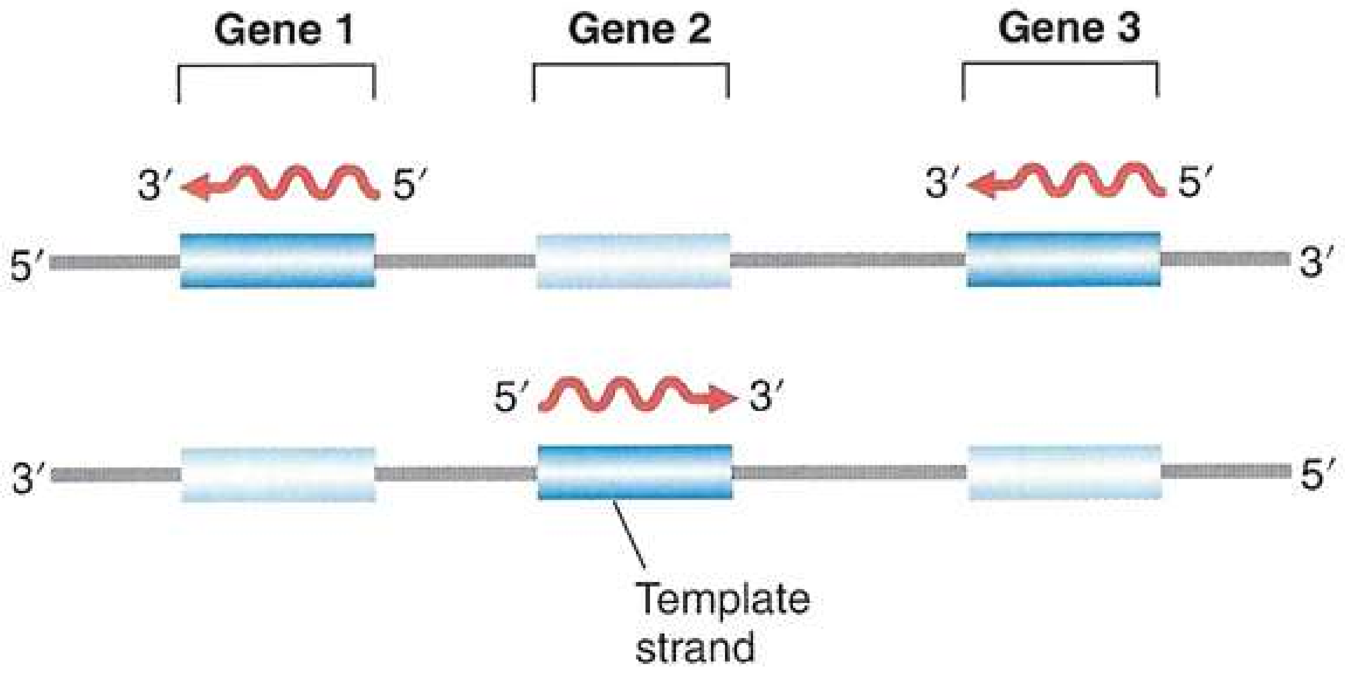
- There are three distinct classes of RNA polymerases in eukaryotic cells. All are large enzymes with multiple subunits. Each class of RNA polymerase recognizes particular types of genes.
- RNA polymerase I- Synthesizes the precursor of the large ribosomal RNAs (28S, 18S and 5.8S).
- RNA polymerase II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



Large Network of Interactions for DNA binding

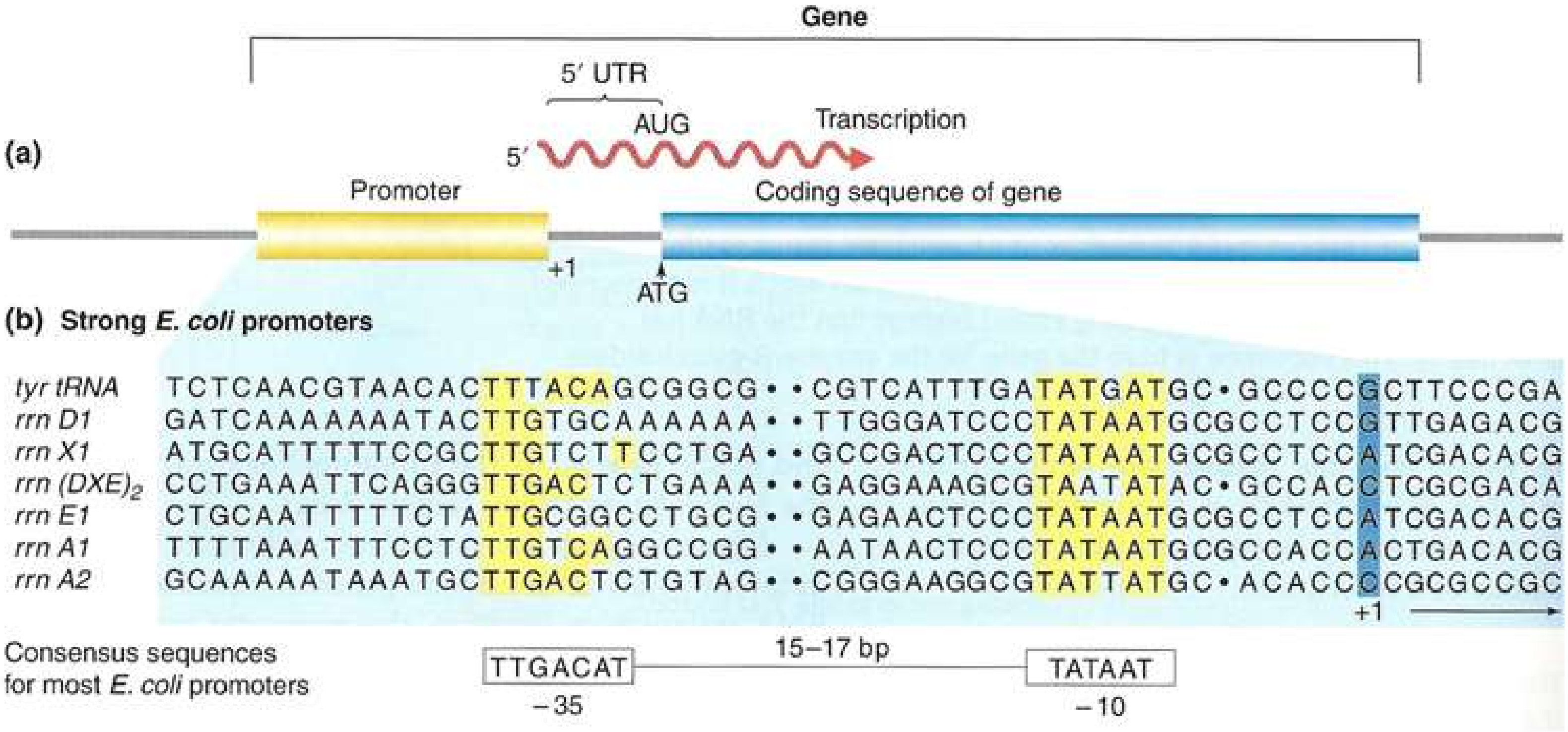
Transcription: DNA → RNA

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*):



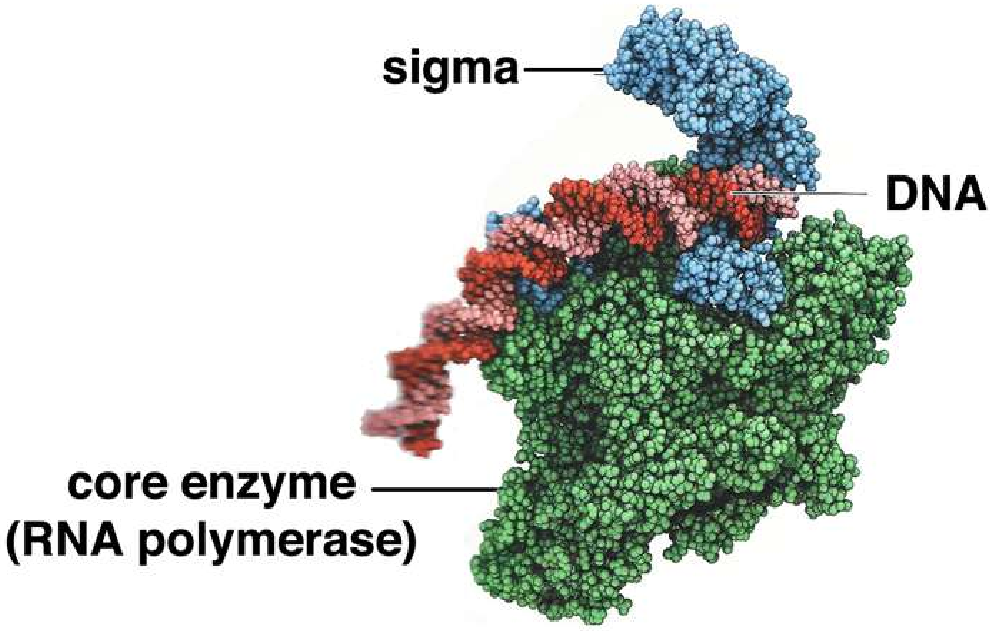
the transcription start site does not exactly correspond to the codon that is used to initiate translation – base 1 (+1). Transcription begins upstream of this sequence. The RNA between the „base 1” and the start of translation some distance downstream is called the 5' untranslated region or 5' UTR.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

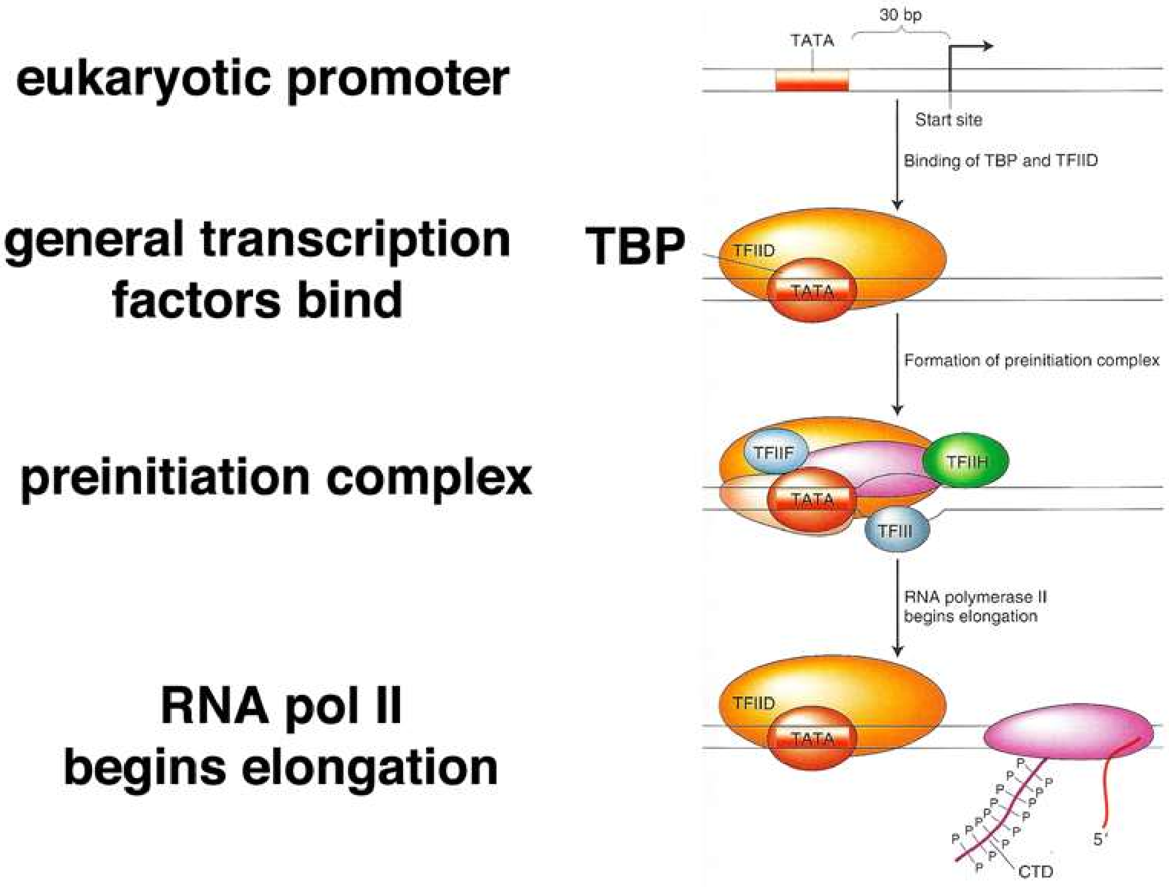
Transcription: DNA → RNA

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.

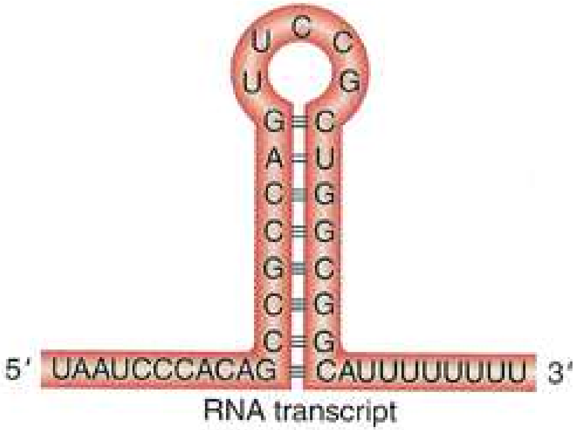
Once the preinitiation complex is formed, RNA polymerase is phosphorylated and released from the complex to begin RNA strand elongation.



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.

Bacterial transcription terminator



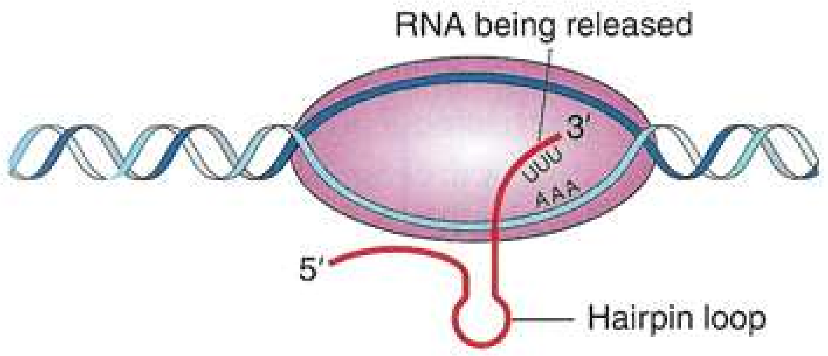
In the intrinsic mechanism, show above, there is a self-complementary sequence past the end of the coding sequence that forms a hairpin loop once it is transcribed. The base-paired part of the loop is very GC-rich, so the hairpin is stable. The presence of this structure interferes with RNA eleongation,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

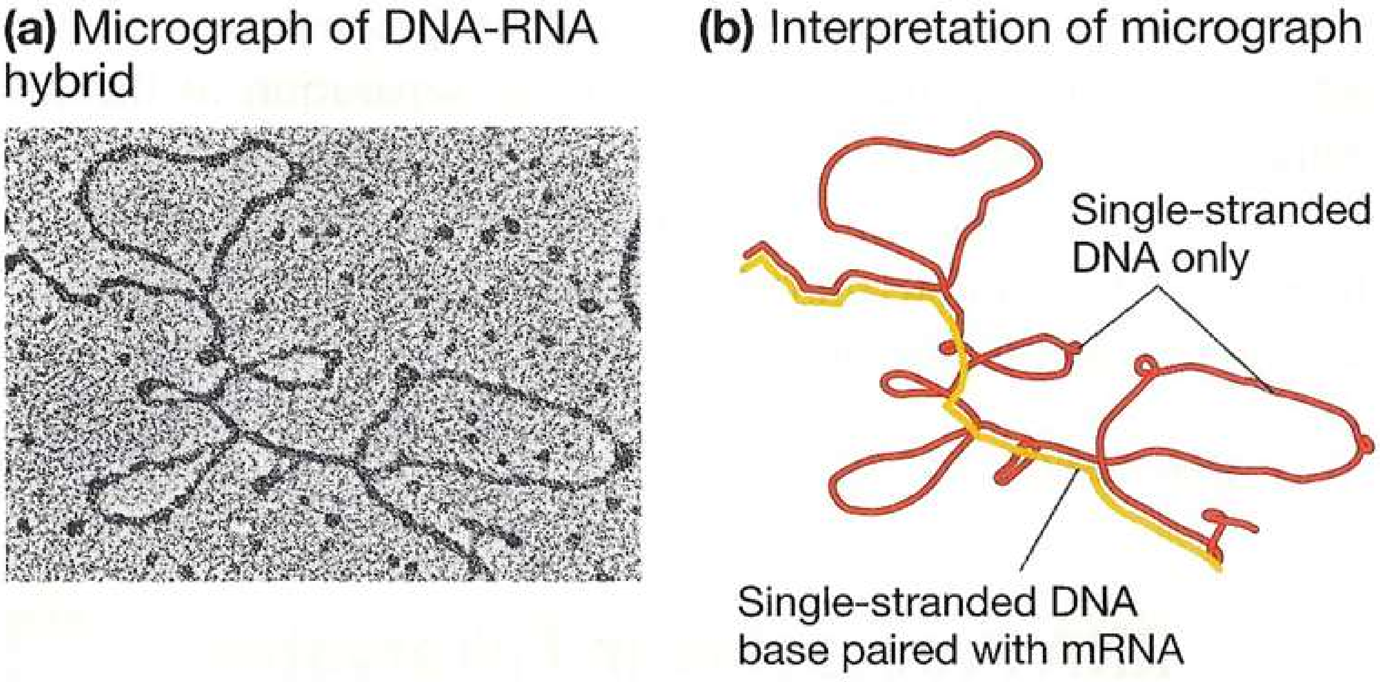
Eukaryotic transcription termination is similar to that of prokaryotes (polyT, or recruiting CPSF/CTSF).

Bacterial termination: intrinsic mechanism



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

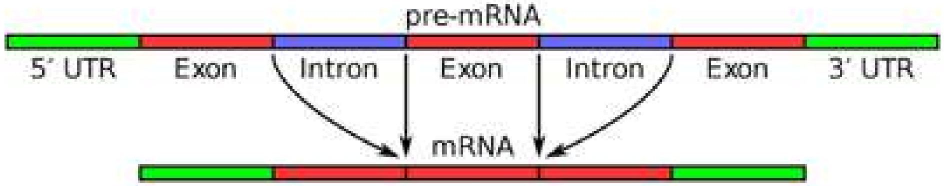
(b) Interpretation of micrograph

Eucaryotic genes are discontinuous!!!

mRNA processing

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

Eukaryotic mRNA processing



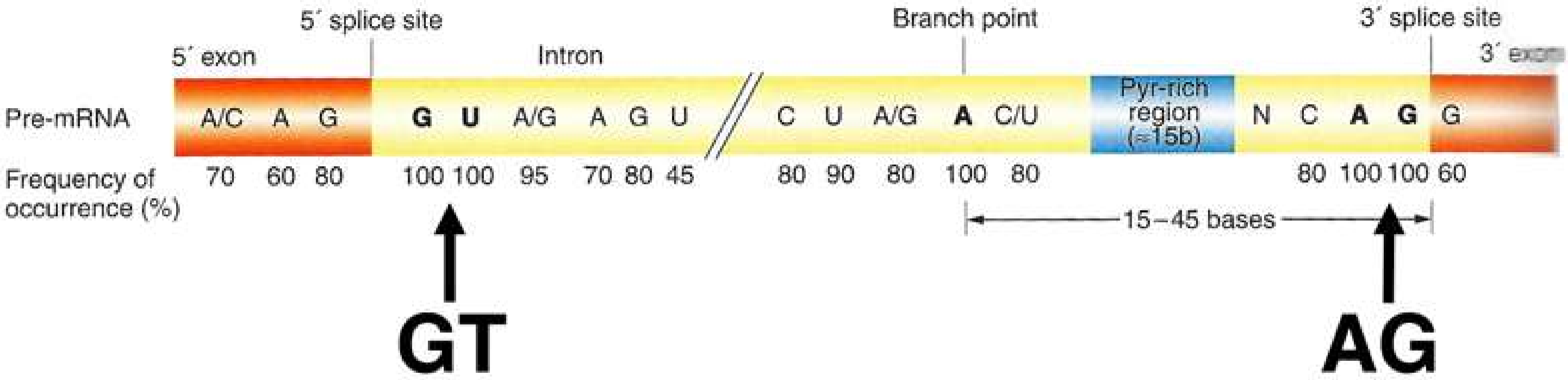
introns are removed (splicing)

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.

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

mRNA processing

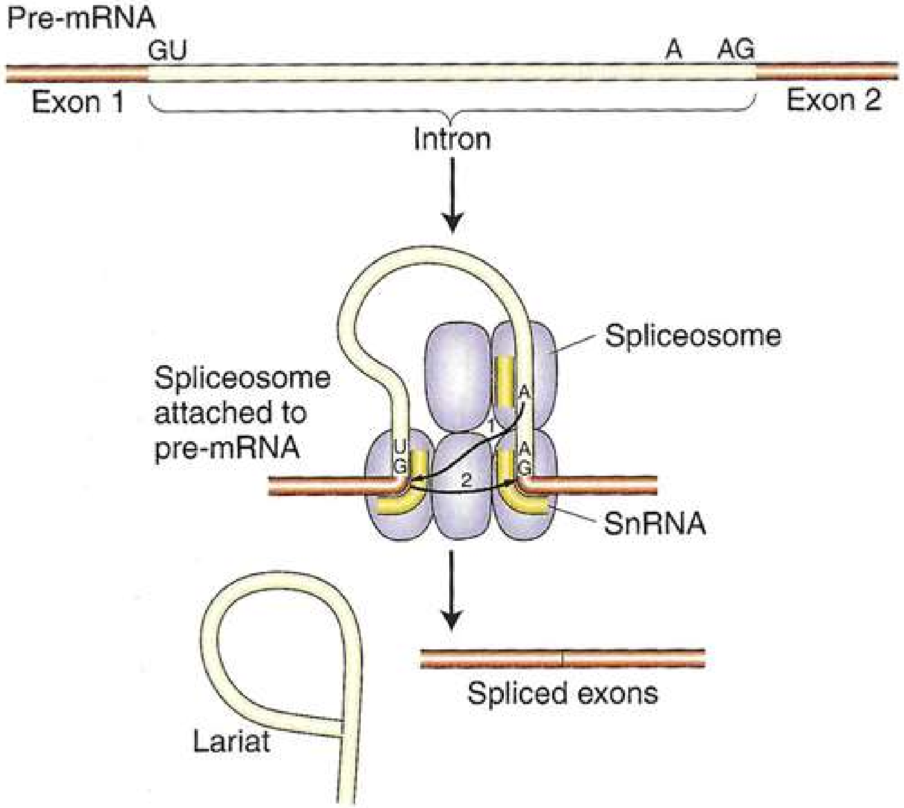
Sequencing of many eukaryotic genes reveals a consensus sequence for splice sites to remove introns



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.

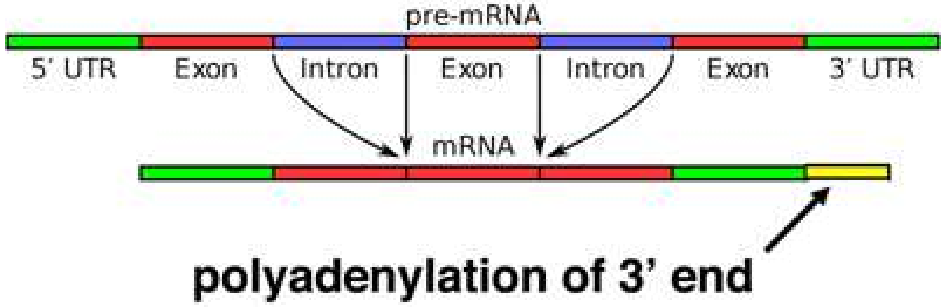
mRNA processing

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.

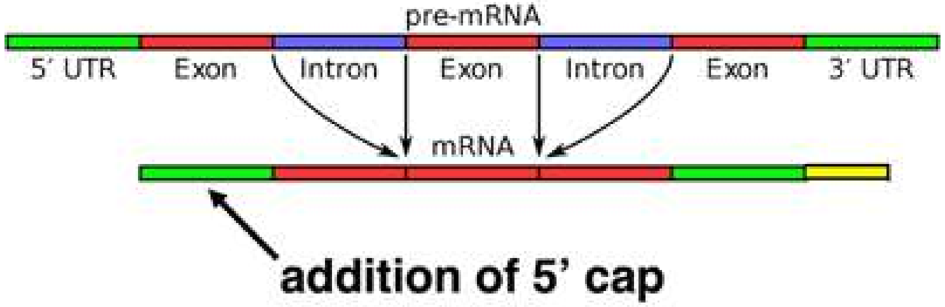


mRNA processing

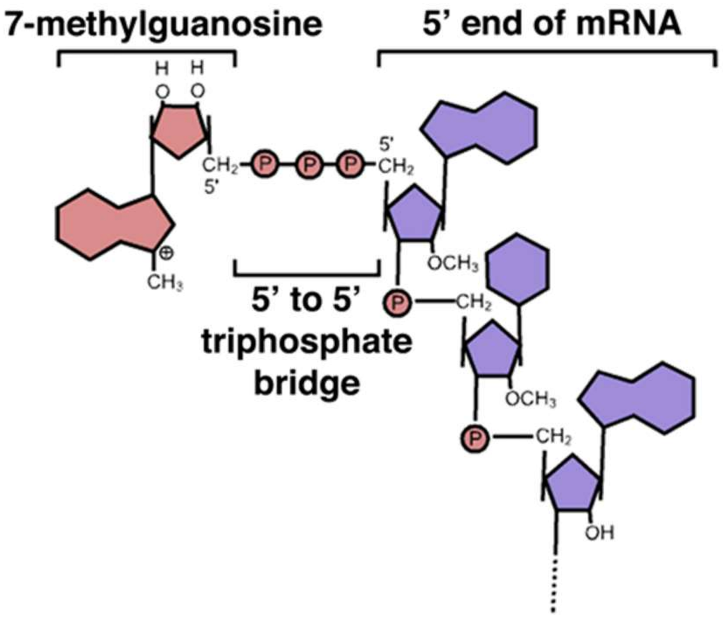
Eukaryotic mRNAs are also modified at the 3' end by the addition of a poly-A tail.



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

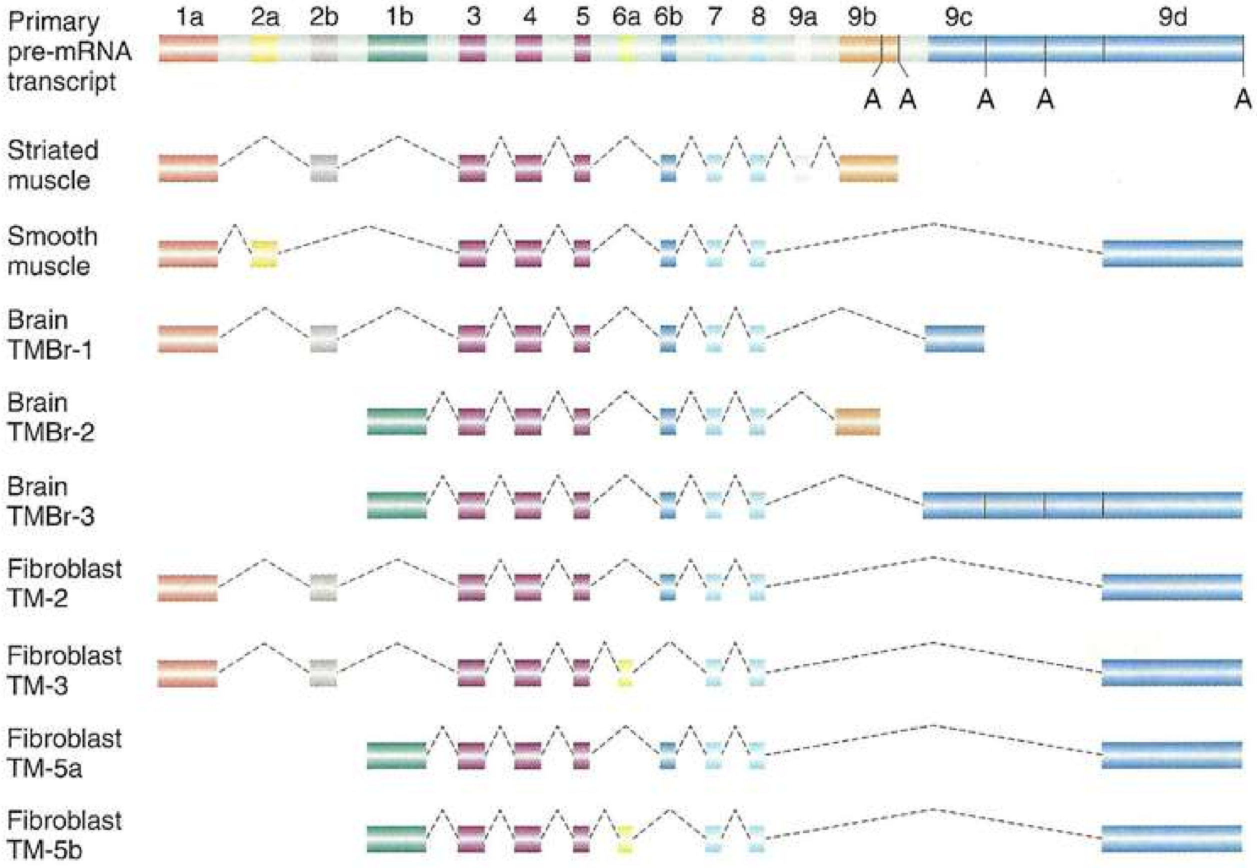


Eukaryotic 5' cap is a modified GTP

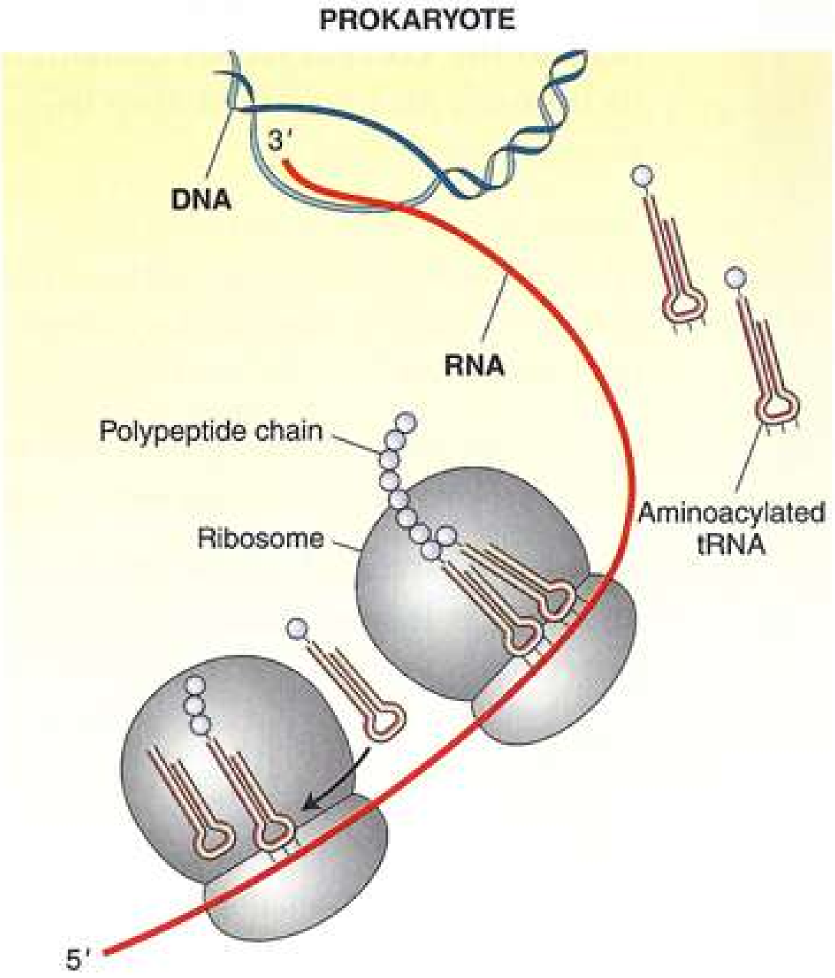


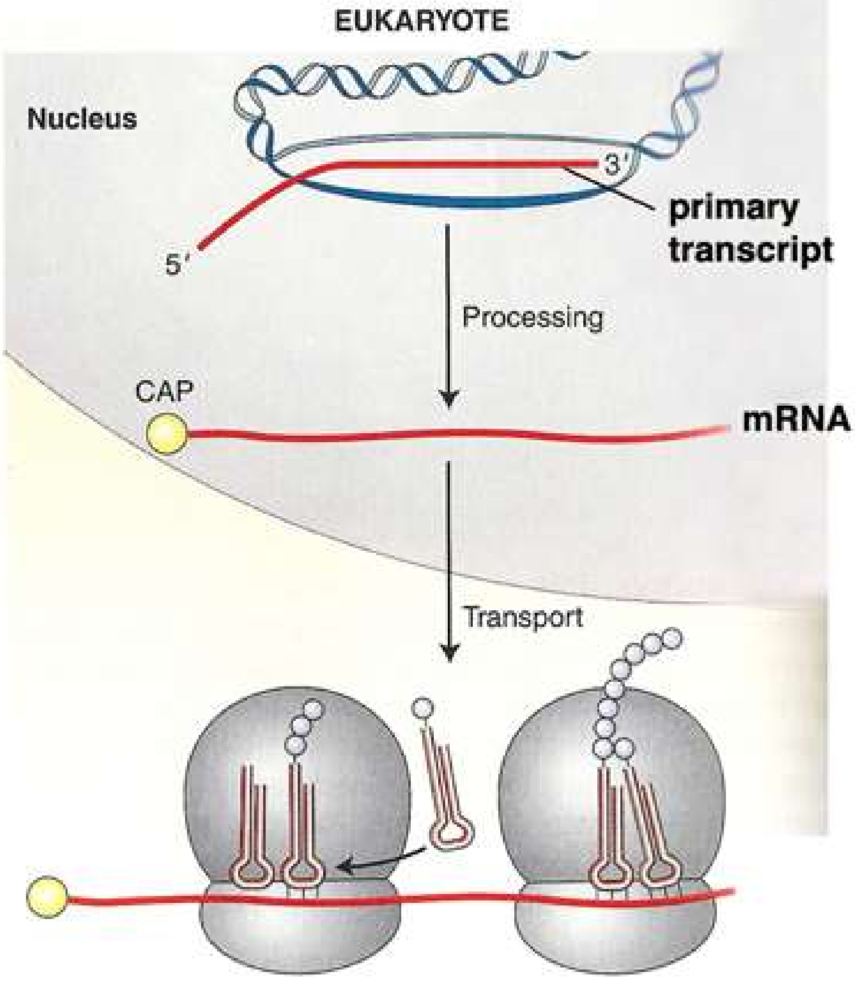
Alternative splicing - tropomyosin

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



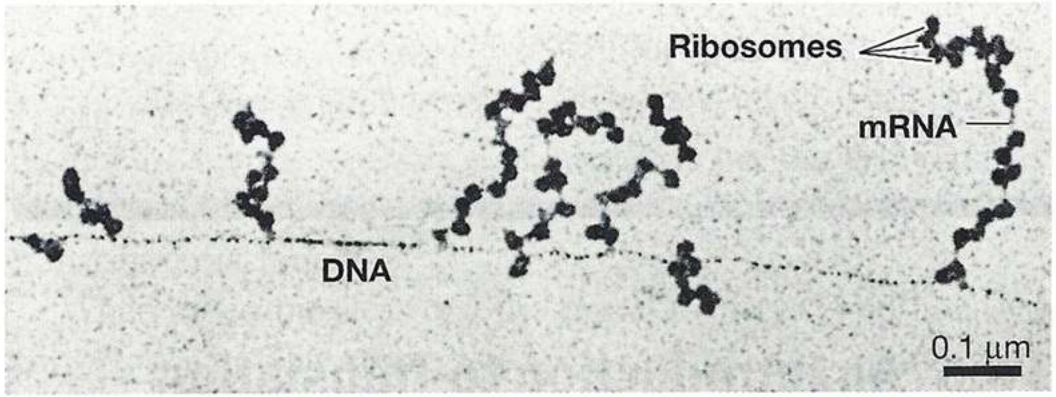


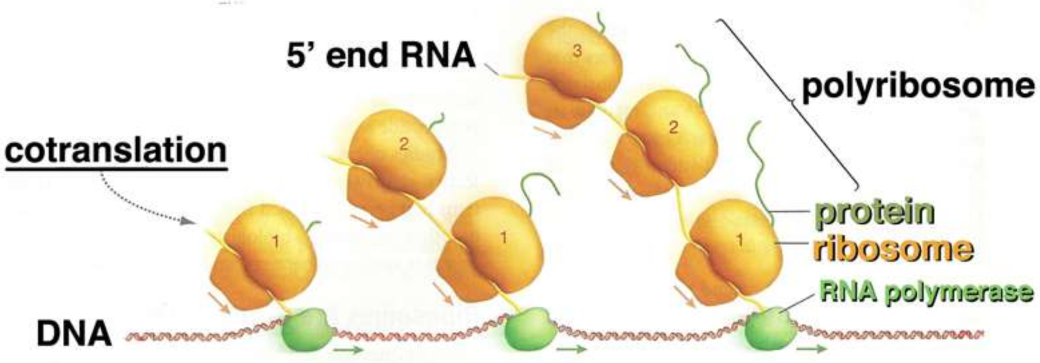
https://www.youtube.com/watch?v=TfYf_rPWUdY

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

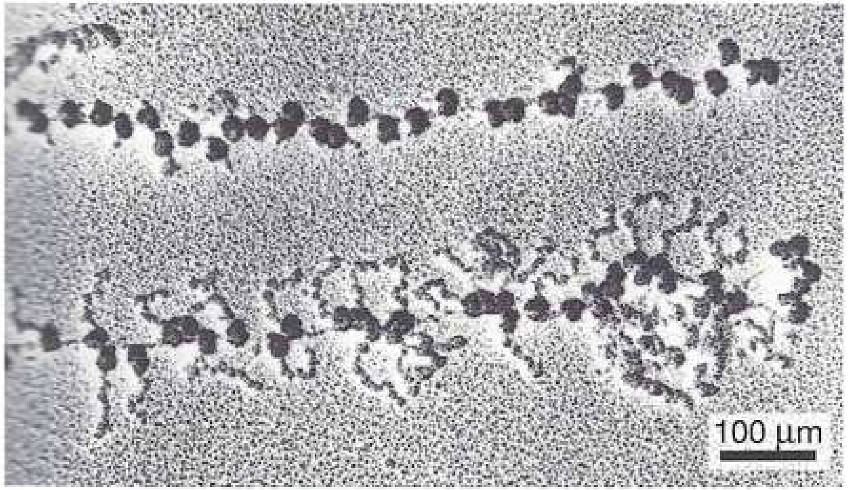
Translation: RNA → proteins

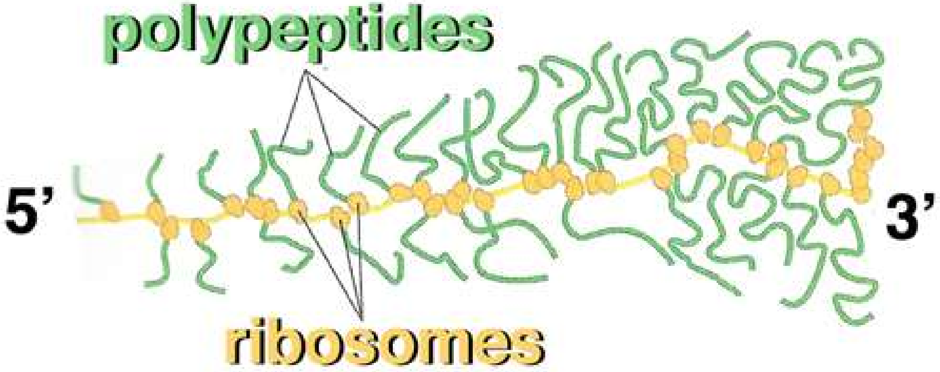
The coupling of transcription and translation in bacteria





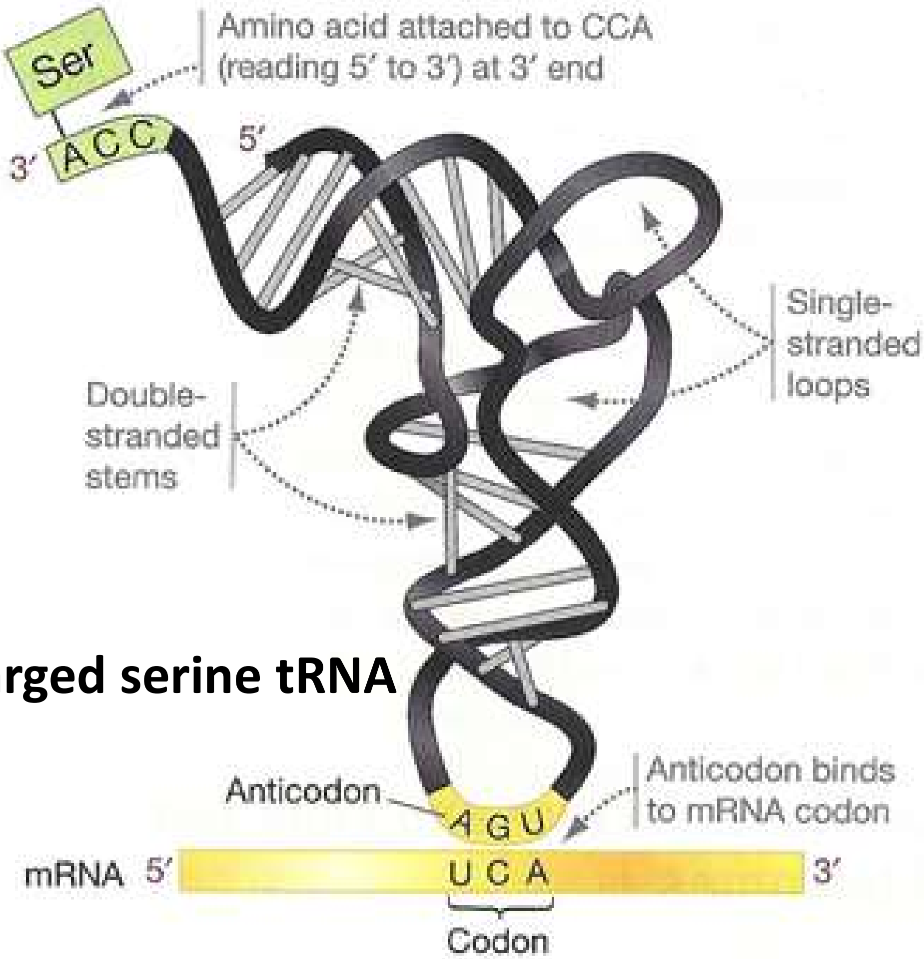
Eukaryotic polyribosomes





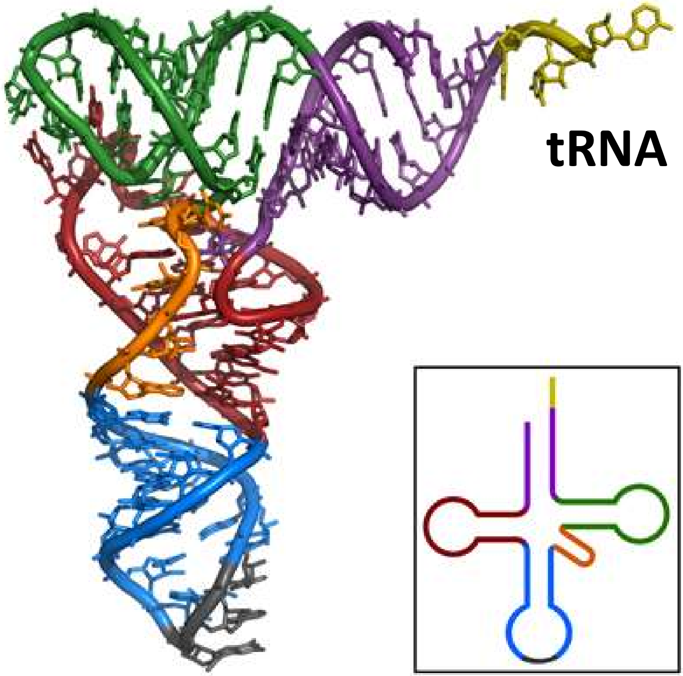
5' end of the mRNA, because there are shorter protein tails on the ribosomes at that end

Translation: RNA → proteins



Charged serine tRNA

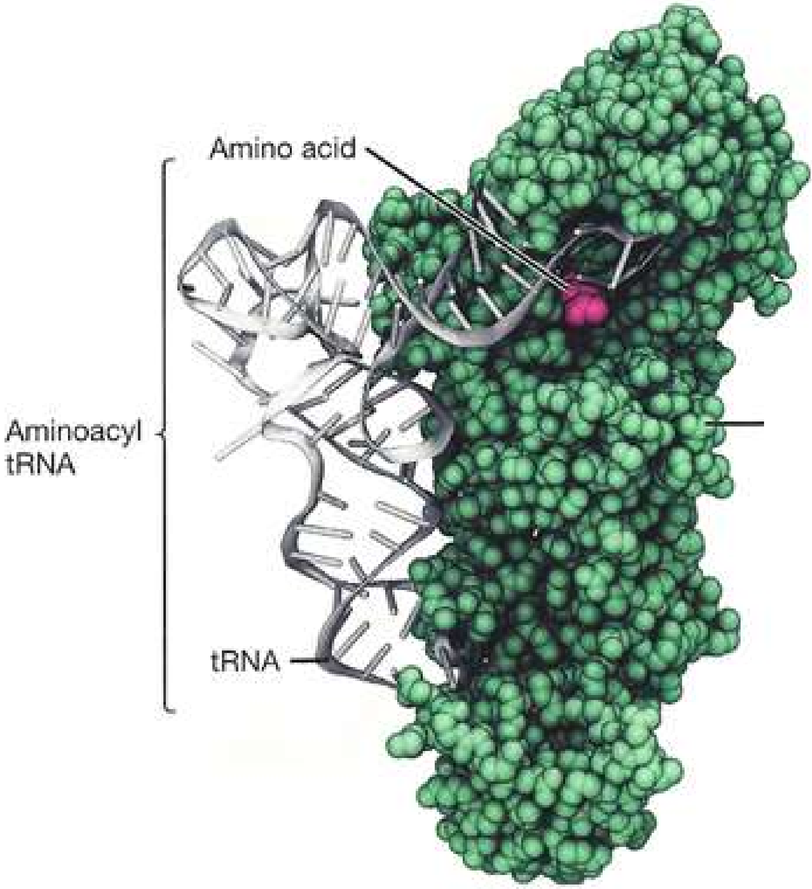
This shows a "charged" serine tRNA, covalently attached to the amino acid serine at its 3' end, with the anticodon paired to a serine codon



tRNA in 3D

Translation: RNA → proteins

Aminoacyl tRNA synthethase



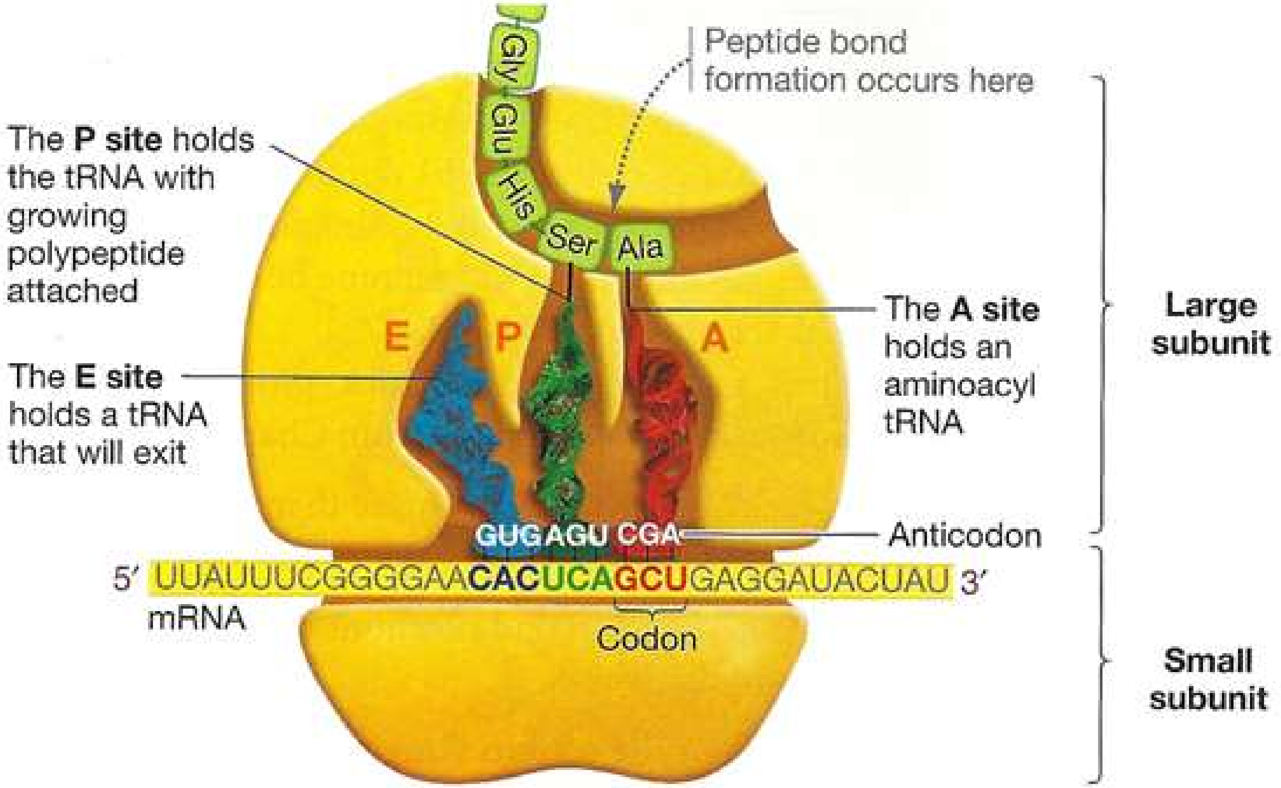
A special set of enzymes "charges" tRNAs, attaching the correct amino acid to particular tRNAs.

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

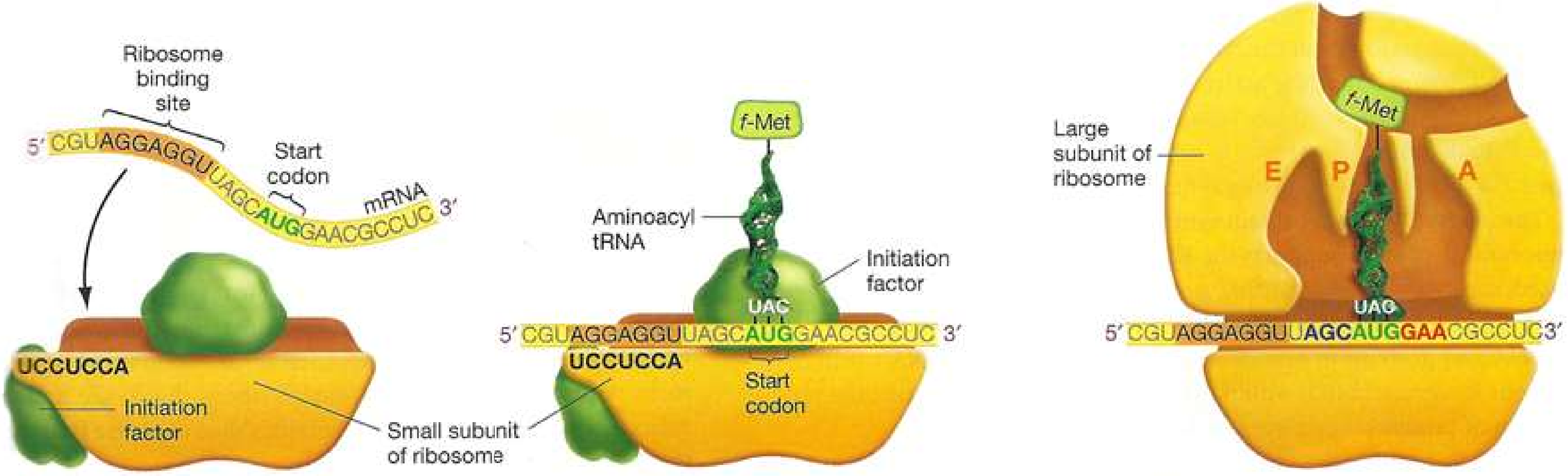
Translation: RNA → proteins



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



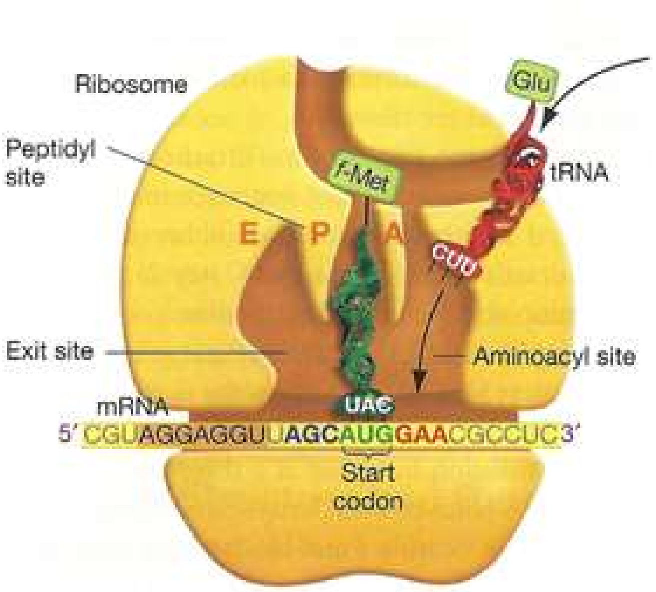
1. mRNA binds to small subunit. Ribosome binding site sequence binds to a complementary sequence in an RNA molecule in the small subunit of the ribosome, with the help of protein initiation factors.

2. Initiator aminoacyl tRNA binds to start codon.

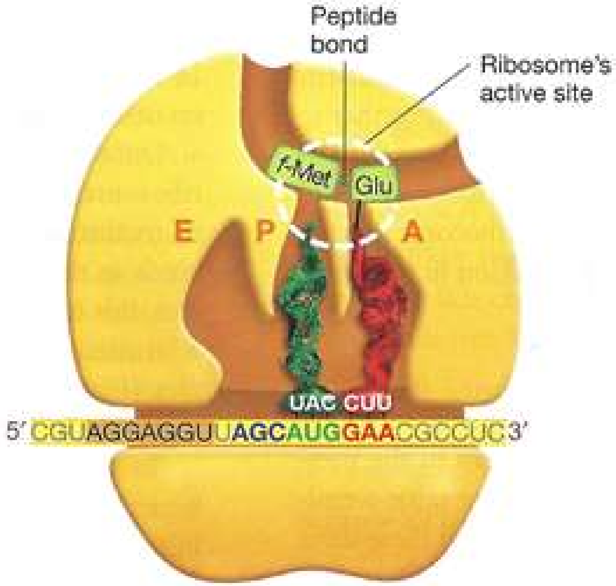
3. Large subunit of ribosome binds, completing ribosome assembly. Translation begins.

Translation: RNA → proteins

Translation elongation



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

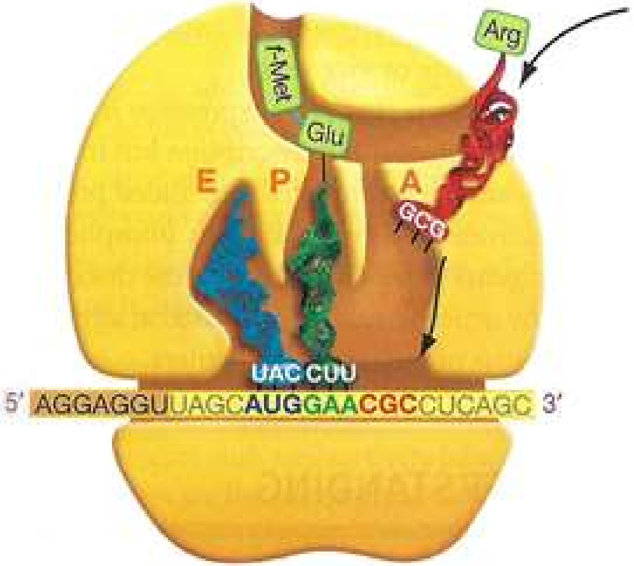


2. Peptide bond formation
The amino acid attached to the tRNA in the P site is transferred to the tRNA in the A site.

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.

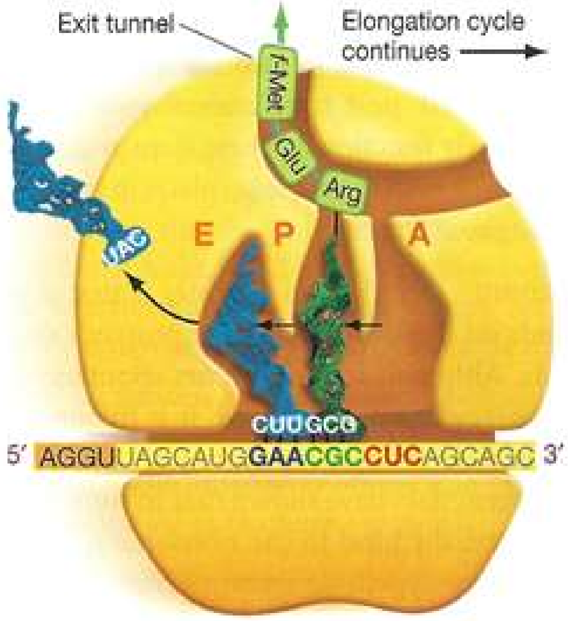
Translation: RNA → proteins

Translation elongation



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

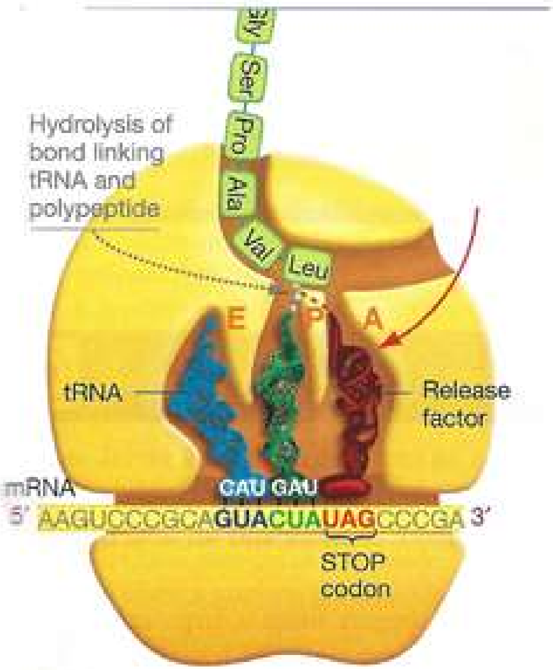
5. Peptide bond formation
The polypeptide chain attached to the tRNA in the P site is transferred to the aminoacyl tRNA in the A site.



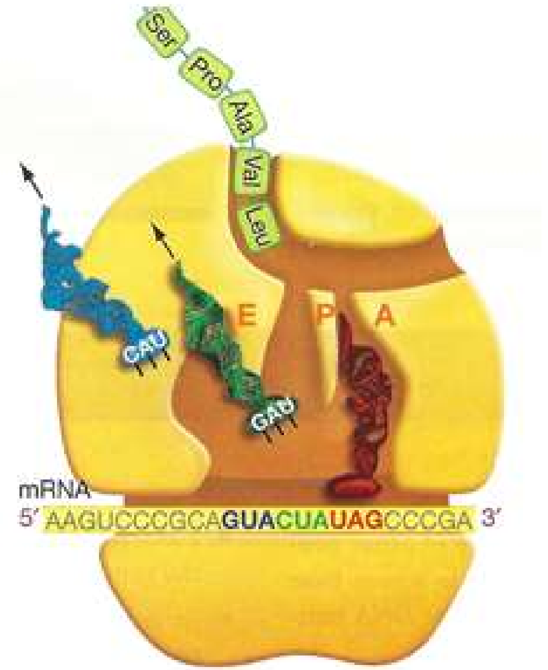
6. Translocation
mRNA is ratcheted through the ribosome again. The tRNA attached to polypeptide chain moves into P site. Empty tRNA from P site moves to E site, where tRNA is ejected. The A site is empty again.

Translation: RNA → proteins

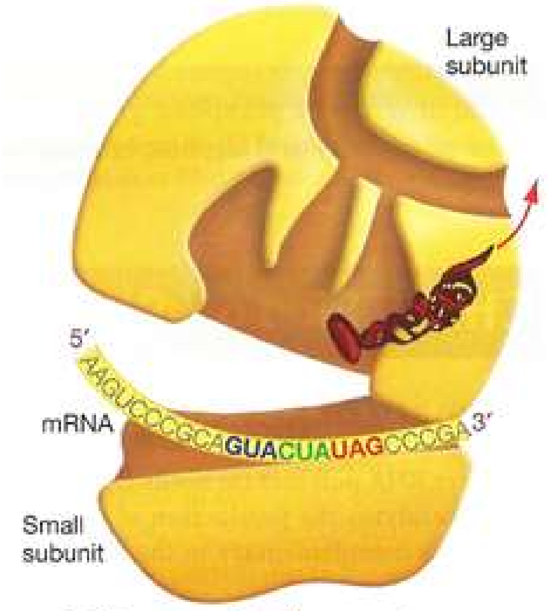
Translation termination



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...

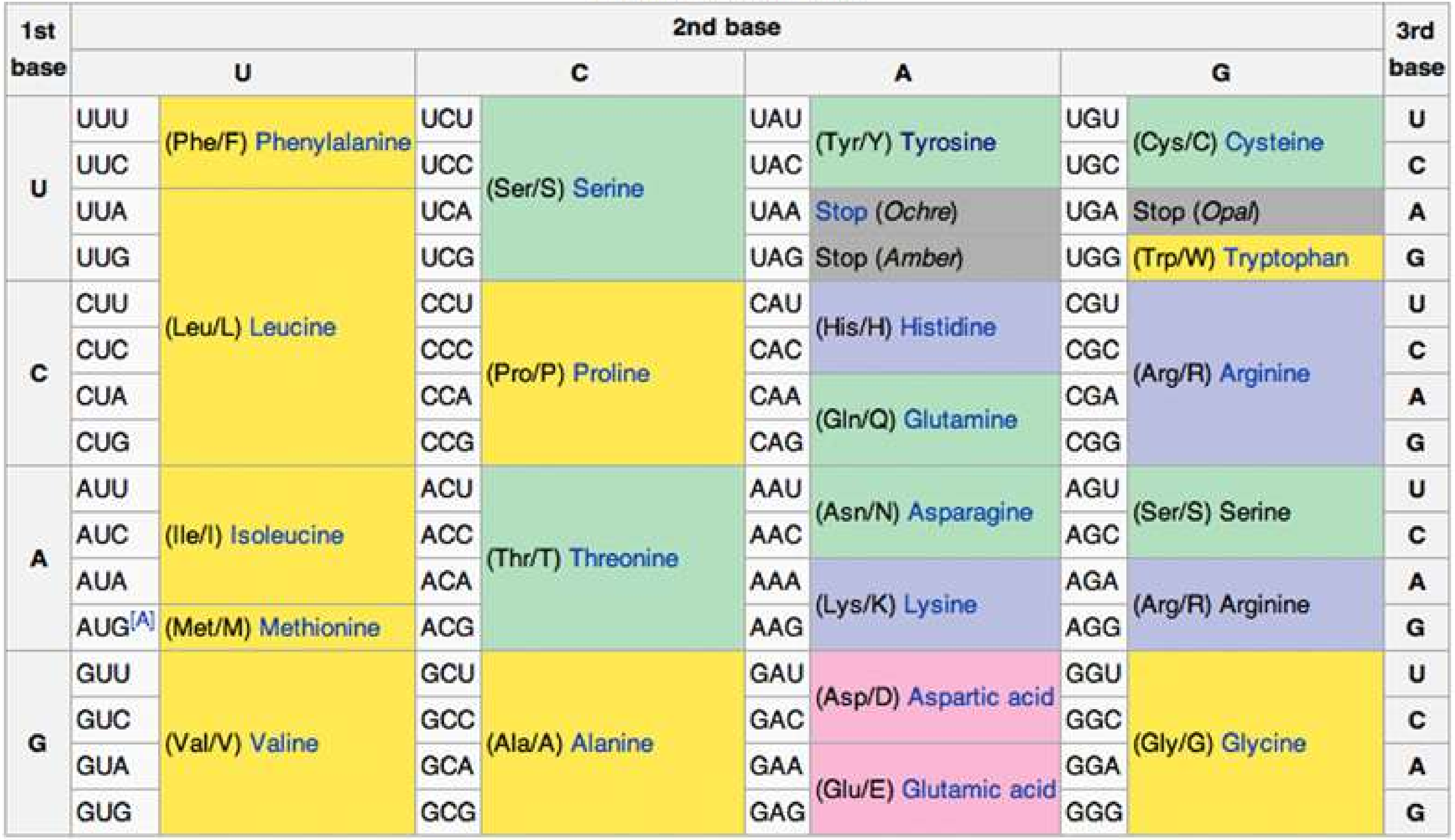


3. Ribosome subunits separate. ...when the ribosome separates from the mRNA, and the two ribosomal subunits dissociate. The subunits are ready to attach to the start codon of another message and start translation anew.

Translation: RNA → proteins – the genetic code

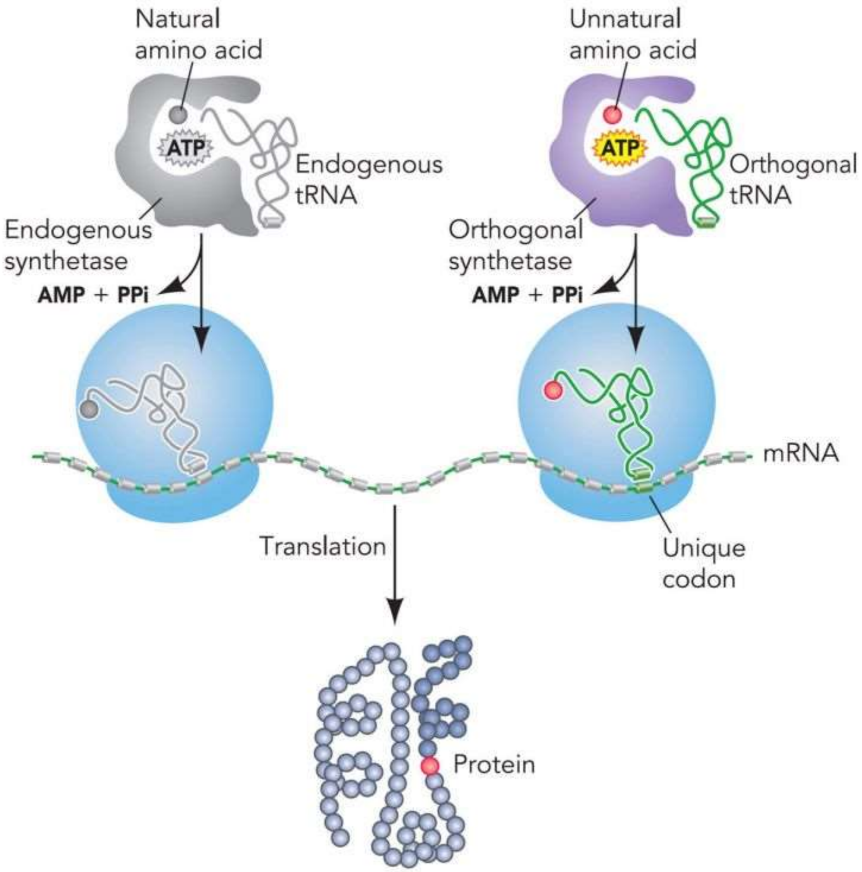
nonpolar polar basic acidic (stop codon)

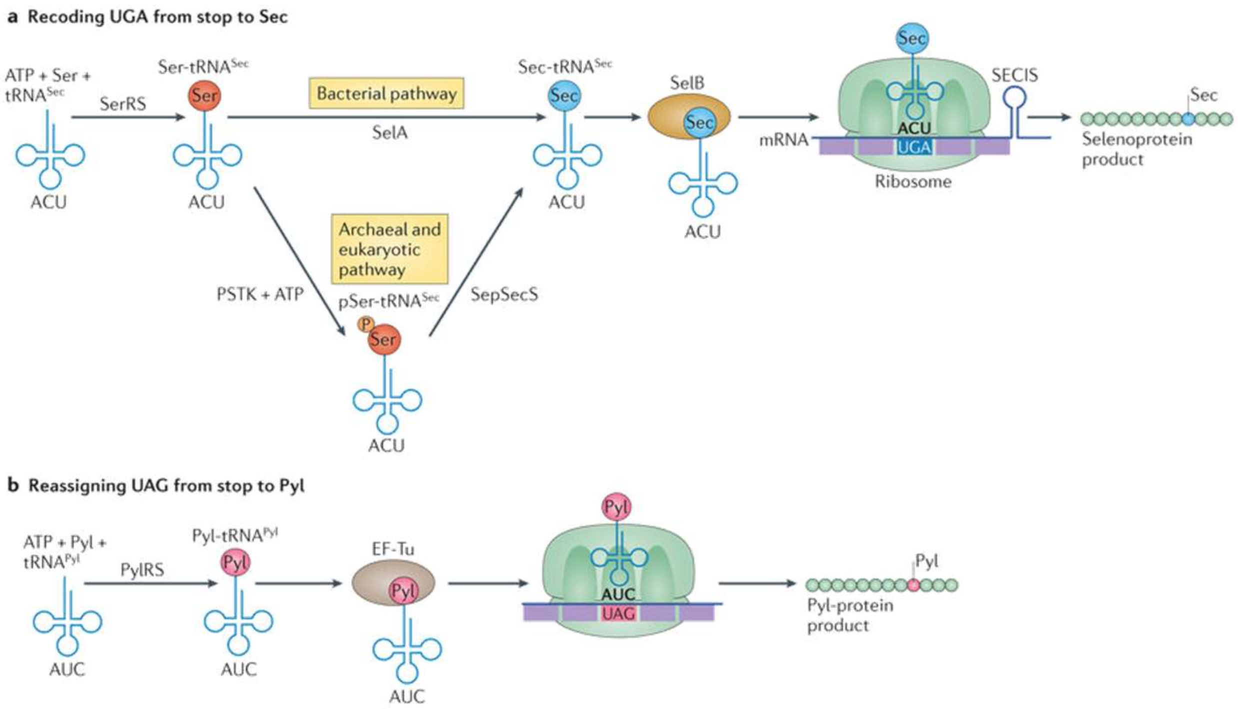
Standard genetic code



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

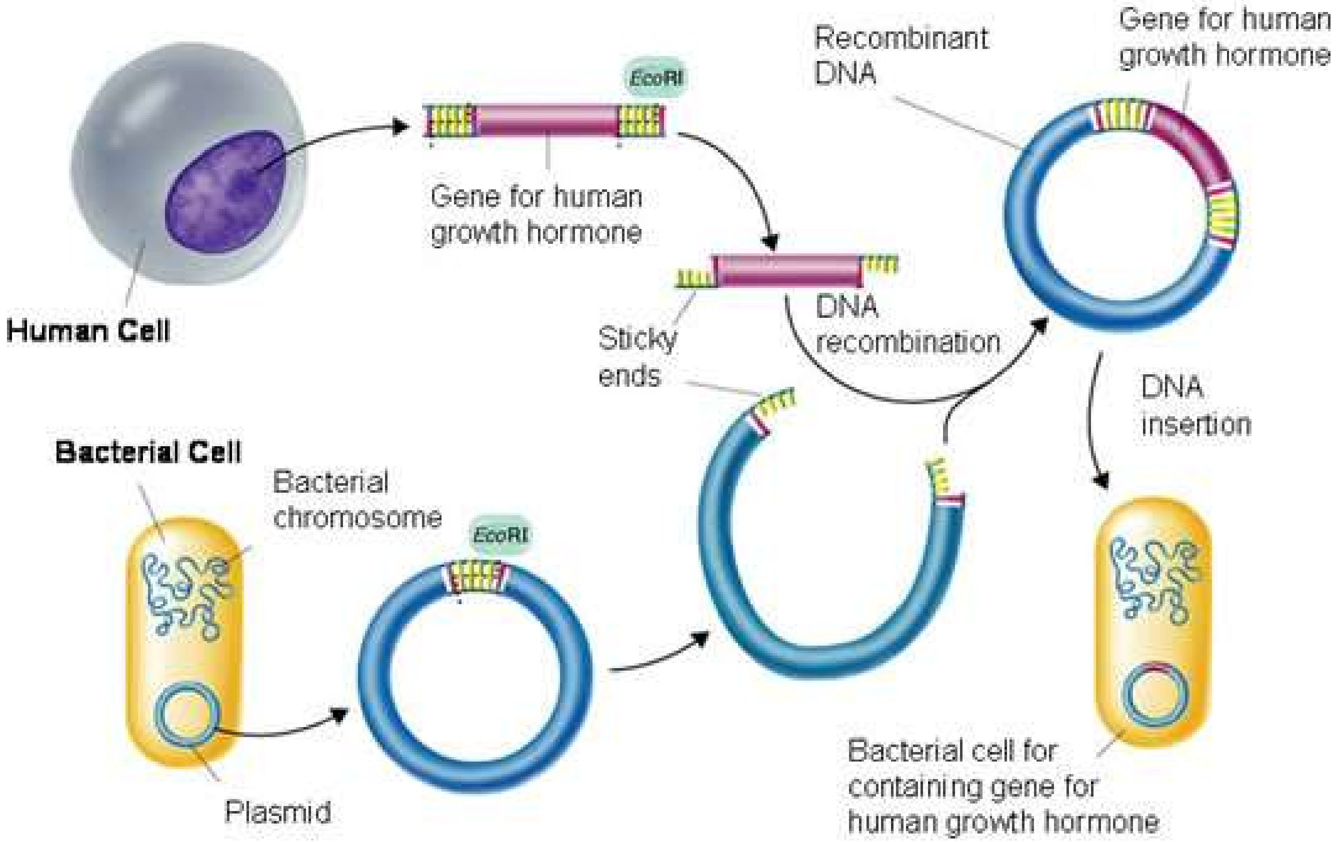
Expanded genetic code





Nature Reviews | Microbiology

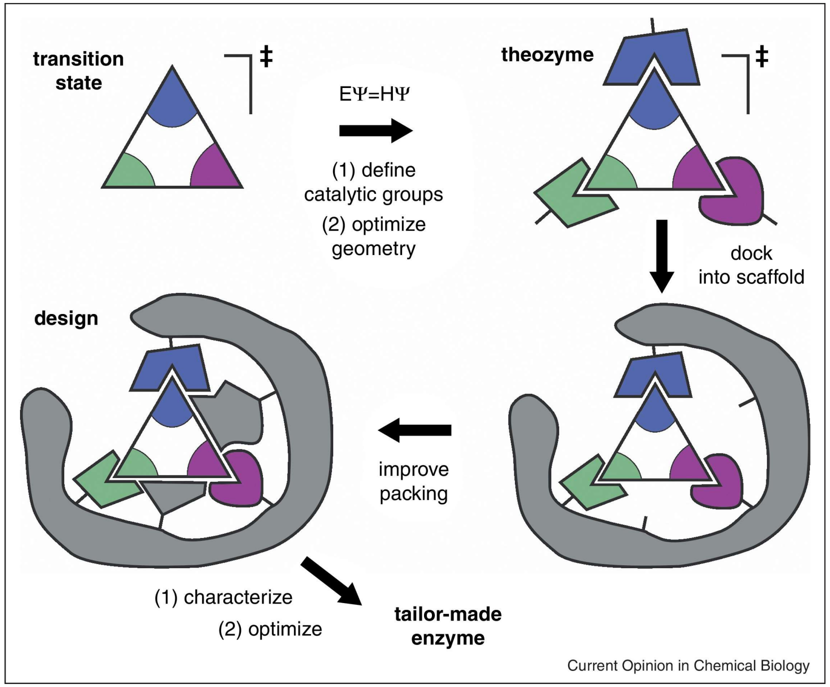
Recombinant proteins



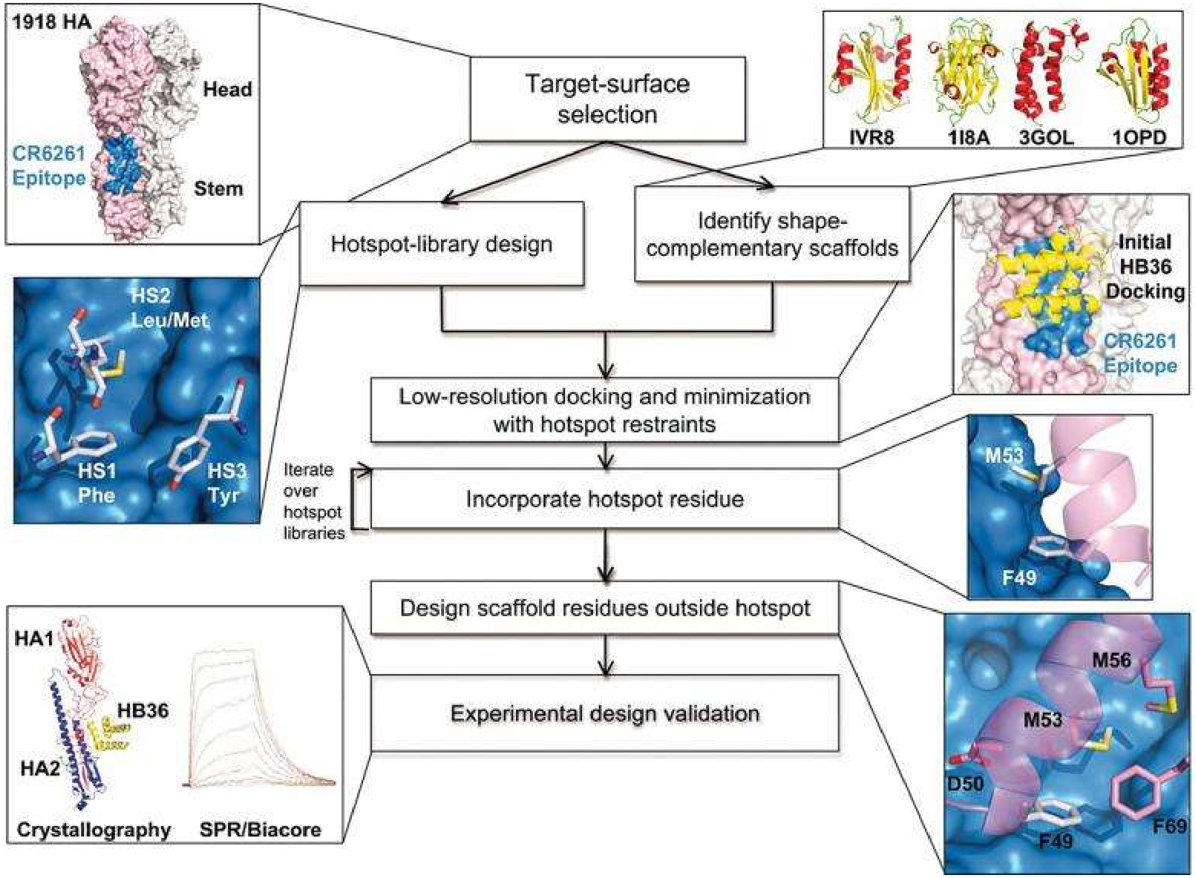
Recombinant insulin:

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

de novo enzyme design

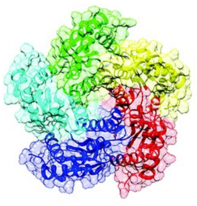


Current Opinion in Chemical Biology

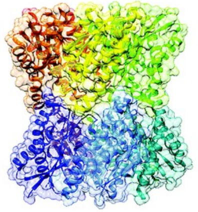


For interested in details:
Introduction to protein design – Part1: https://www.youtube.com/watch?v=0LetJMbu7uY&ab_channel=iBiology
Introduction to protein design – Part2: https://www.youtube.com/watch?v=ZrAwWx7meTk&ab_channel=iBiology

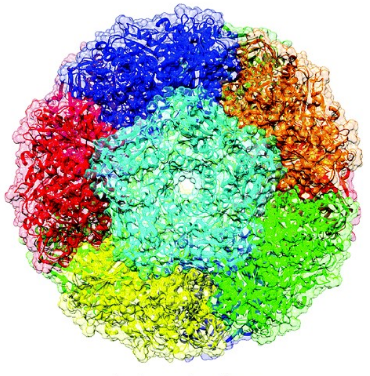
Evolvability of proteins from thermophiles



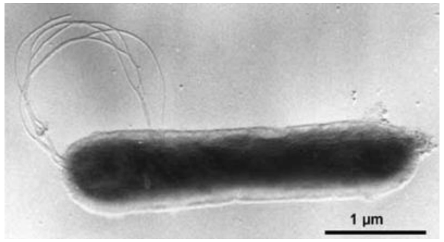
pentamer
Schizosaccharomyces pombe



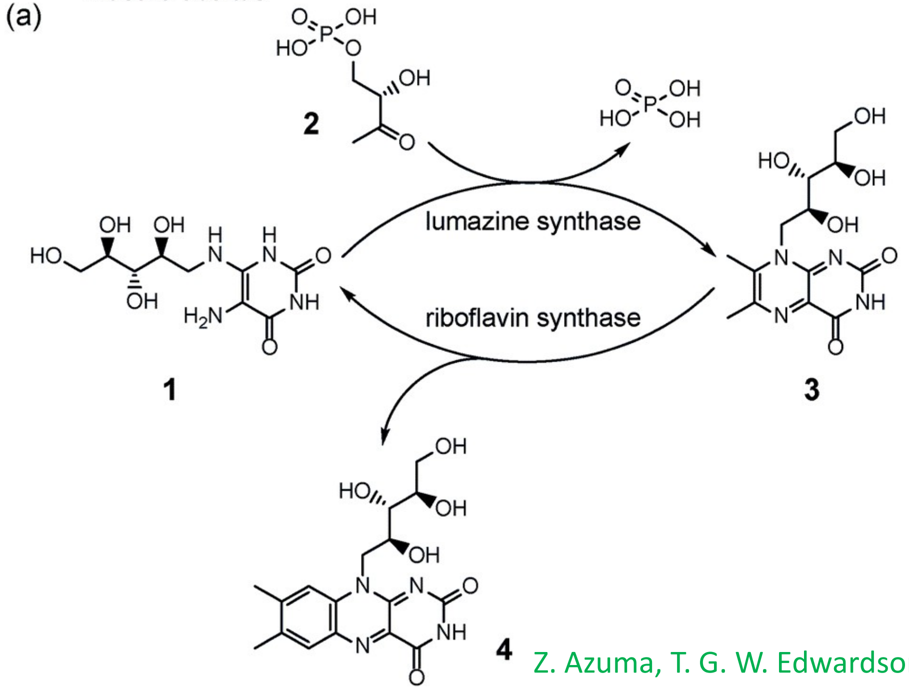
dimer of pentamers
Brucella abortus

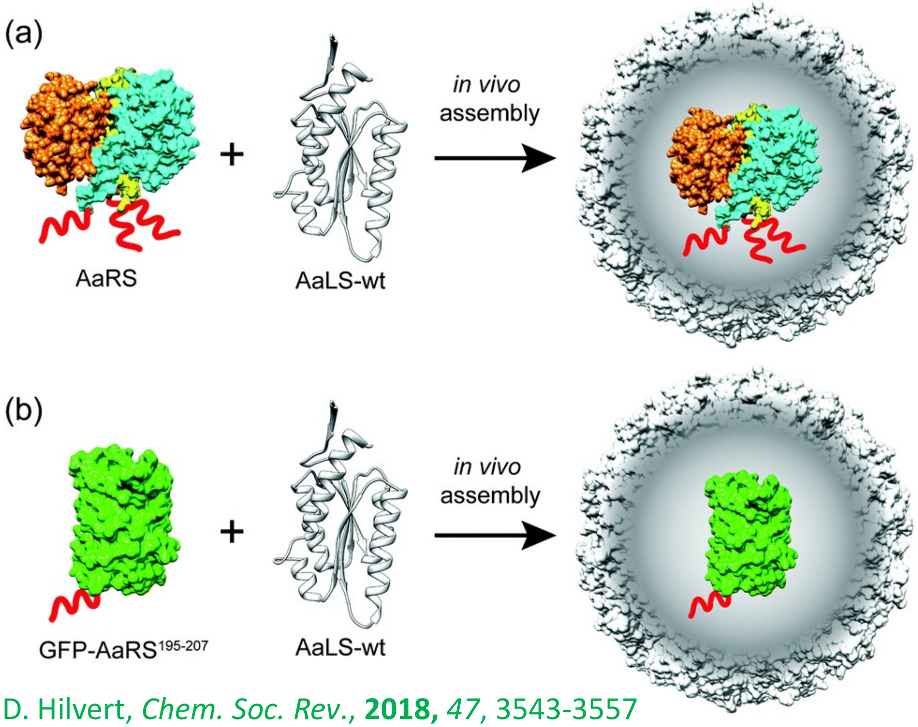


dodecahedron
formed from 12 pentamers
Aquifex aeolicus

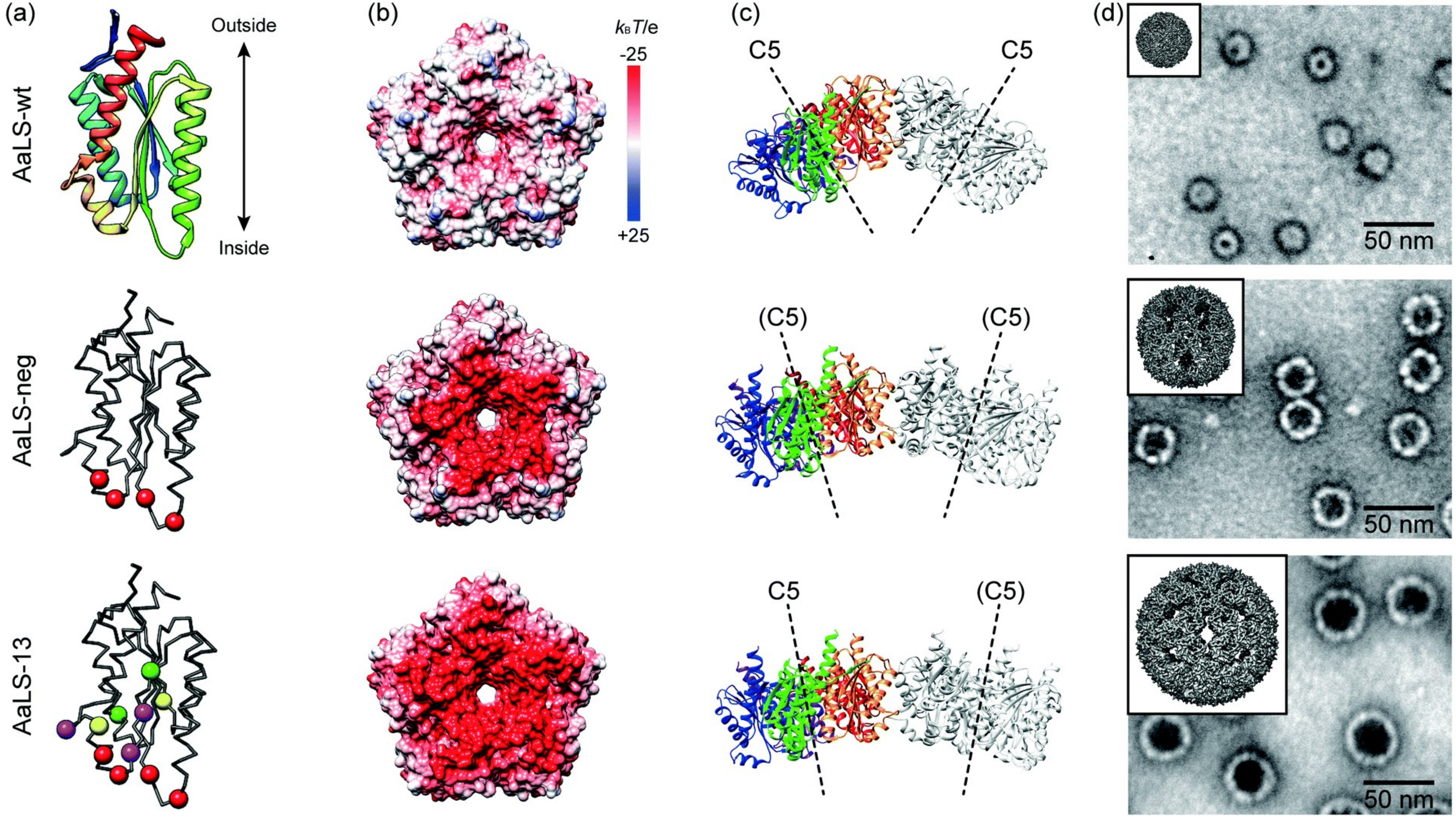


Aquifex aeolicus (Aa) – thermophilic bacteria

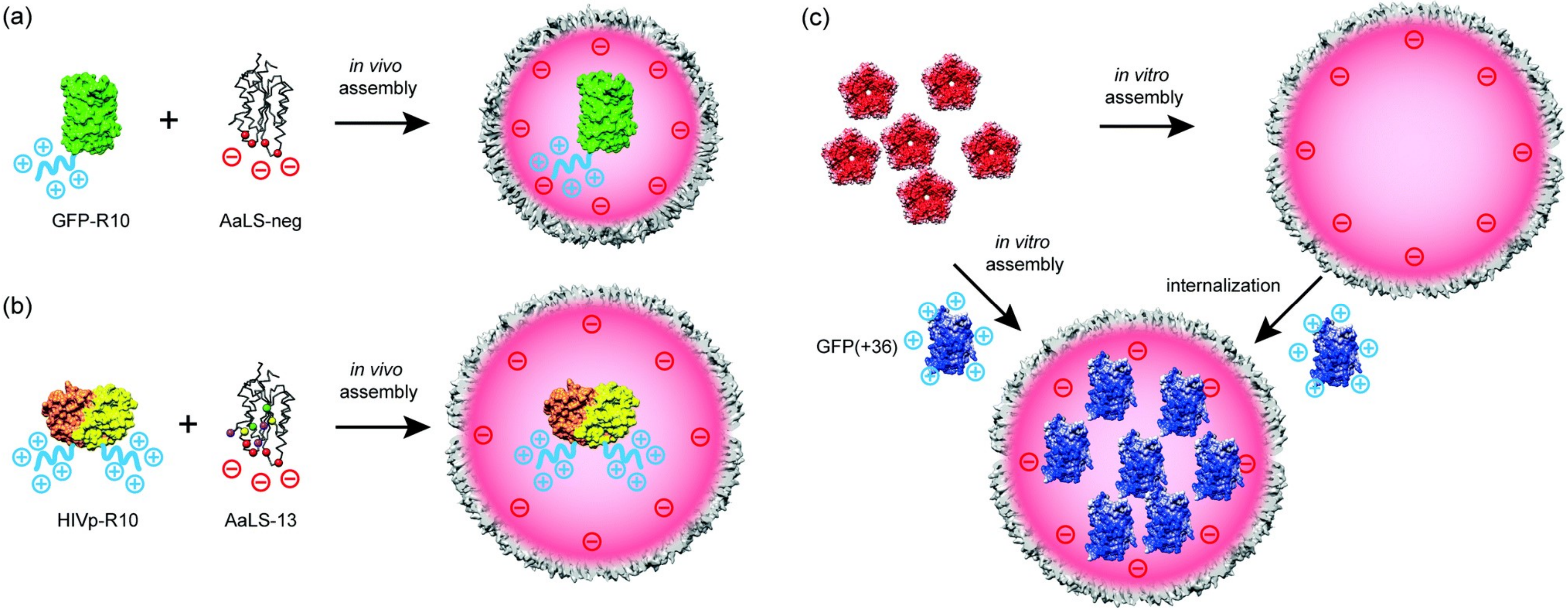




Z. Azuma, T. G. W. Edwardson and D. Hilvert, *Chem. Soc. Rev.*, **2018,** *47*, 3543-3557



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;

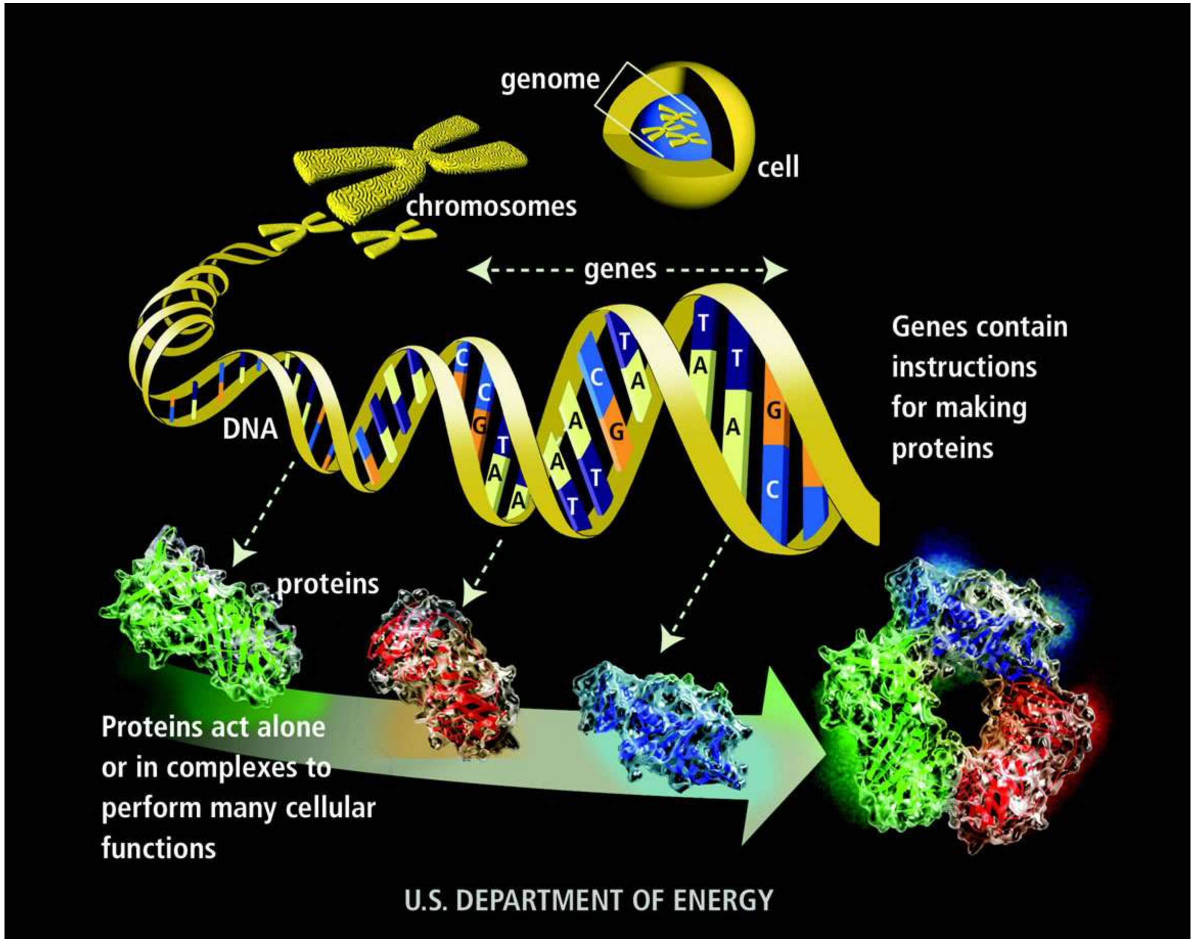
design of artificial organisms

CHAPTER 1

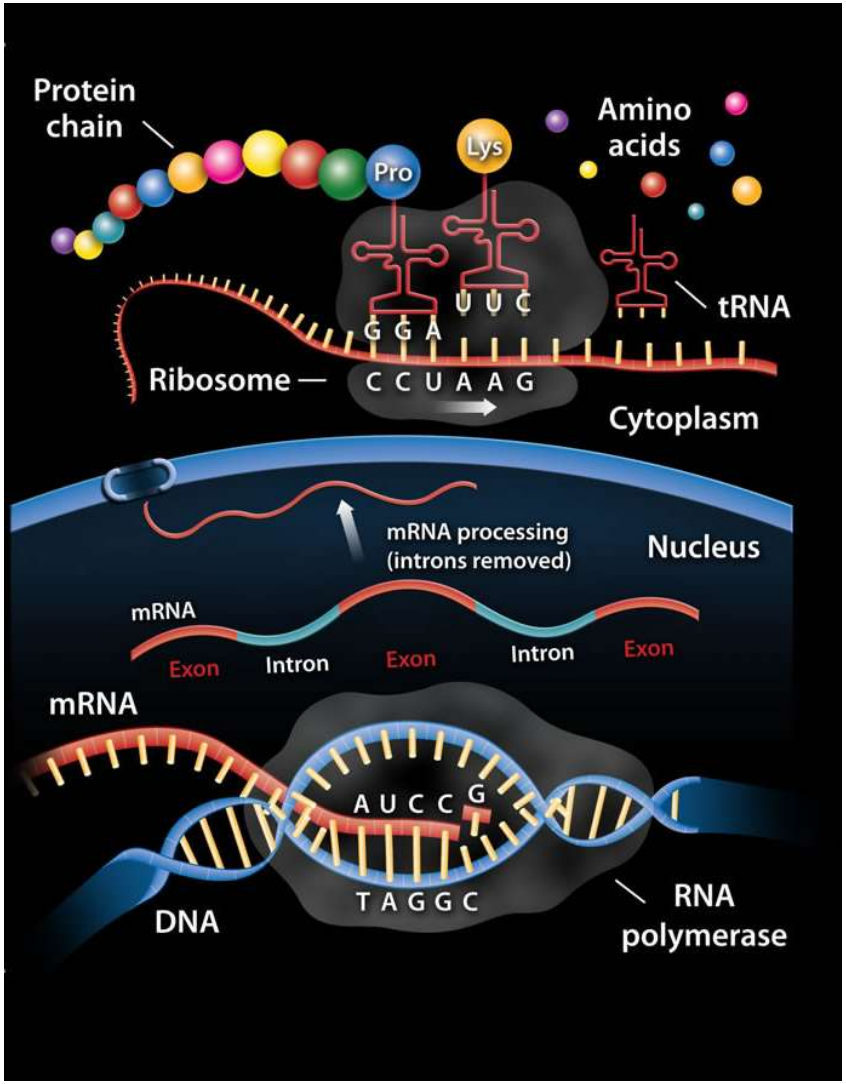
OLIGONUCLEOTIDES

The less common side of RNA

From DNA to proteins



U.S. DEPARTMENT OF ENERGY

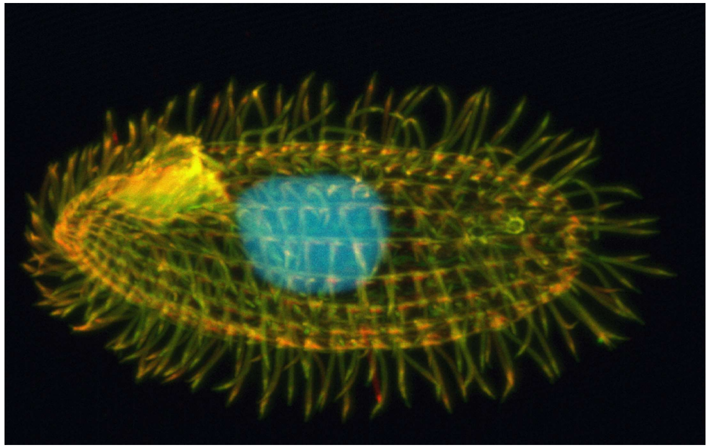


Ribozymes

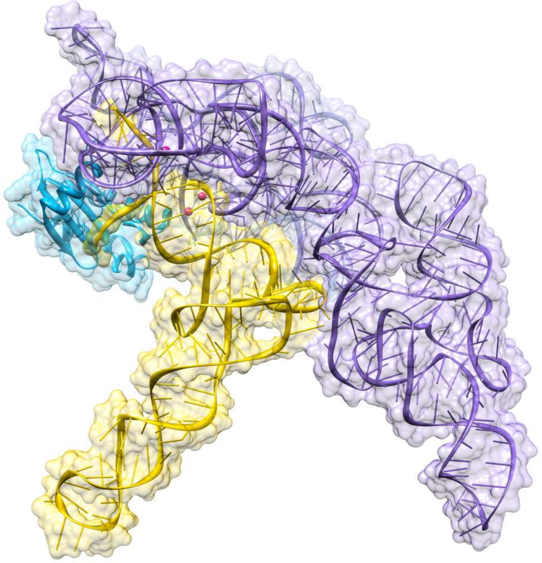
Ribozymes – Ribonucleic acid enzymes

1989 – Thomas Cech and Sidney Altman – Nobel Prize in chemistry for discovery of catalytic RNA

Thomas R. Cech was studying RNA splicing in the ciliated protozoan *Tetrahymena thermophila*
Sidney Altman and Norman Pace were studying the bacterial RNase P complex.

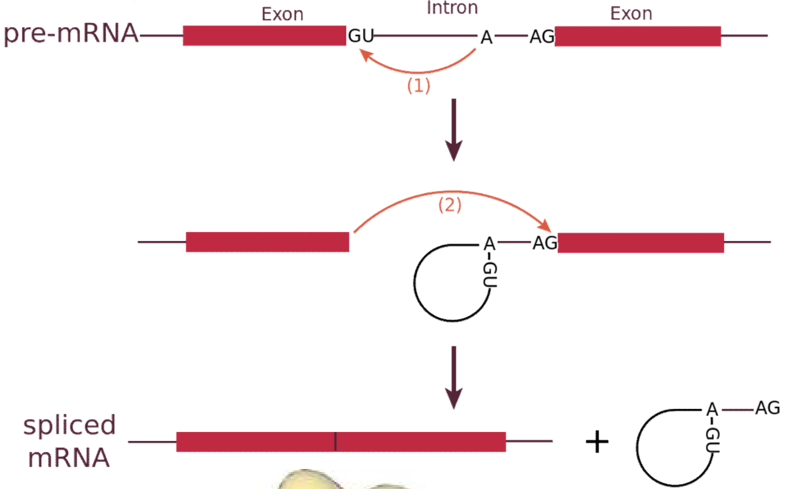


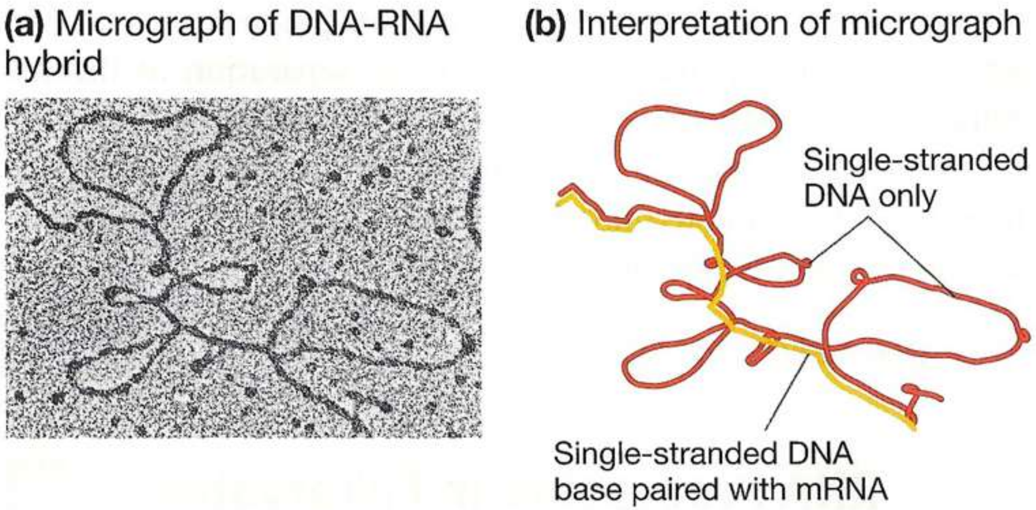
Tetrahymena thermophila



Bacterial RNAse P

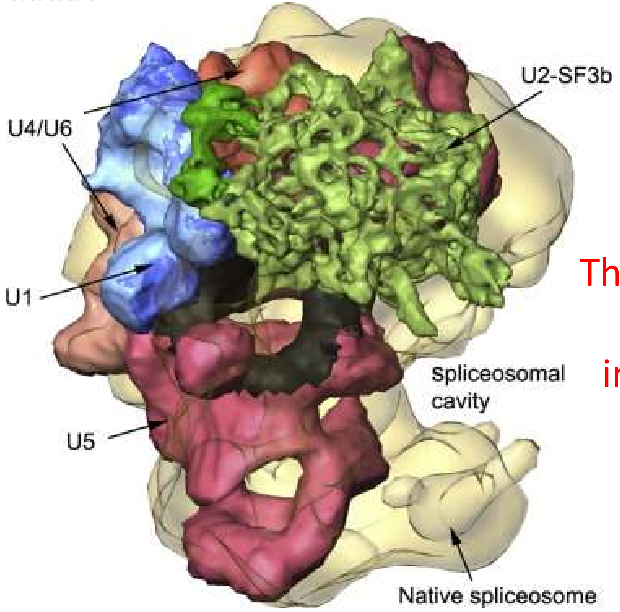
mRNA processing




(a) Micrograph of DNA-RNA hybrid

(b) Interpretation of micrograph

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.



The spliceosome carries out the removal of introns as RNA lariats

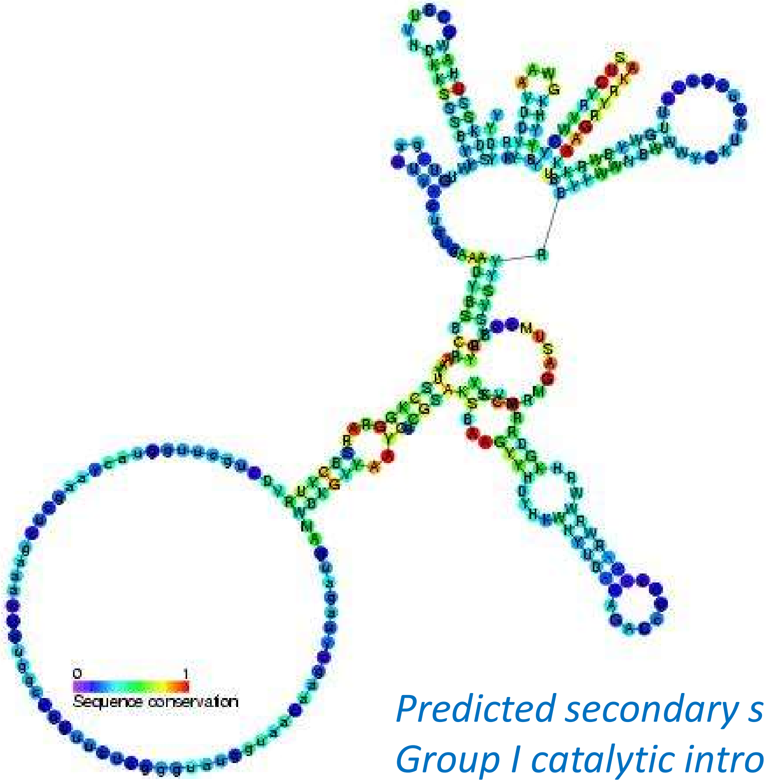
Spliceosome – a complex of ribonucleoproteins

RNA splicing

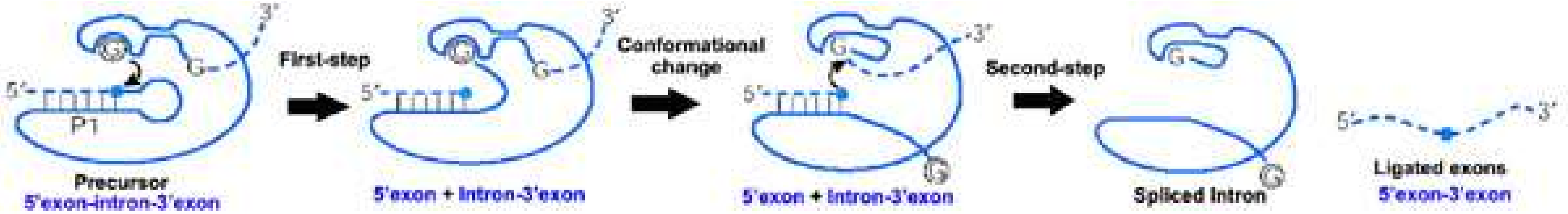
Self-splicing RNA introns

RNA splicing in *Tetrahymena* was taking place also in absence of the spliceosome - the ‚negative control' obtained after protease digestion also spliced.

In contrary to the spliceosome, the ***catalytic*** motif ***does not*** contain protein part, ***only RNA***.
First known example of a ***ribozyme*** – ribonucleic acid-composed enzyme analogue.



Predicted secondary structure and sequence conservation of Group I catalytic intron

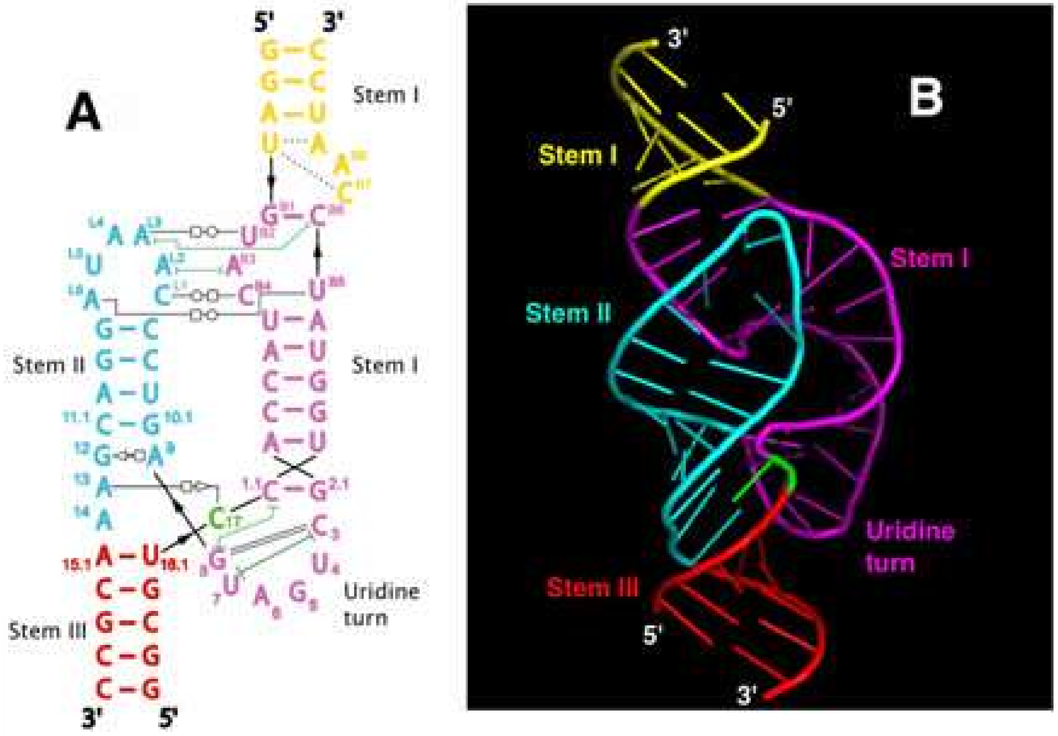


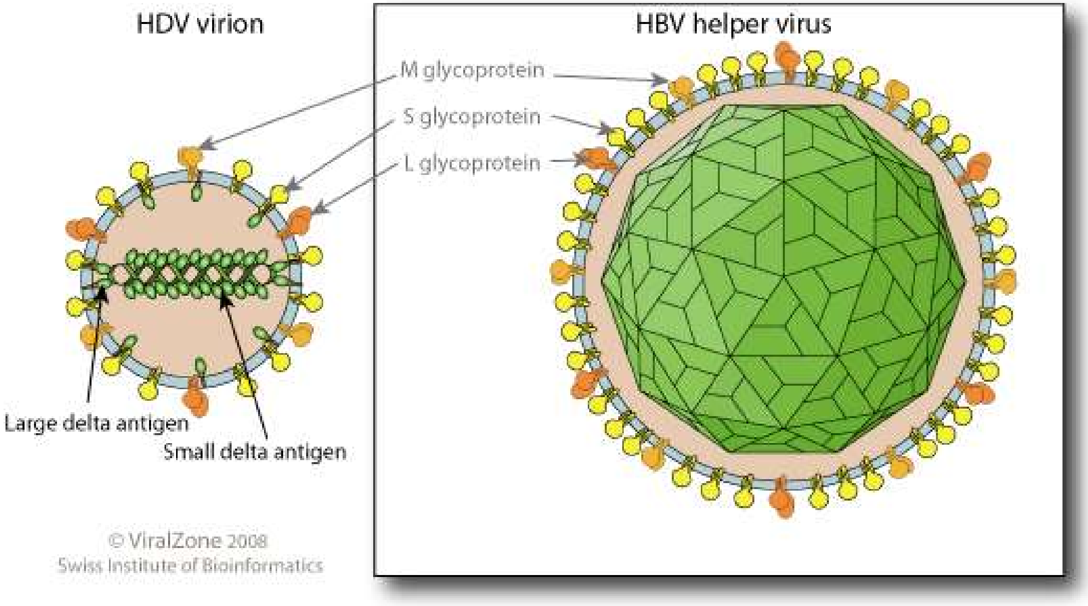
Ribozymes

Hammerhead ribozyme

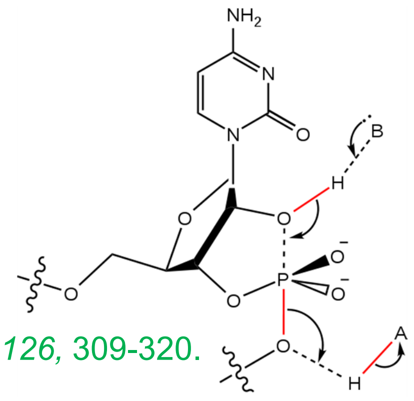
The hammerhead ribozyme is a RNA molecule motif that catalyzes reversible cleavage and joining reactions at a specific site within an RNA molecule (model system; targeted RNA cleavage experiments)

The *hepatitis delta virus (HDV) ribozyme* is a non-coding RNA found in the hepatitis delta virus that is necessary for viral replication and is thought to be the only catalytic RNA known to be required for viability of a human pathogen.





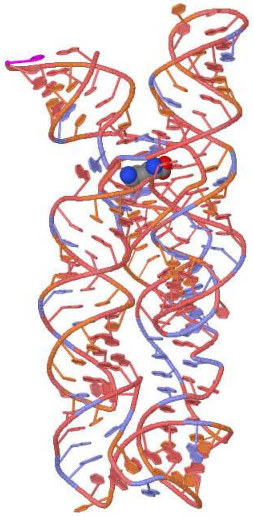
© ViralZone 2008
Swiss Institute of Bioinformatics



M. Martick and W. G. Scott, *Cell* **2006**, *126*, 309-320.

Riboswitches

2002 - (Breaker and Nudler) – discovery of a nucleic acid-based genetic regulatory element – ***riboswitch***.



The lysine riboswitch

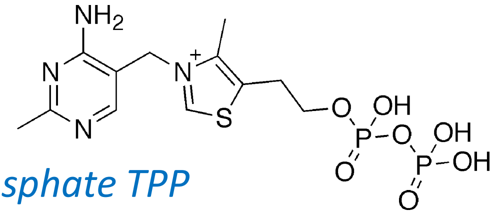
Riboswitches - naturally occurring regulatory segments of mRNA that bind small molecules specifically. The binding results in a change in production of the proteins encoded by the mRNA

Before discovery of ***riboswitches*** only ***proteins*** were supposed to do so in the biological context.

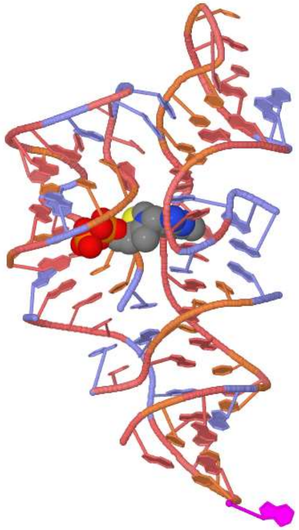
Most known ***riboswitches*** occur in bacteria, but functional riboswitches of one type (the TPP riboswitch) have been discovered in archaea, plants and certain fungi.

Riboswitches exist in all domains of life, and therefore are likely that they might represent ancient regulatory systems or fragments of ***RNA-world ribozymes*** whose binding domains remained conserved throughout the evolution

The **TPP riboswitch** (THI element and Thi-box riboswitch), is a highly conserved RNA secondary structure. It binds directly to thiamine pyrophosphate (TPP, a form of the vitamin B1, an essential coenzyme) to regulate gene expression through a variety of mechanisms in archaea, bacteria and eukaryotes.



Thiamine pyrophosphate TPP



The 3D structure of TPP riboswitch (by Benjamin Schuster-Böckler)

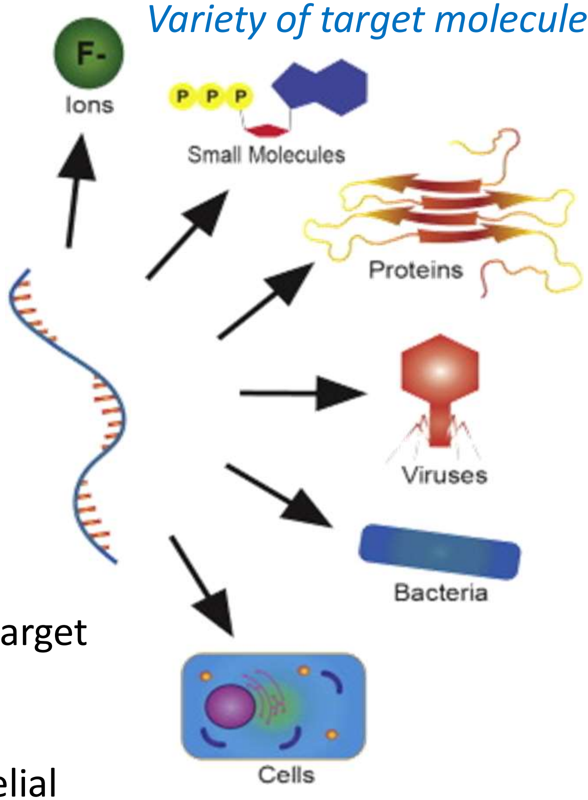
Aptamers

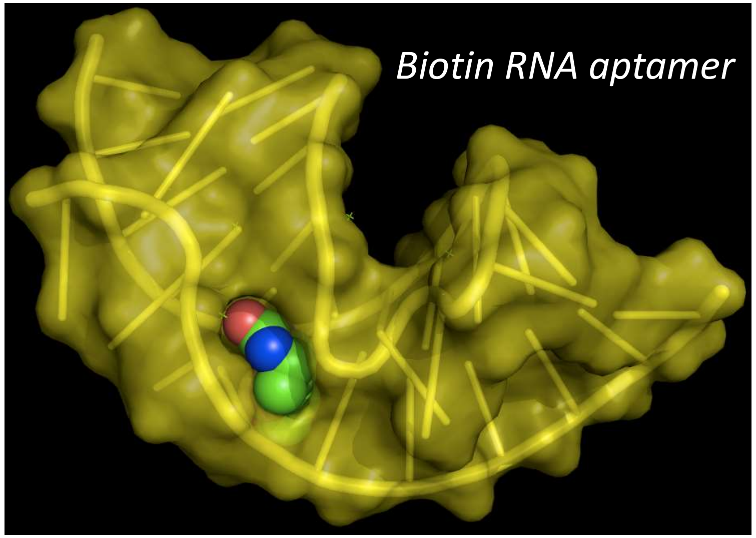
Aptamers (from the Latin *aptus* – fit, and Greek *meros* – part) are ***oligonucleotide*** or ***peptide*** molecules that bind to a specific target molecule.

Aptamers are usually created by selecting them from a large random sequence pool, but natural aptamers also exist in riboswitches.

- ***DNA or RNA or XNA aptamers*** – oligonucleotide strands (usually short)
- ***Peptide aptamers*** - one (or more) short variable peptide domains, attached at both ends to a protein scaffold.

Variety of target molecules





Fdardel

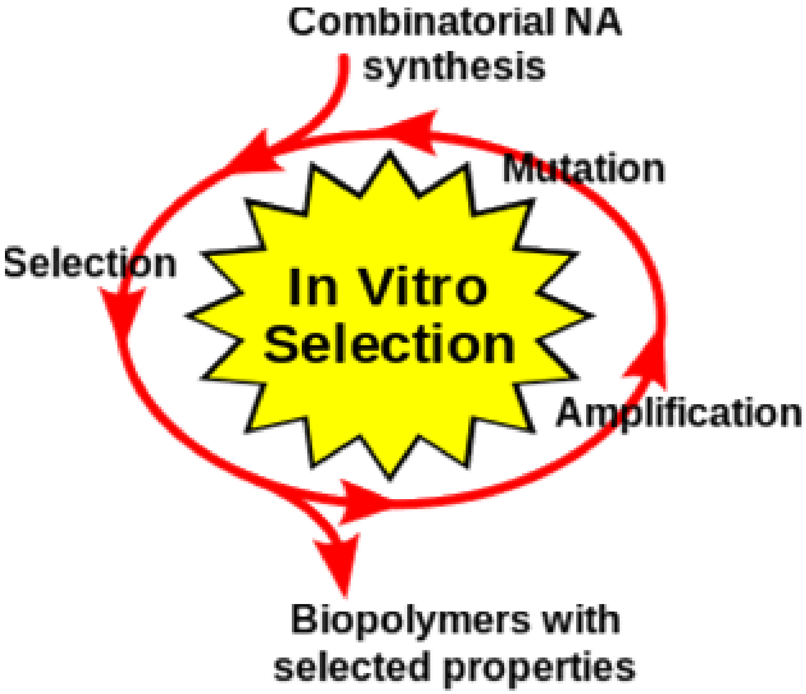
Aptamers were evolved for a variety of target ligands:

- small molecules (ATP and adenosine)
- proteins: prions and vascular endothelial growth factor (VEGF) - **MACUGEN**,
- tumor cells.

Additionally, SELEX has been utilized to obtain highly specific catalytic DNA or **DNAzymes**. Several metal-specific **DNAzymes** have been reported including the GR-5 DNAzyme (**lead-specific**), the CA1-3 DNAzymes (**copper-specific**), the 39E DNAzyme (**uranyl-specific**) and the NaA43 DNAzyme (**sodium-specific**).

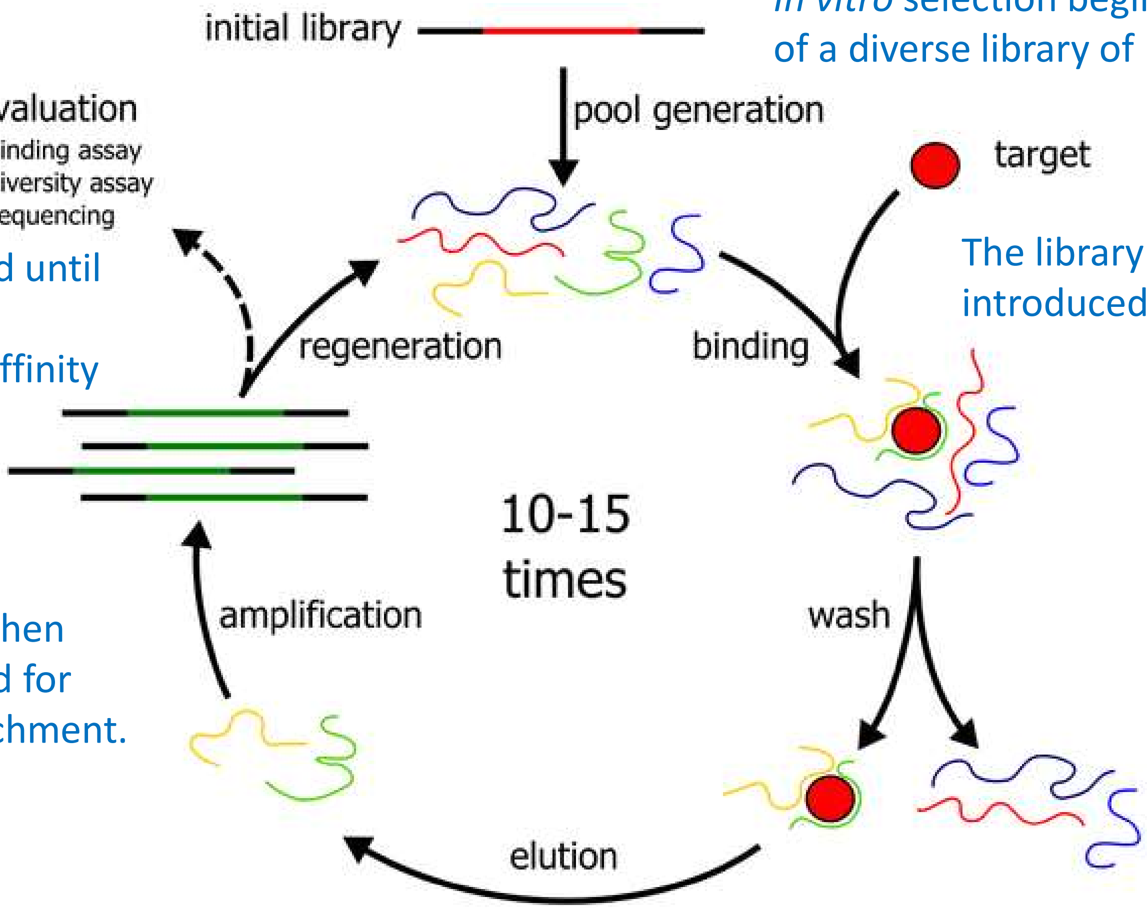
Systematic evolution of ligands by exponential enrichment - SELEX

1990 – Gold *et al.* – selection of RNA ligands against T4 DNA polymerase
1990 – J. Szostak *et al.* – selecting RNA ligands towards organic dyes



A general overview of in vitro selection protocol. NA stands for Nucleic Acids (DNA, RNA) which start as a random pool, and are enriched through the selection process

Systematic evolution of ligands by exponential enrichment - SELEX



In vitro selection begins with the generation of a diverse library of DNA or RNA molecules.

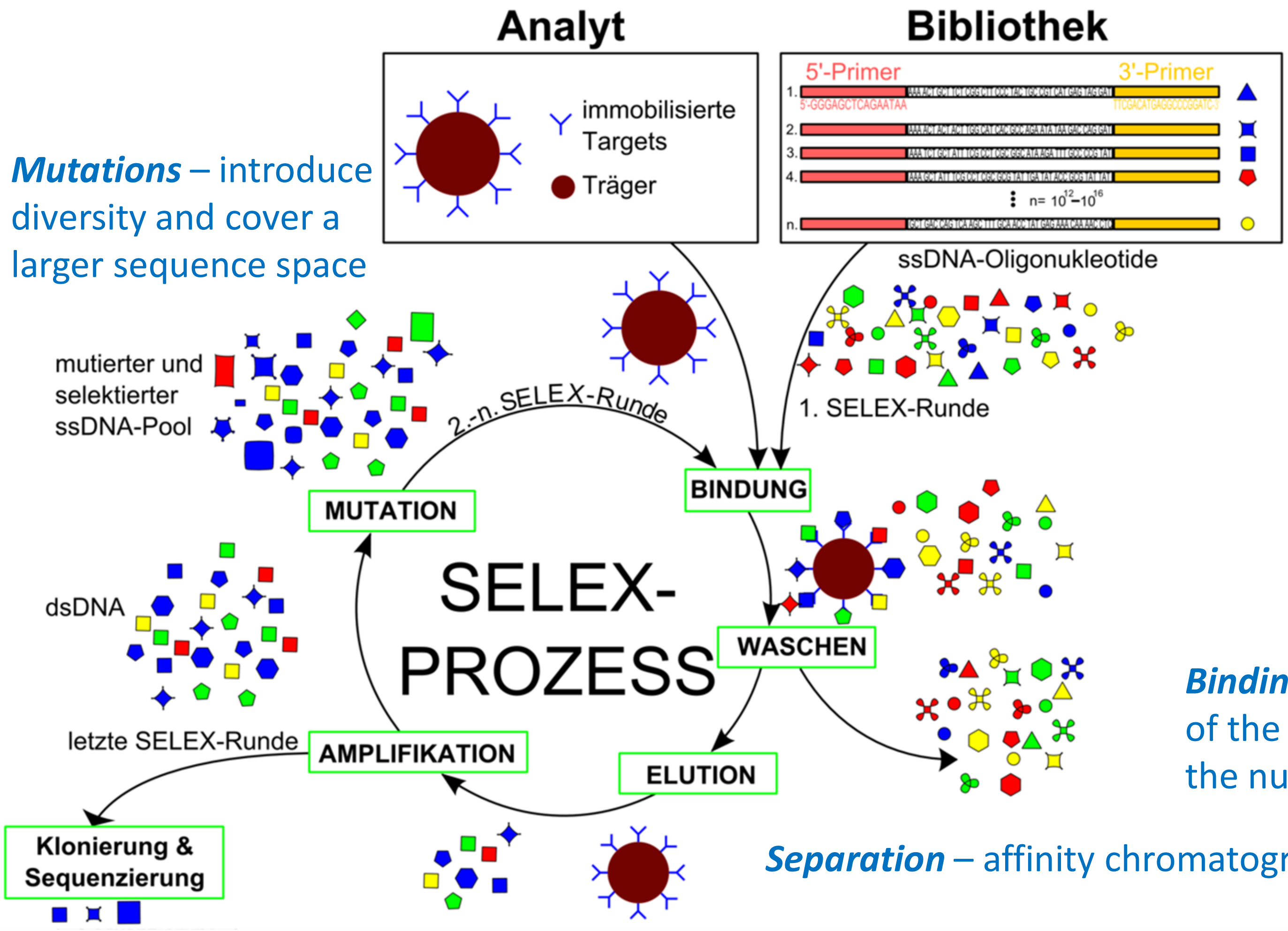
The library is then introduced to a target ligand

Multiple rounds are performed until the library converges on to a collection of sequences with affinity for the target molecule.

The bound sequences are then collected and PCR amplified for subsequent rounds of enrichment.

Sequences demonstrating affinity towards the target molecule are isolated from any unbound sequences.

Systematic evolution of ligands by exponential enrichment - SELEX



Mutations – introduce diversity and cover a larger sequence space

The initial library – usually contains sequences of 20-200 randomised nucleotides

The 5'- and 3'-termini are composed of constant primer sequences. Between them, the random sequences of *n* nucleotides result in 4^n combinations:

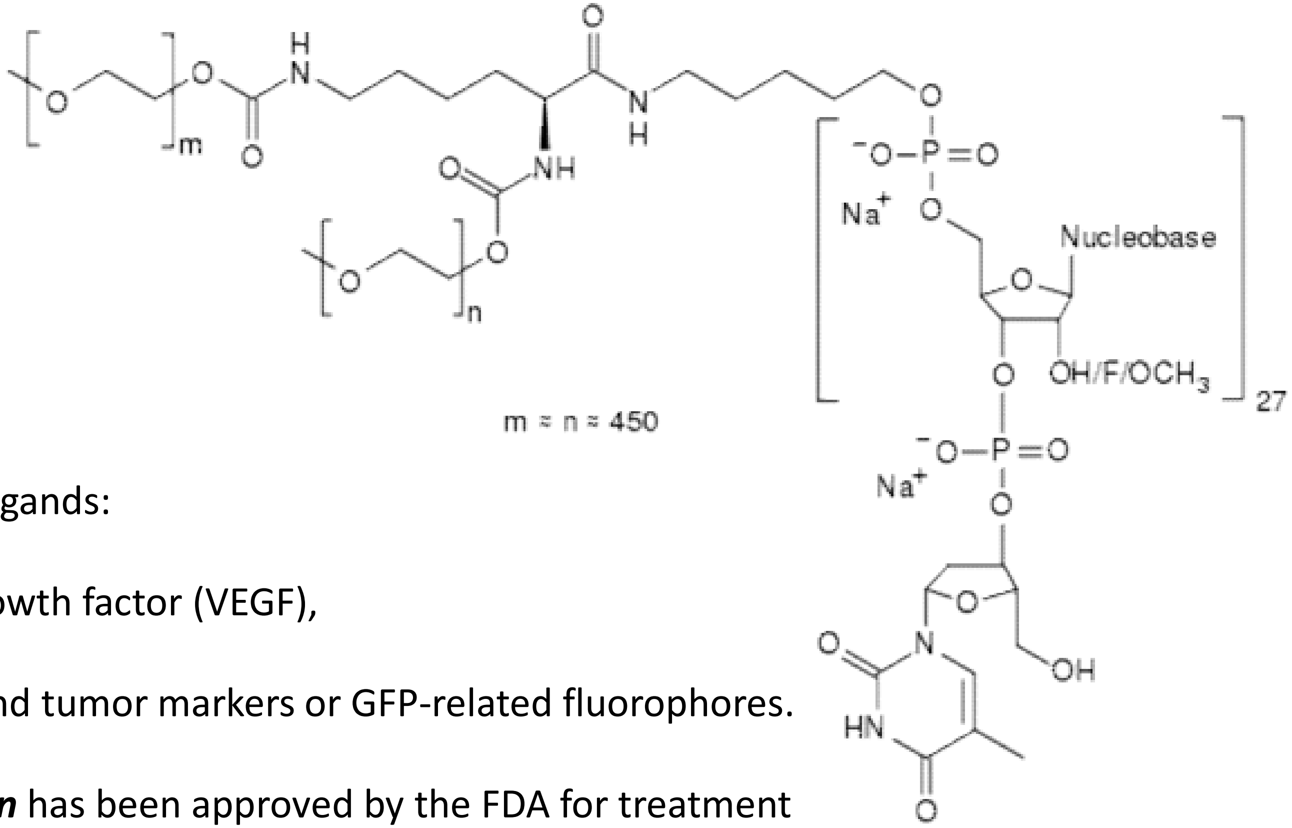
- n = 25 – a library of c.a. 10^{15} sequences,
- n = 50 – a library of c.a. 10^{30} sequences,
- n = 75 – a library of c.a. 10^{45} sequences,
- n = 100 – a library of c.a. 10^{60} sequences.

Binding strongly depends on the secondary structure of the oligonucleotides, which in turn depends on the nucleobase sequence

Separation – affinity chromatography

Amplification – reverse transcription RNA→DNA (only for RNA aptamers) + PCR (for RNA and DNA aptamers)

Systematic evolution of ligands by exponential enrichment - SELEX



Aptamers were evolved for a variety of target ligands:

- small molecules (ATP and adenosine)
- proteins: prions and vascular endothelial growth factor (VEGF),
- tumor cells.

Clinical uses are suggested by aptamers that bind tumor markers or GFP-related fluorophores.

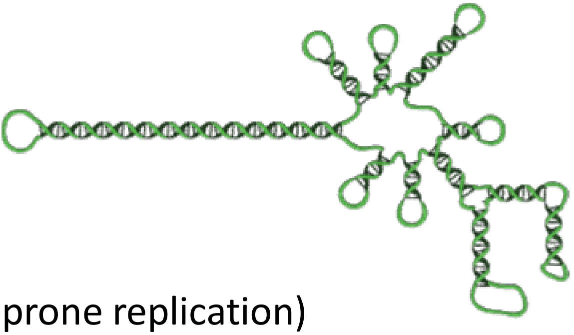
A VEGF-binding aptamer trade-named ***Macugen*** has been approved by the FDA for treatment of macular degeneration.

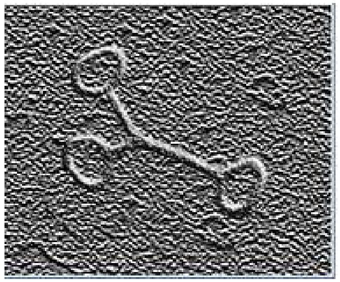
Additionally, SELEX has been utilized to obtain highly specific catalytic DNA or DNAzymes. Several metal-specific DNAzymes have been reported including the GR-5 DNAzyme (lead-specific), the CA1-3 DNAzymes (copper-specific), the 39E DNAzyme (uranyl-specific) and the NaA43 DNAzyme (sodium-specific).

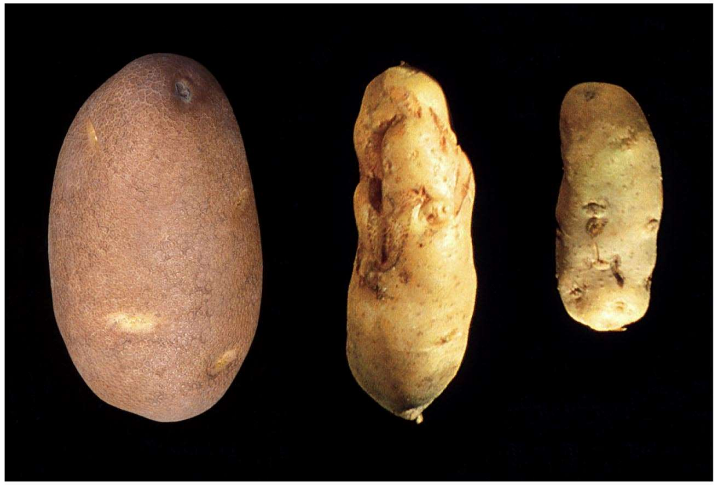
Macugen

Viroids

Viroids ("subviral pathogens„) are mostly plant pathogens, which consist of short stretches of highly complementary, circular, single-stranded, and non-coding RNA without a protein coat. Viroids are extremely small - 246 to 467 nucleobases (genomes of smallest viruses start from 2,000 nucleobases). Viroids are plausible "living relics" of the RNA world.







PSTVd-infected potatoes (right)

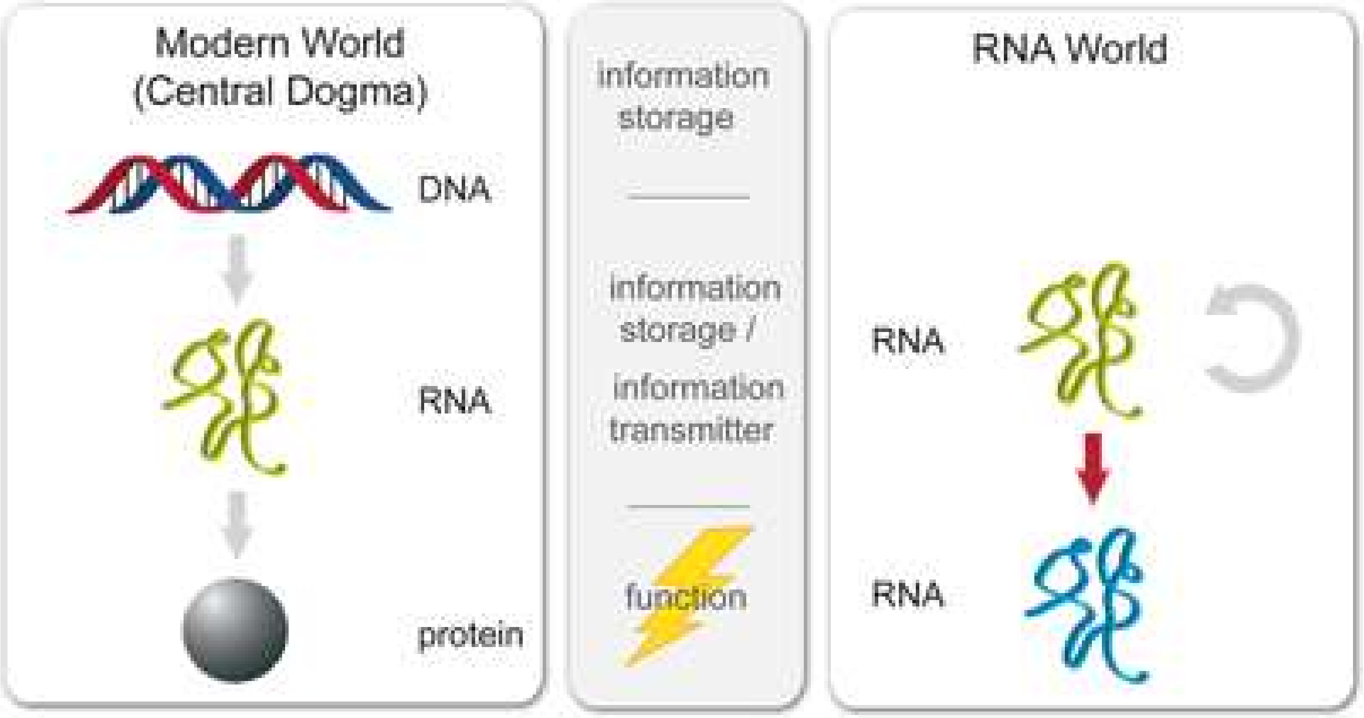
Viroid properties:

- small size (error-prone replication)
- high G-C content, (stability and replication fidelity)
- circular structure (complete replication without genomic tags)
- lack of protein-coding ability, consistent with a ribosome-free habitat; and replication mediated in some by ribozymes—the fingerprint of the RNA world.

Putative secondary structure of the PSTVd viroid

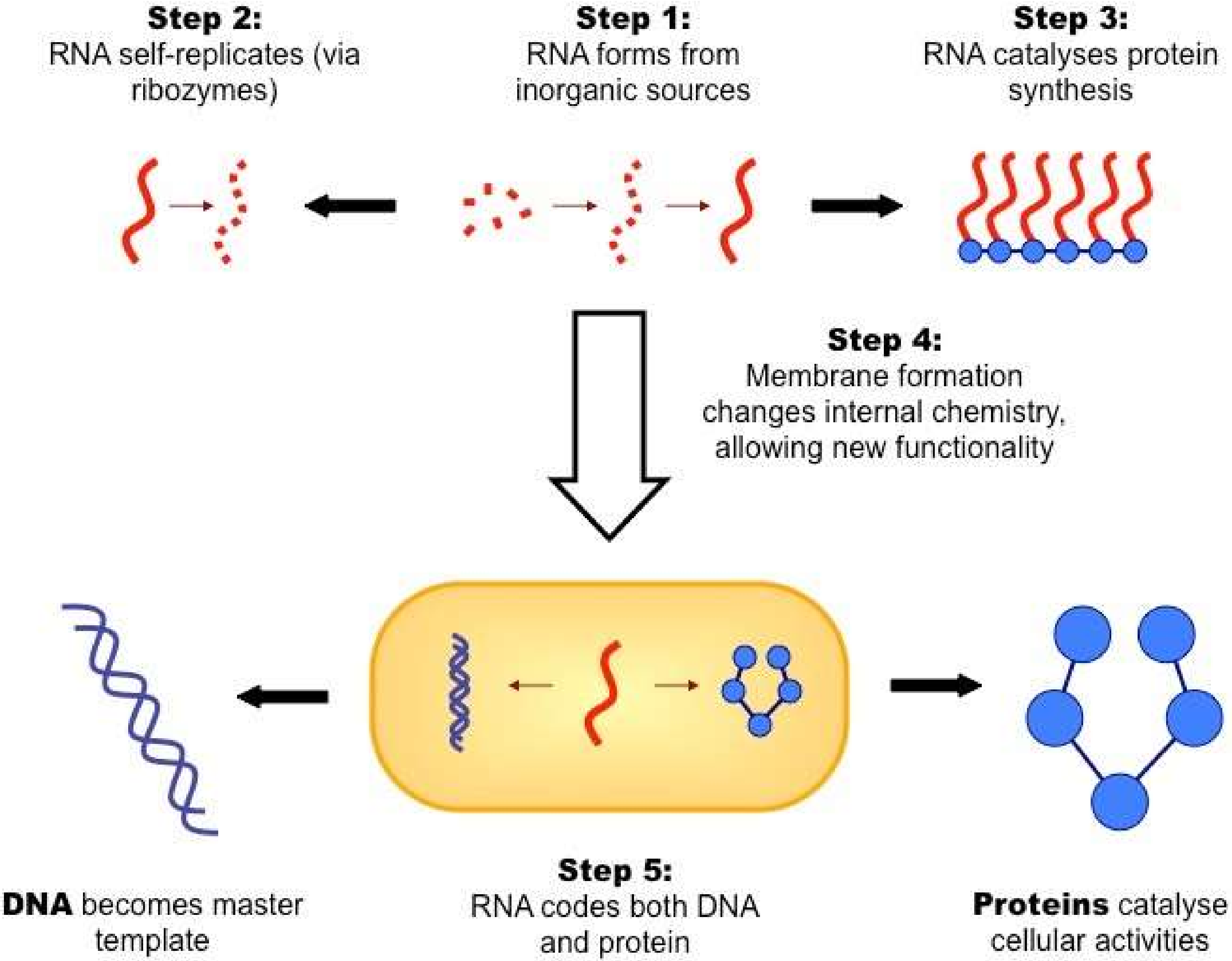
No virion reported. Viroids do not encode for proteins

The RNA world



Conceptual idea that there was a period in the early history of life on Earth when RNA (or its structurally simplified analogue) carried out most of the information processing and metabolic transformations needed for biology to emerge from chemistry

The RNA world



Step 2:
RNA self-replicates (via ribozymes)

Step 1:
RNA forms from inorganic sources

Step 3:
RNA catalyses protein synthesis

Step 4:
Membrane formation changes internal chemistry, allowing new functionality

DNA becomes master template

Step 5:
RNA codes both DNA and protein

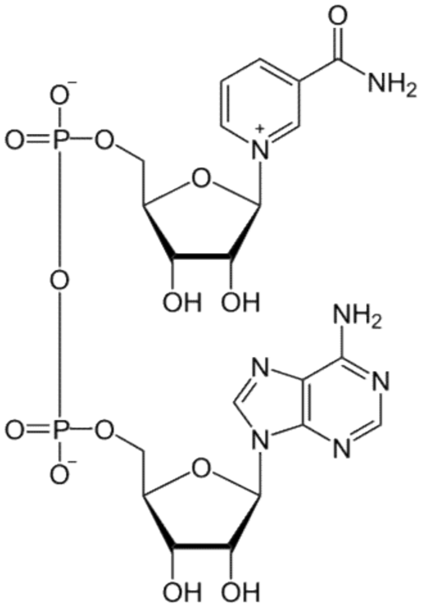
Proteins catalyse cellular activities

The RNA world

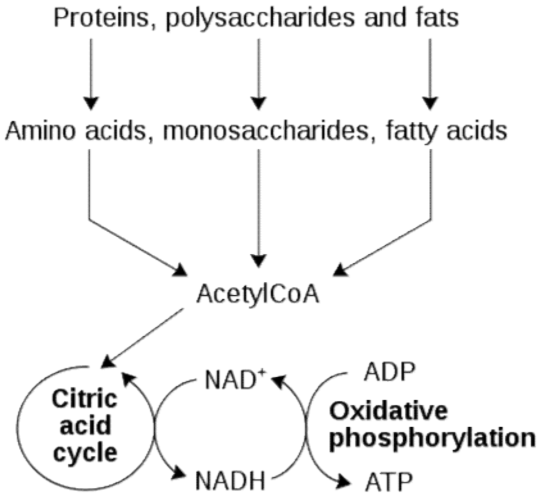
Crick, Orgel and Woese speculated in 1968 that, because RNA can form secondary structures, it has both a genotype and a phenotype and is a good candidate for the emergence of life

F. H. C. Crick *J. Mol. Biol.* **1968**, *38,* 367-379, L. E. Orgel *J. Mol. Biol.* **1968**, *38,* 381-393

Ribonucleotide coenzymes currently used by many proteins may be molecular „fossils" from the primordial RNA-based metabolism



Nicotinamide adenine dinucleotide (**NAD**$^+$)



Proteins, polysaccharides and fats

Amino acids, monosaccharides, fatty acids

AcetylCoA

Citric acid cycle

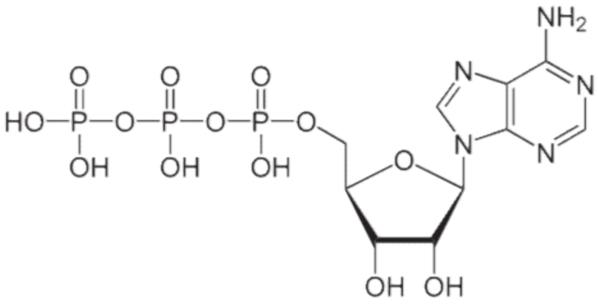
NAD$^+$

NADH

ADP

Oxidative phosphorylation

ATP

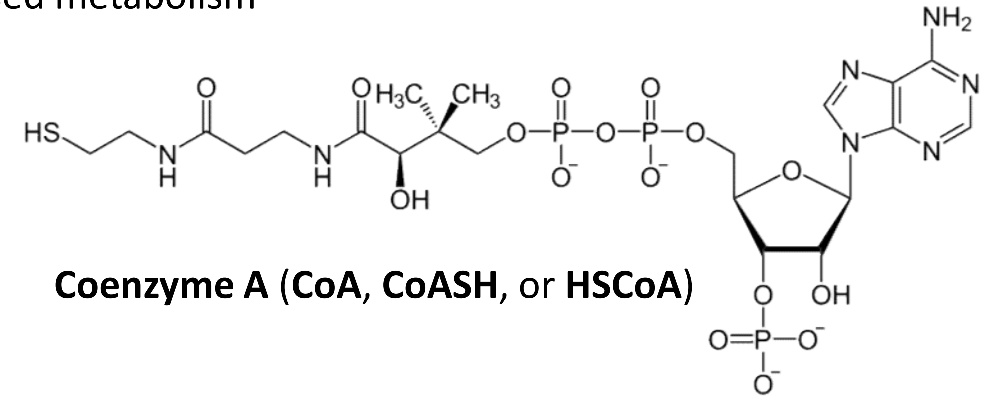


Adenosine triphosphate (**ATP**)

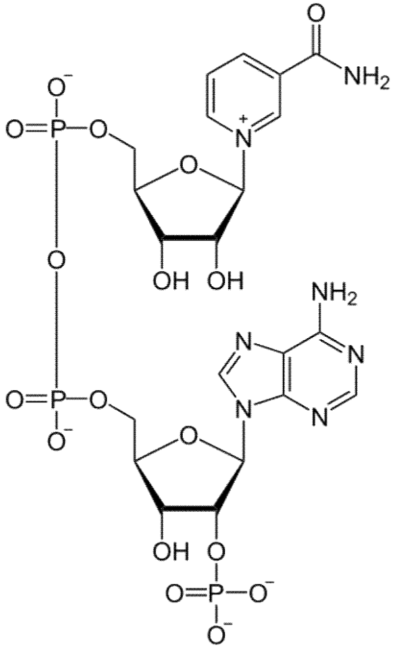
H. B. White III *J. Mol. Evol.* **1976**, *7,* 101-104

The RNA world

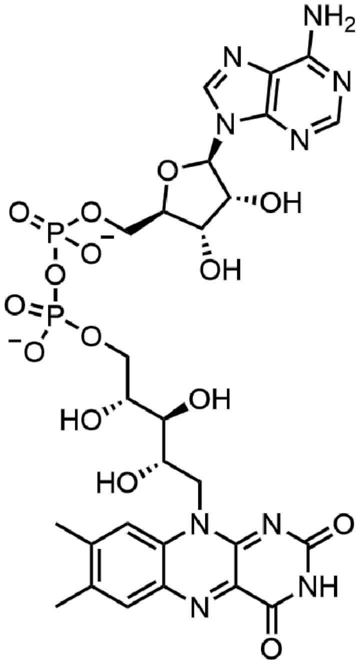
Ribonucleotide coenzymes now used by many proteins may be molecular „fossils” from the primoridal RNA-based metabolism



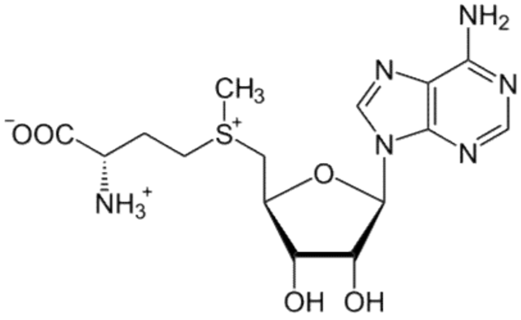
Coenzyme A (**CoA**, **CoASH**, or **HSCoA**)



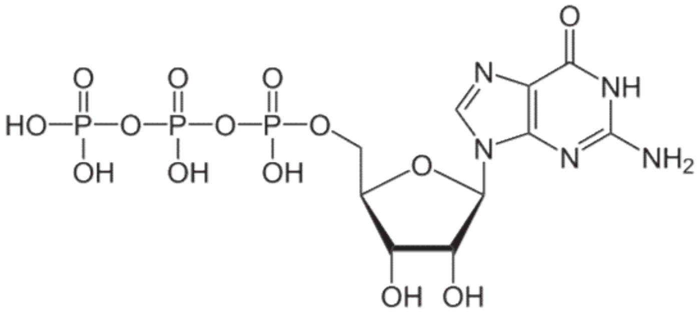
Nicotinamide adenine dinucleotide phosphate (NADP$^+$)



flavin adenine dinucleotide (**FAD**)



***S*-Adenosyl methionine**

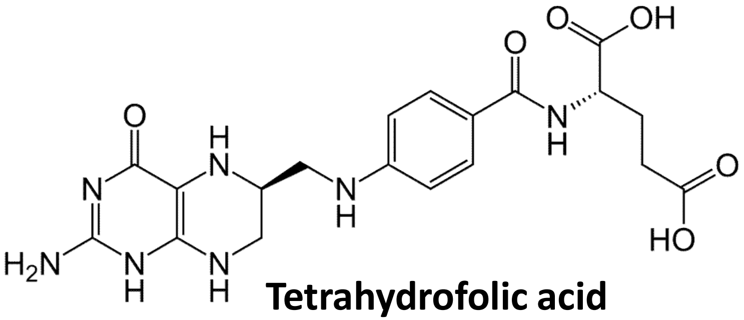


Guanosine-5'-triphosphate (**GTP**)

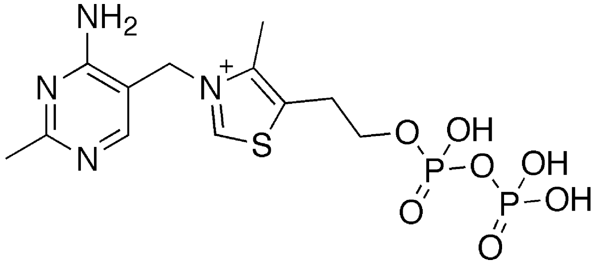
H. B. White III *J. Mol. Evol.* **1976**, *7,* 101-104

The RNA world

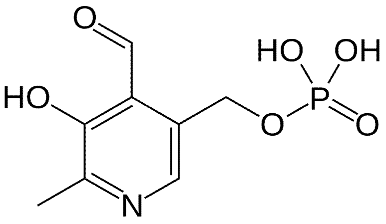
Other coenzymes contain cyclic nitrogen-containing bases that can also derive from nucleotides



Tetrahydrofolic acid



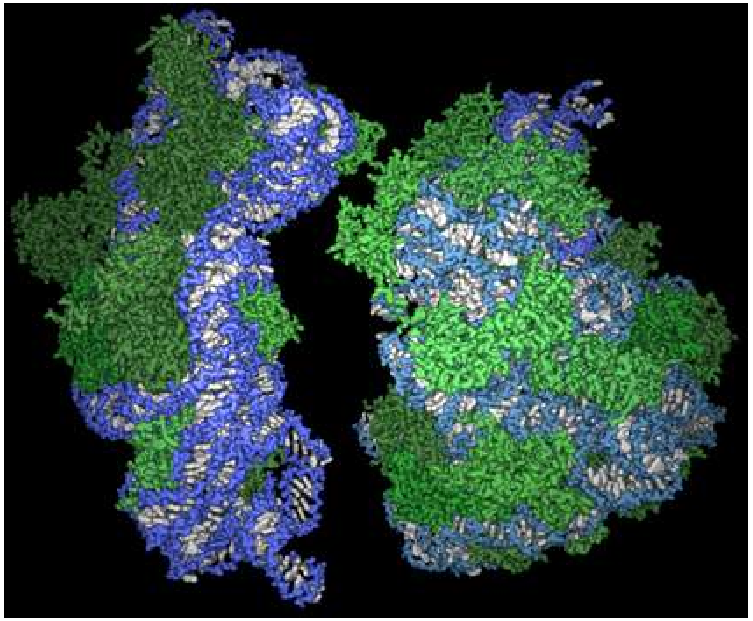
Thiamine pyrophosphate (TPP or ThPP) – Vit. B_1



Pyridoxal phosphate (PLP) – Vit. B_6

H. B. White III *J. Mol. Evol.* **1976**, *7,* 101-104

Ribosome



Ribosome: green - proteins, blue and white - RNA

The ribosome is a molecular machine, found within all living cells, that serves as the site of biological protein synthesis (translation). Ribosomes link amino acids together in the order specified by messenger RNA (mRNA) molecules.

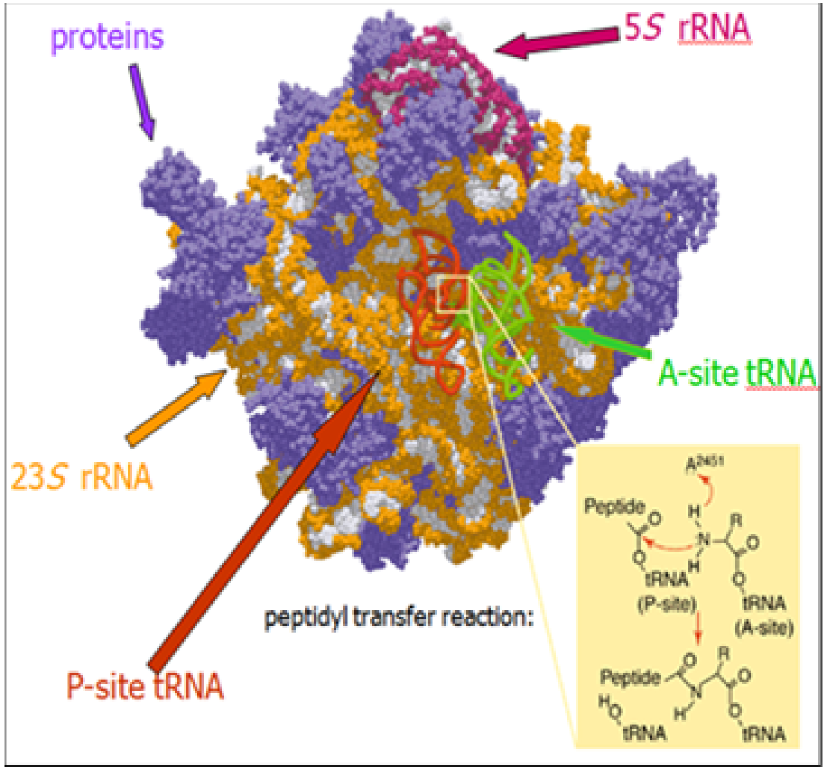
Ribosome is structurally highly conserved among all living species – most likely present in LUCA

Ribosomes:
- the *small ribosomal subunit*, which reads the RNA
- the *large subunit*, which joins amino acids to form a polypeptide chain.

Each subunit is composed of one or more ribosomal RNA (rRNA) molecules and a variety of ribosomal proteins.

Ribosome – the ‚smoking gun'

Ribosome is a ribozyme!

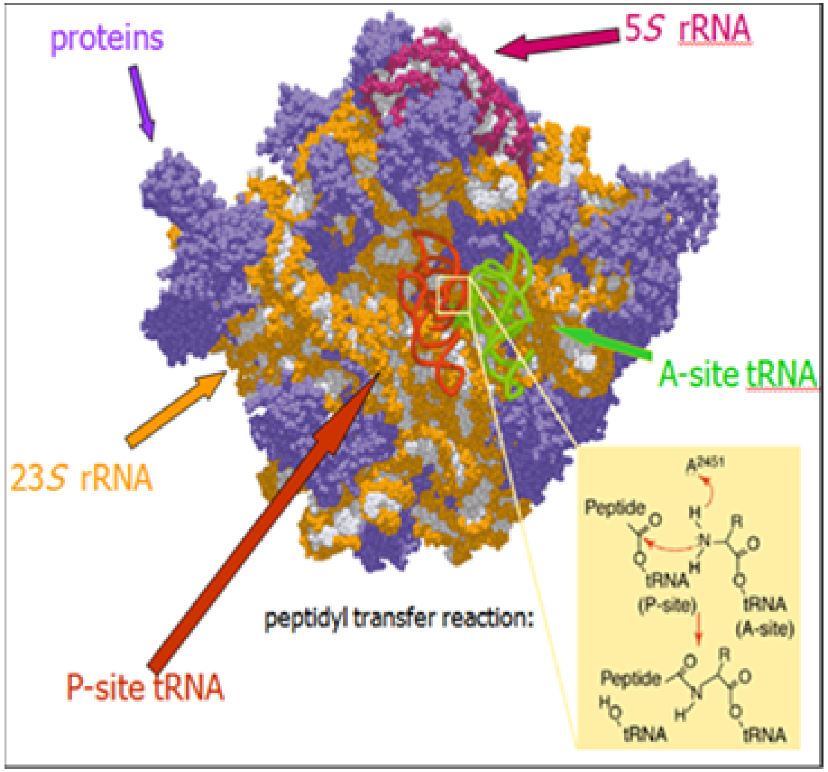


No protein is present within 18 Angstroms from the active site → proteins play a structural role, but DO NOT CATALYZE THE ACYL TRANSFER PROCESS

T. Cech *Science.* **2000**, *289,* 878-879

Ribosome – the ‚smoking gun'

Ribosome is a ribozyme!



The proto-ribosomes in the RNA world – first as a self-replicating complex, later evolved the ability to synthesize proteins with emerging amino acids.

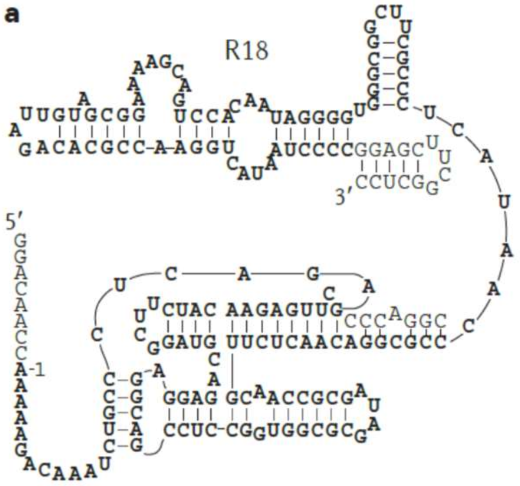
Early proto-ribosomes were self-replicating complexes:
the rRNA had informational, structural, and catalytic purposes –
it coded for tRNAs and proteins needed for ribosomal self-replication.

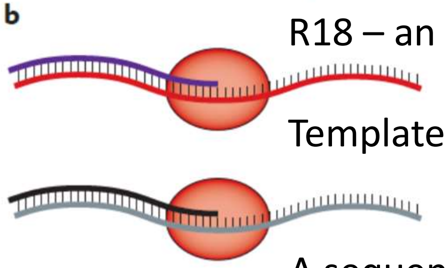
Emerging amino acids interacted with catalytic RNA:
increased scope and efficiency of catalytic RNA molecules.

→ Ability to synthesize peptide bonds was caused by the evolutionary pressure to increase its capacity for self-replication by incorporating proteins into the catalysis

The RNA world

RNA-dependent RNA polymerase ribozyme – Replicase - the ‚holy Grail' of the RNA world

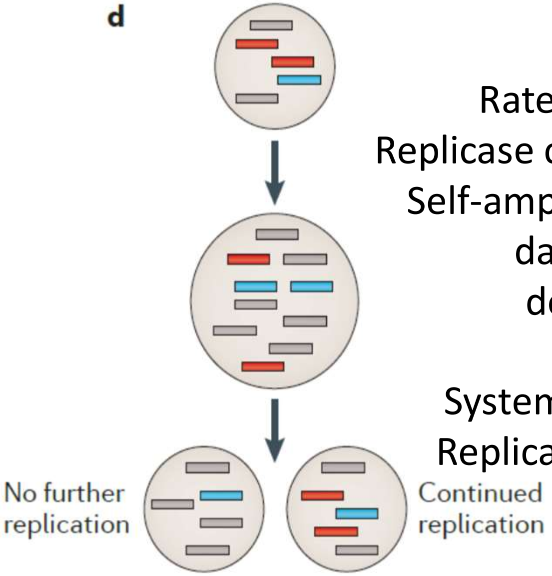




R18 – an artificial polymerase evolved from the class I ligase ribozyme.

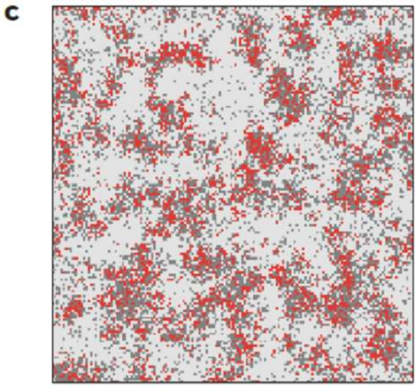
Template: another copy of itself (red) or an unrelated sequence (grey).

A sequence of 206 nt was copied (fidelity 97.4%) at low temperatures by an engineered R18 mutant – first ribozyme capable to synthesize RNA oligomers longer than itself (though NO self-replication yet!)



Rate of replication not sensitive on the template's sequence. Replicase could replicate other ribozymes (e.g. with metabolic functions). Self-amplifying replicase needs a working complementary replicase – danger of paraistes (templates that copy themselves but do not contribute to the replication of the polymerase).

Systems of altruistic replicators are destroyed by parasites (grey). Replicators (red) can survive e.g. by diffusion on 2D surfaces (**c**) or selection inside compartments (**d**)



Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E. & Bartel, D. P. *Science* **2001**, *292*, 1319–1325.

Attwater, J., Wochner, A. & Holliger, P. *Nature Chem.* **2013**, *5*, 1011–1018.