

Introduction to molecular biology: DNA



A/Prof Scott Beatson



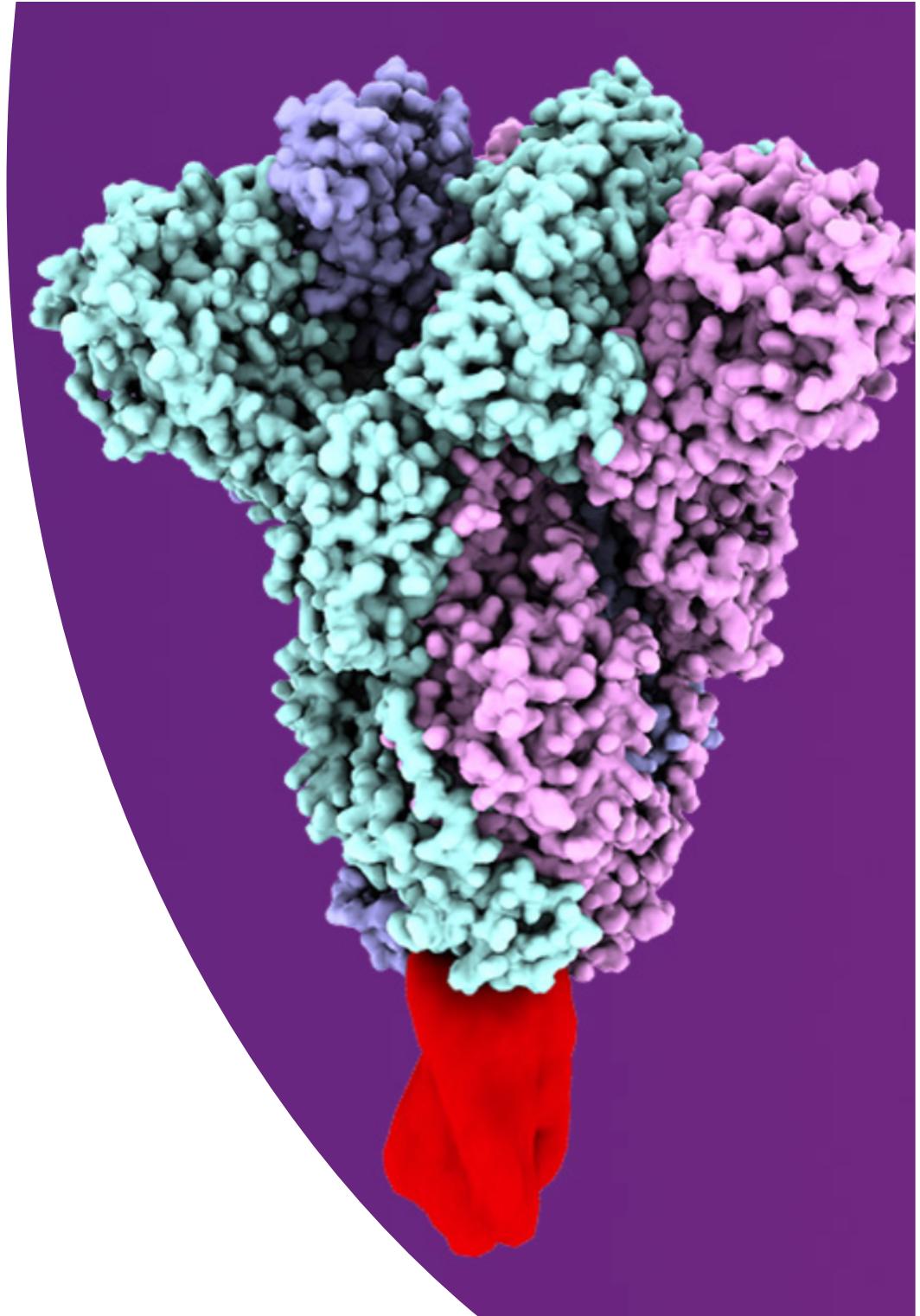
s.beatson@uq.edu.au



BINF6000/SCIE2100

Molecular biology intro

- **Part 1:**
nucleic acids (sequence, structure,
biological role...)
- **Part 2 :**
 - genes and gene expression
- **Part 3:**
 - amino acids and proteins



cell

- basic building blocks of all the living organisms → bio(macro)molecules
→ nucleic acids, proteins, carbohydrates, lipids
- two different cell types based on the structural organization:
 - **prokaryotic** cells (e.g. bacteria, archaea) – NO nucleus and cell organelles!
 - **eukaryotic** cells (e.g. plants, animals) – well defined nucleus and cell organelles (mitochondria, chloroplast...) → more complex system

common features of life forms

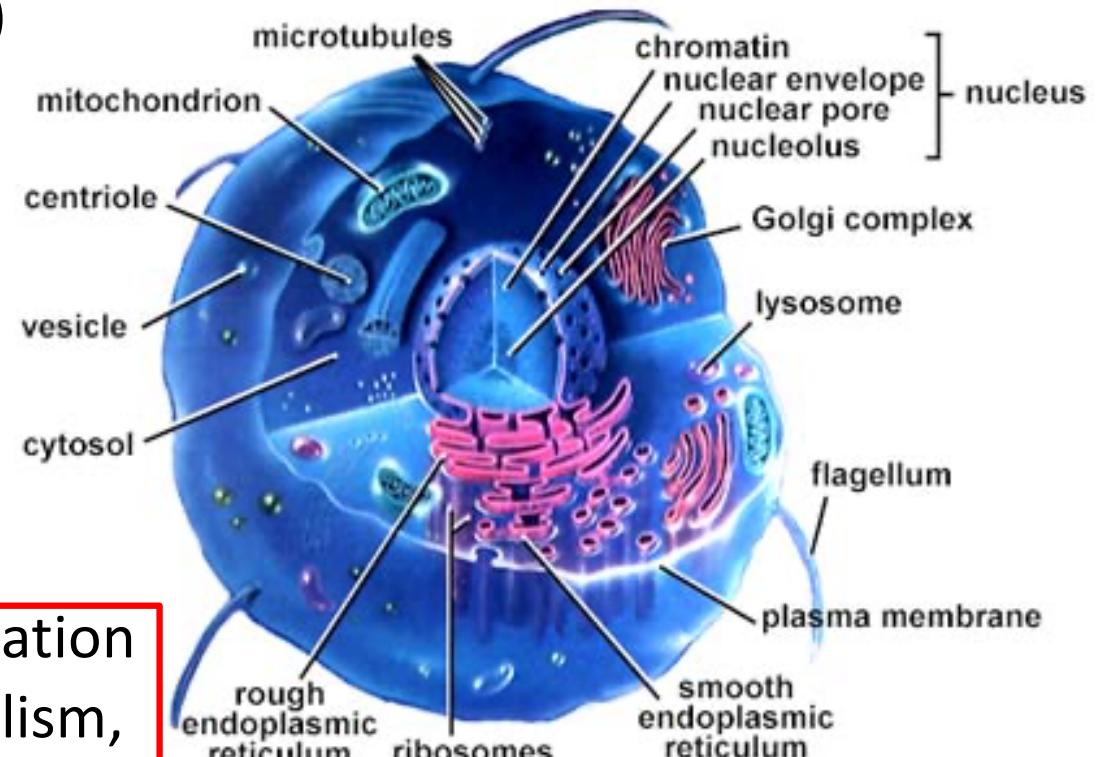
- mostly Carbon, Oxygen, Nitrogen and Hydrogen (and a little bit of sulphur, iron, magnesium, zinc...) combined into basic cellular ingredients:

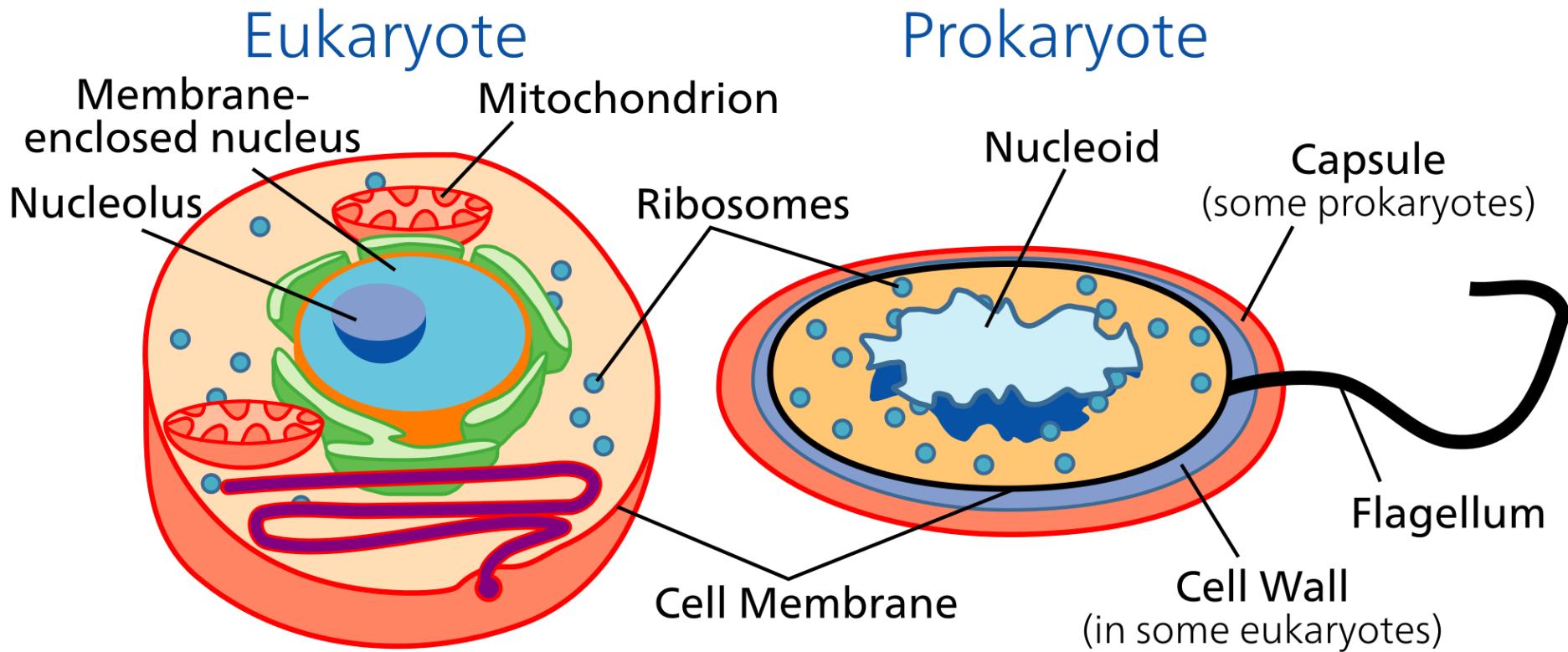
- nucleic acids (DNA, RNA)
- proteins
- carbohydrates (sugars)
- lipids (fatty acids)
- various small molecules

- Different roles for each type of biomolecules:

DNA (RNA) → genetic information
proteins → catalysis, metabolism, transport, structure, motion...

lipids → membrane





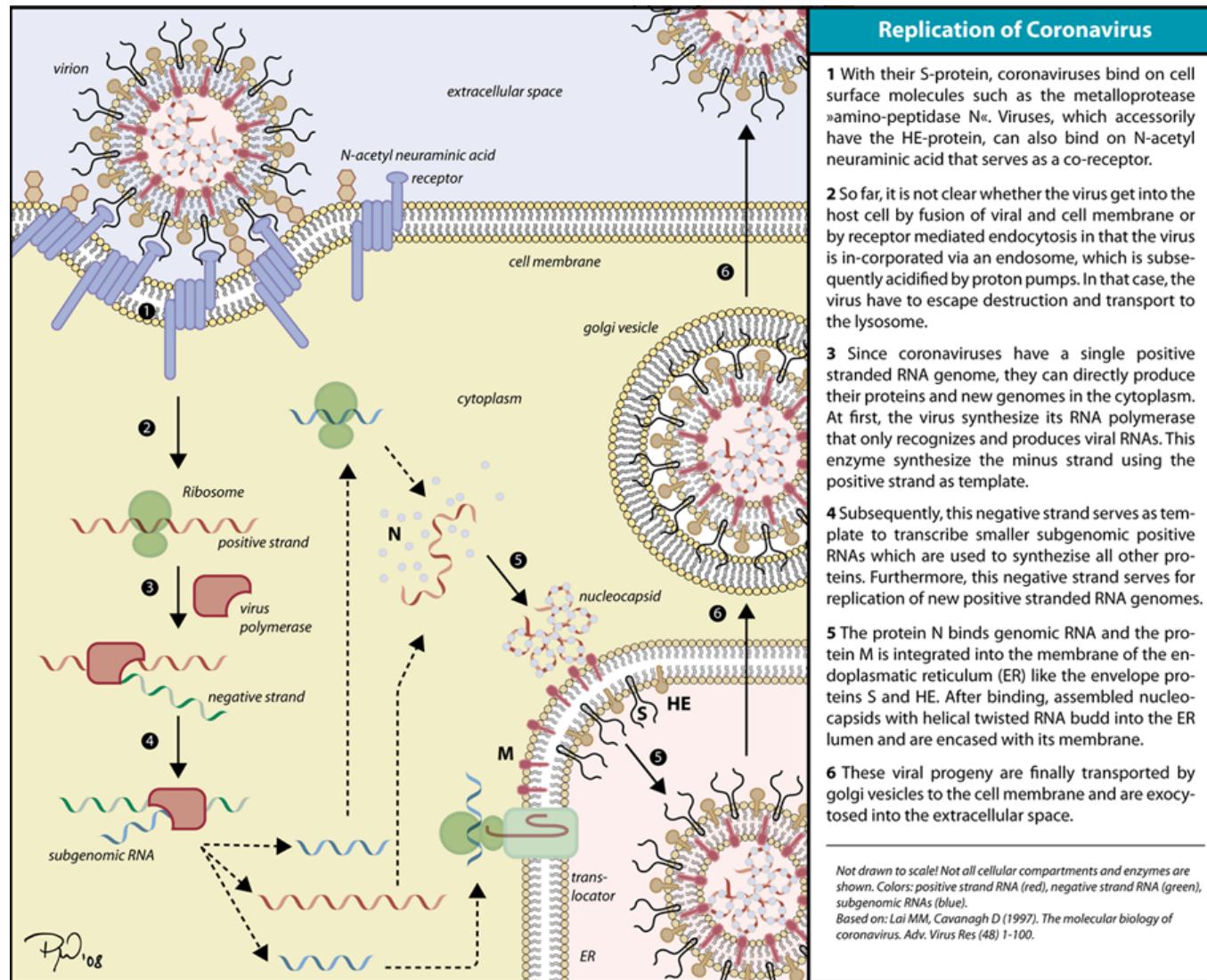
Common features:

- DNA genome
- Protein machines and catalysts
- Ribosomes (protein synthesis)
- ATP as (metabolic) energy source
- Lipid membranes

Distinct features:

- Chromosome structure
- Genome organisation
- Subcellular structures (organelles)
- Metabolic pathways

Coronavirus update: RNA virus



some useful links:

ONLINE TEXTBOOK ON CELL BIOLOGY:

<http://www.nature.com/scitable/ebooks/essentials-of-cell-biology-14749010/contents>

CELL BIOLOGY VIDEOS:

Quick review of eukaryote cell:

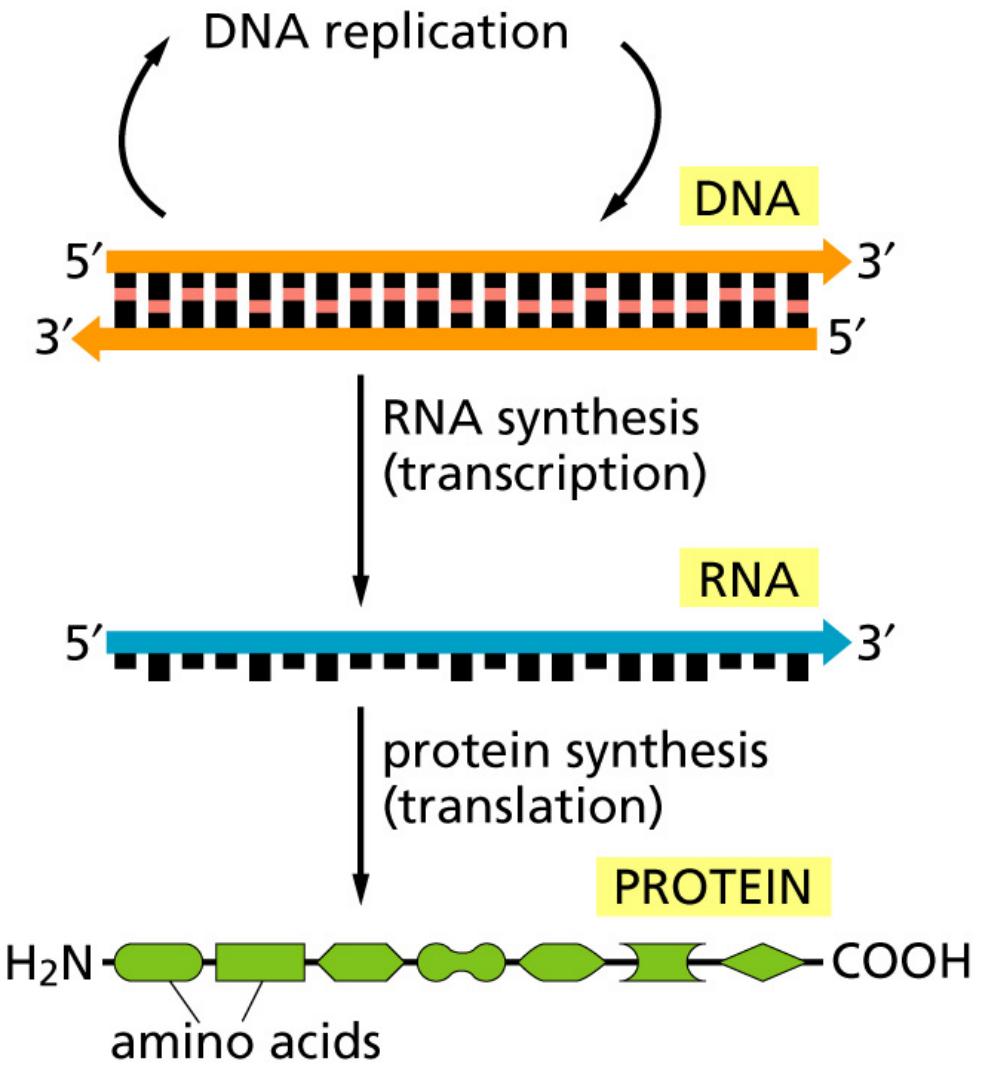
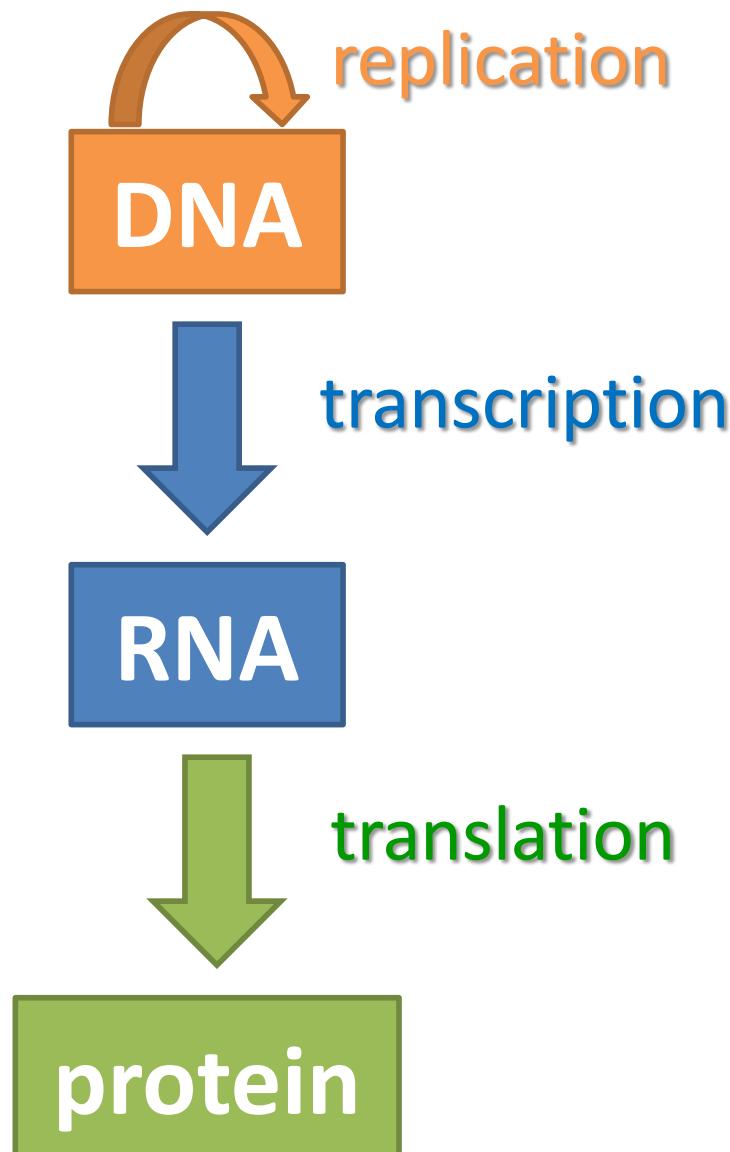
<https://www.youtube.com/watch?v=vwAJ8ByQH2U>

Prokaryote vs Eukaryote cells:

<https://www.youtube.com/watch?v=zZtcMBTQaS4>

central dogma of molecular biology

- link between these molecules:



**“Nothing in
biology makes
sense except in
the light of
evolution.”**

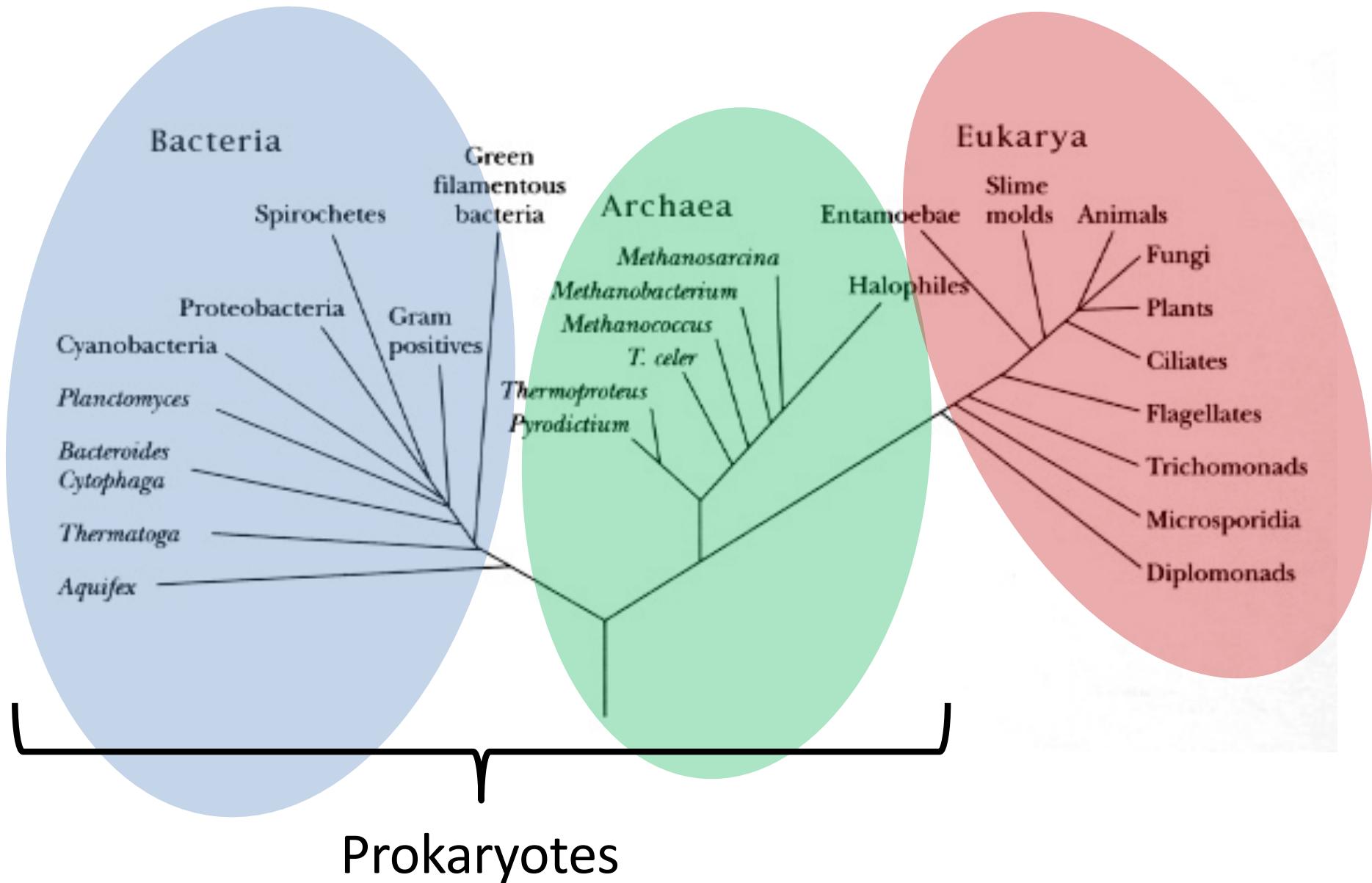
*Theodosius
Dobzhansky*

thelogicofscience.com



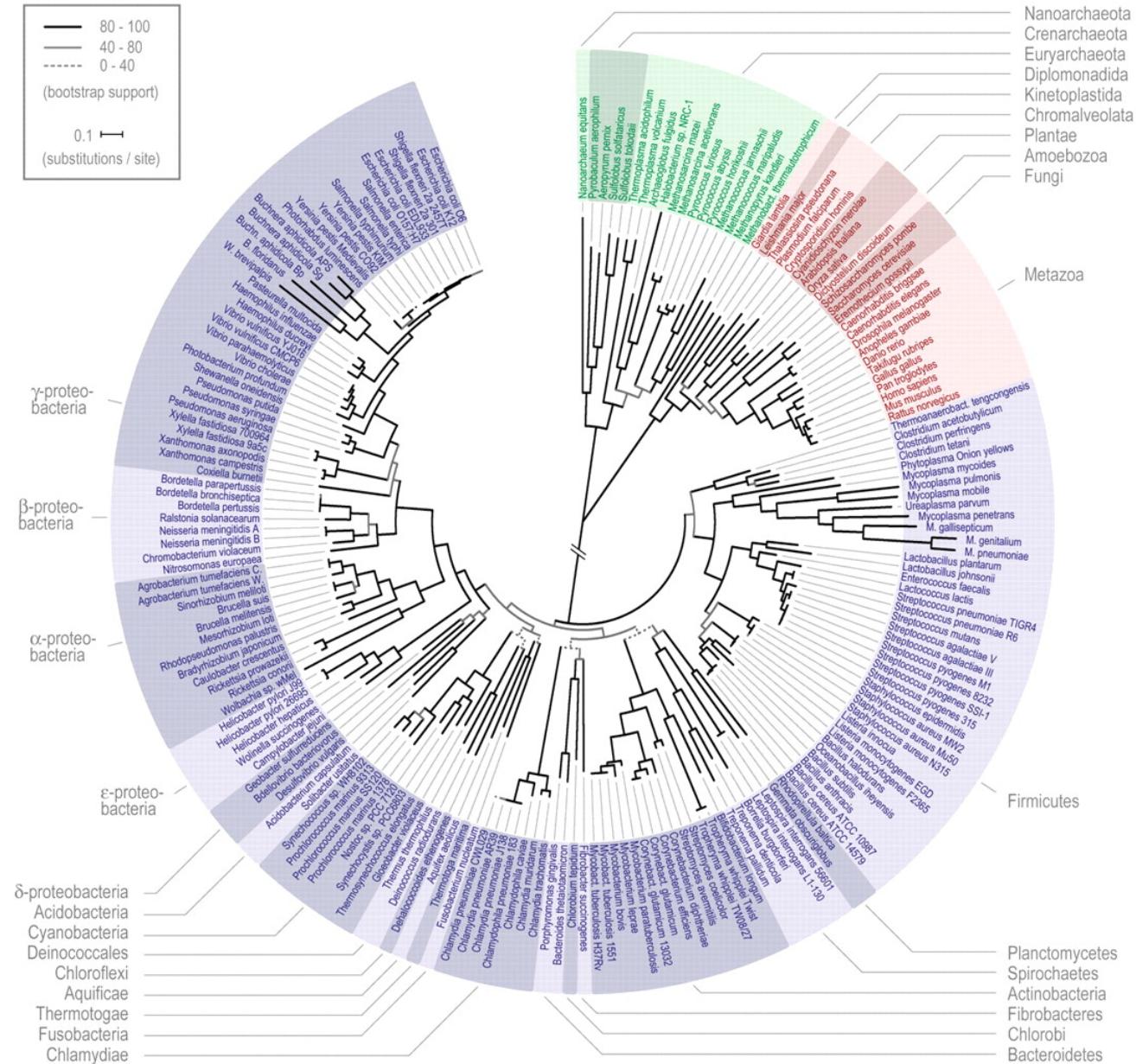
tree of life (3 domains)

domains → kingdoms → phylum → class → order → family → genus → species



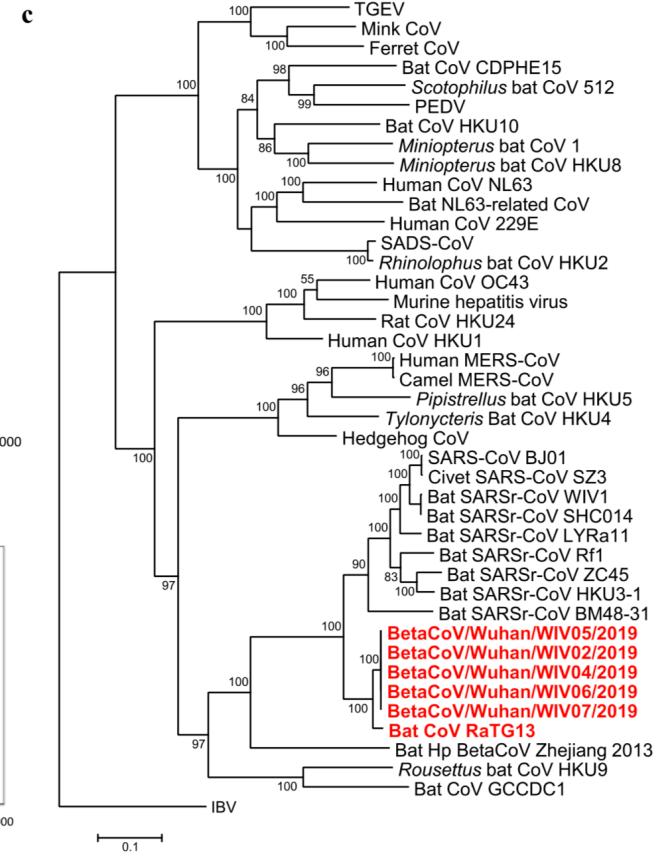
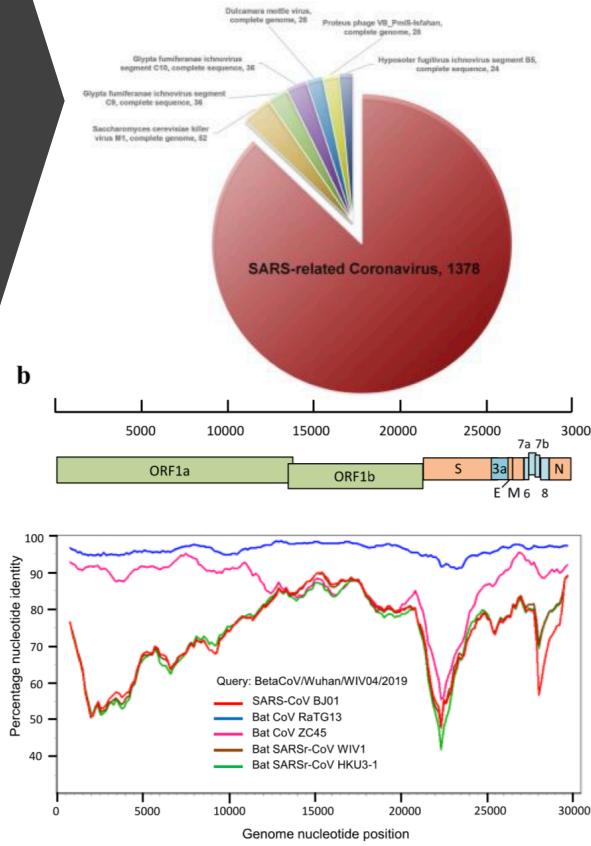
tree of life – high resolution

Ciccarelli et al., (2006). Toward Automatic Reconstruction of a Highly Resolved Tree of Life. *Science*, **311**:1283-1287



Coronavirus update: phylogenetics

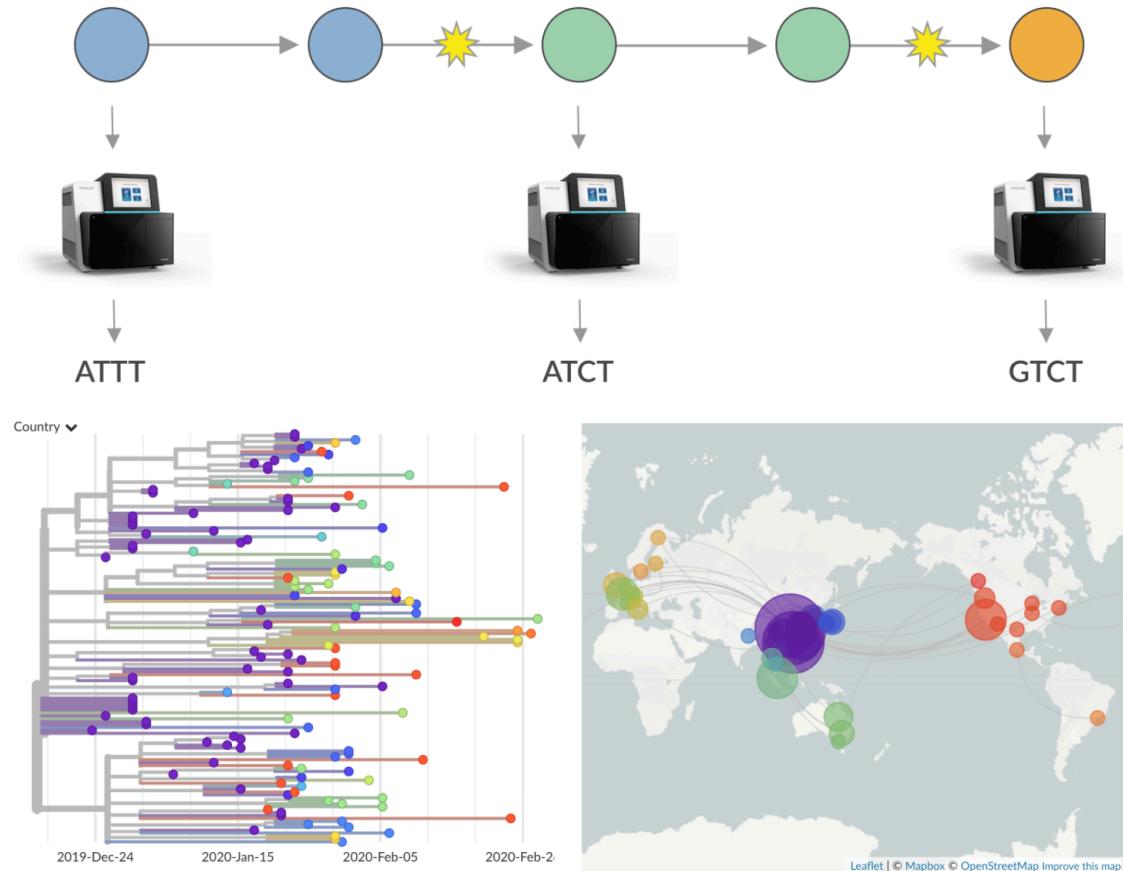
Zhou, P., Yang, XL., Wang, XG. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
<https://doi.org/10.1038/s41586-020-2012-7>



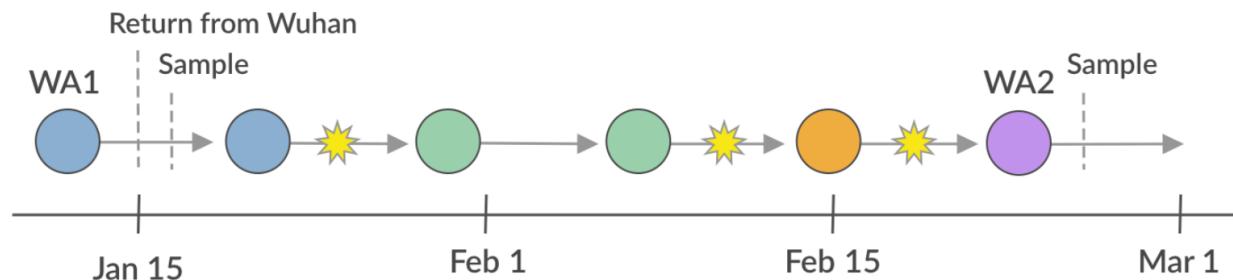
Alpha-CoV

Beta-CoV

Coronavirus update: genome epidemiology

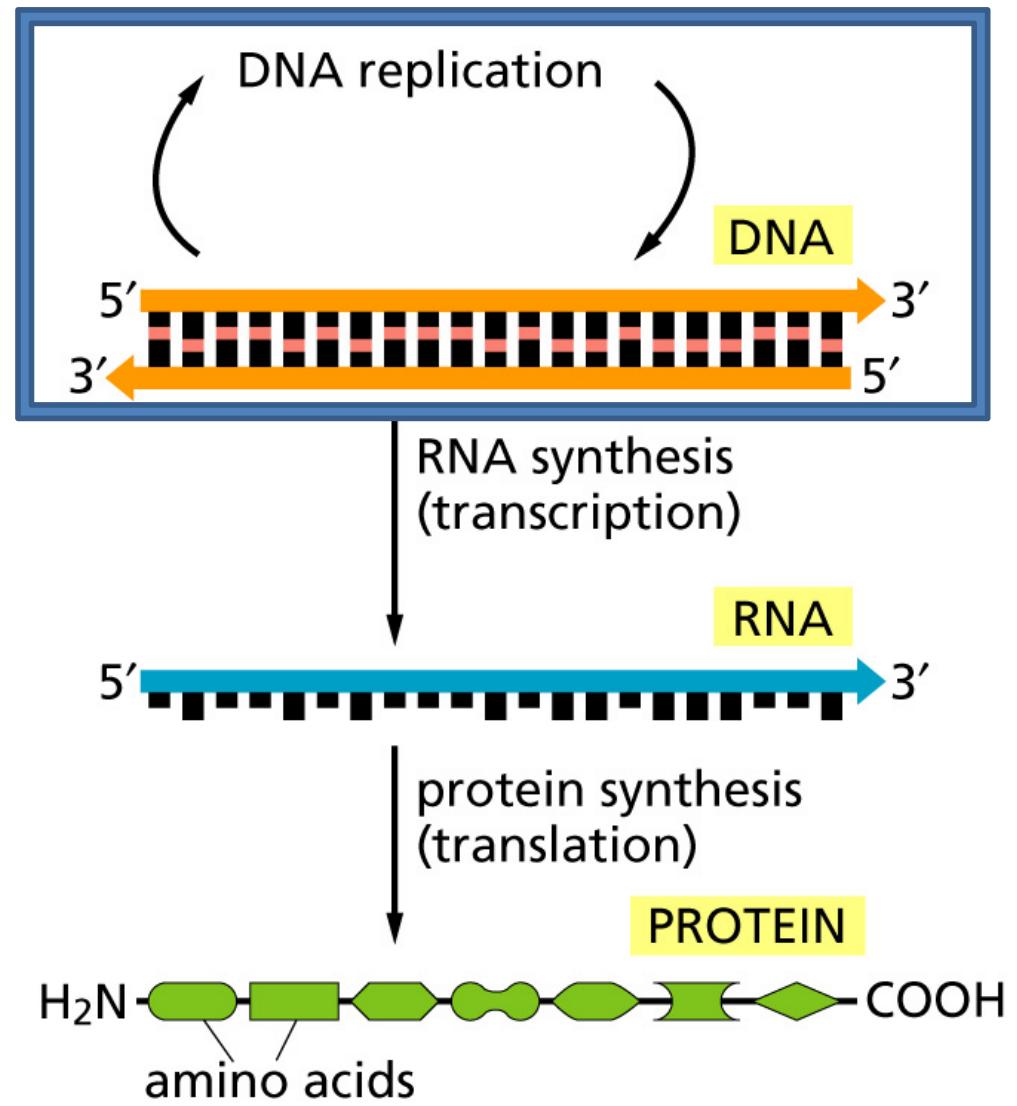
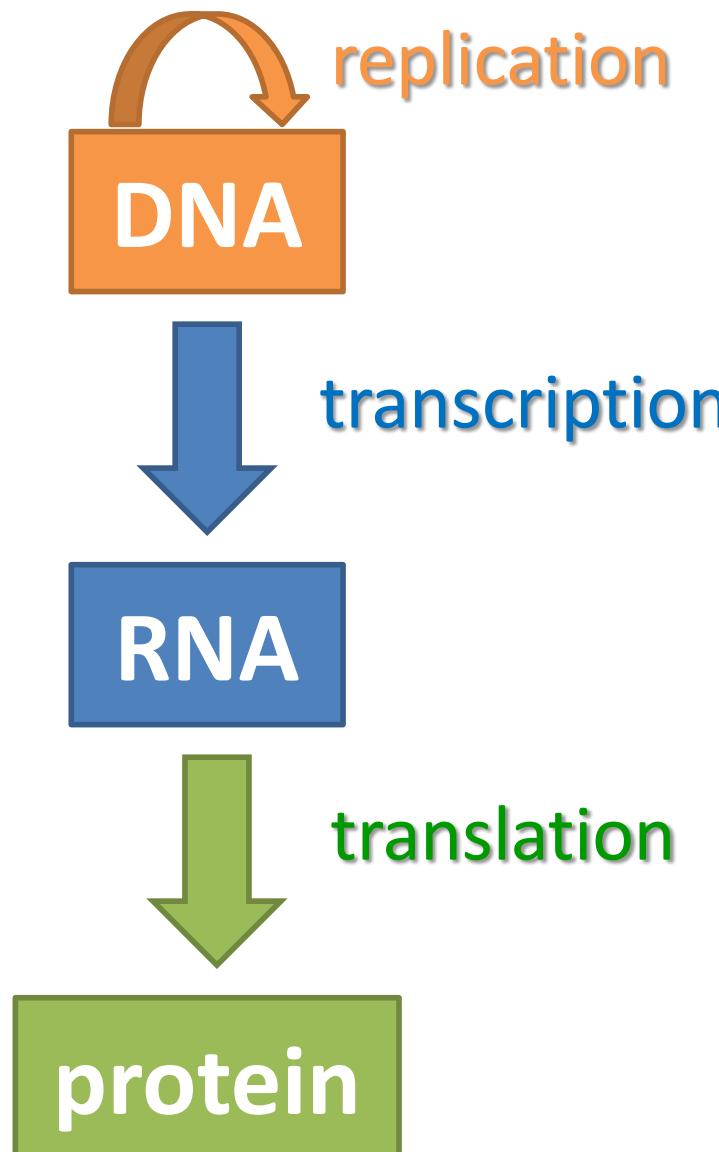


Trevor Bedford blog:
<https://bedford.io/blog/ncov-cryptic-transmission/>



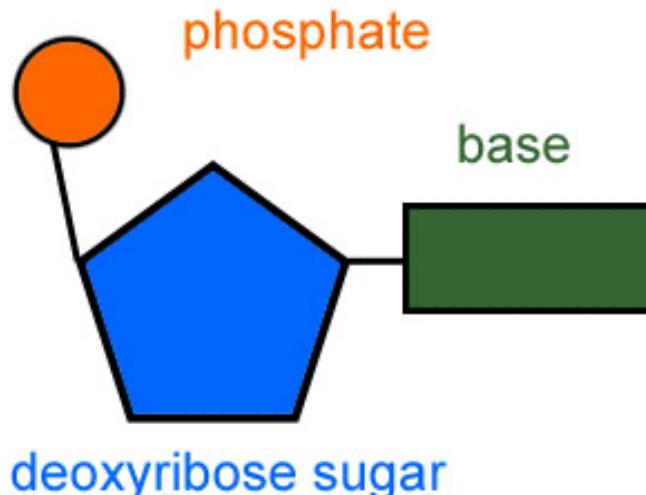
central dogma of molecular biology

- link between these molecules:



nucleic acids

- deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) → biopolymers of 4 different basic units – *nucleotides*

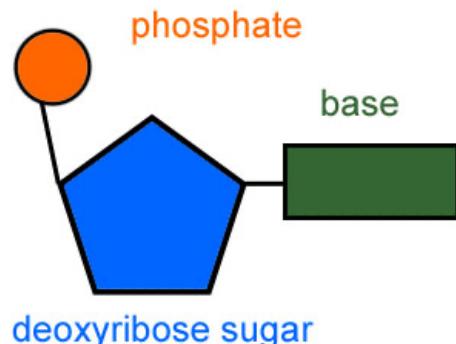


© scienceaid.co.uk

- nucleotide → nucleobase + pentose sugar (deoxyribose/ribose) + phosphate

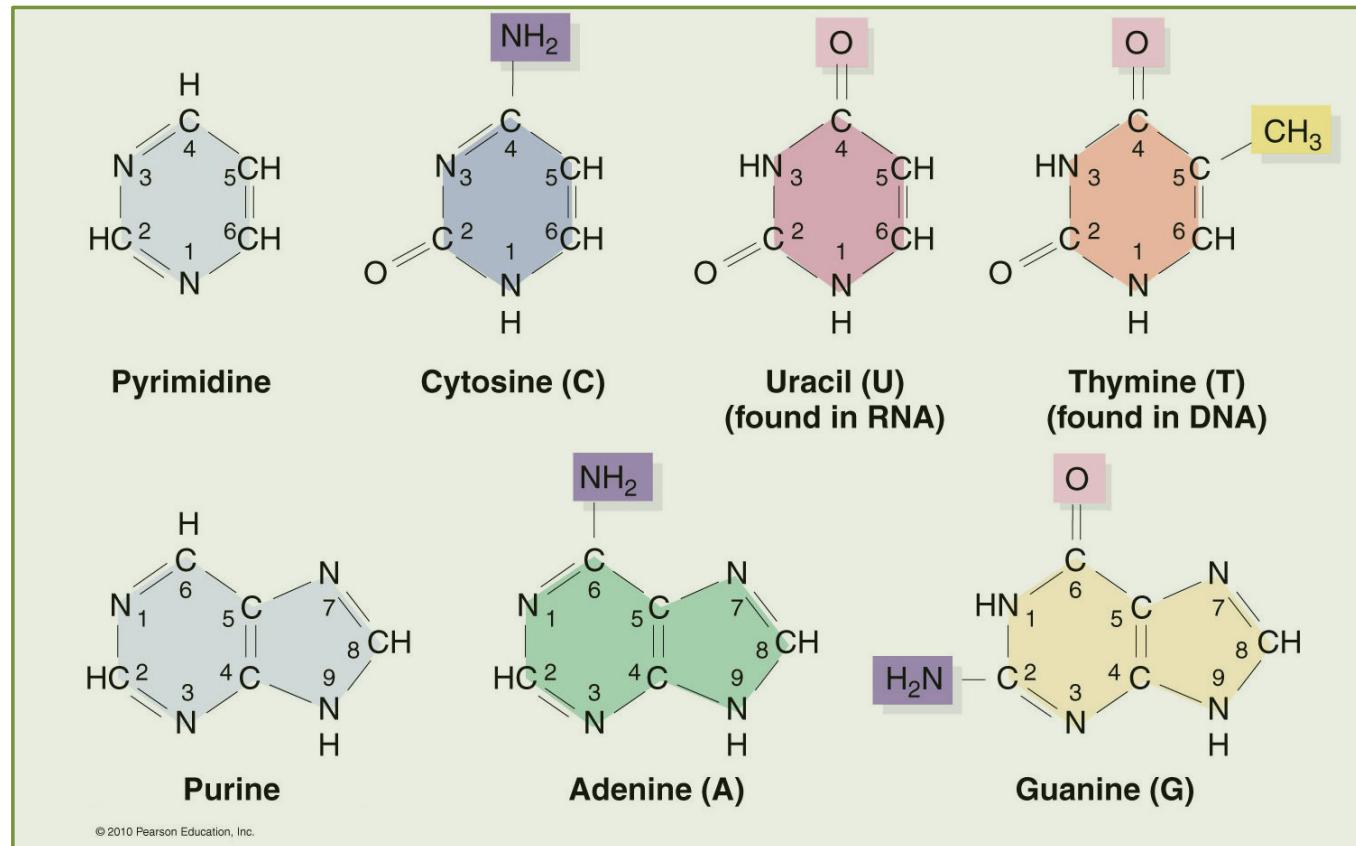
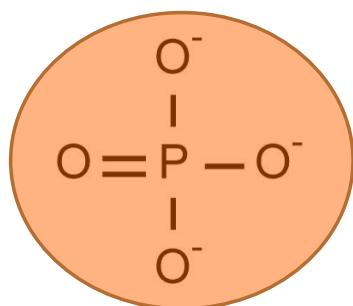
nucleotides

nucleobases

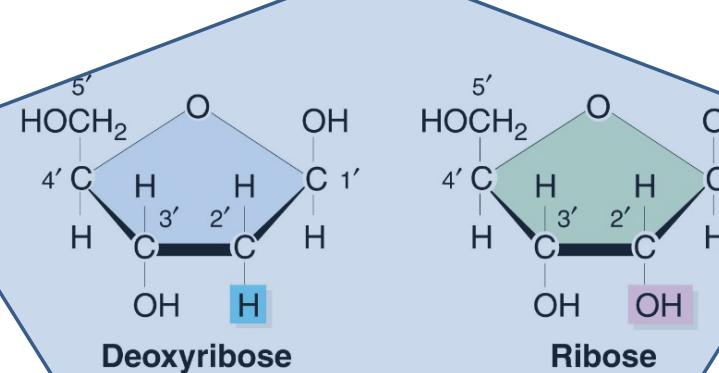


© scienceaid.co.uk

phosphate

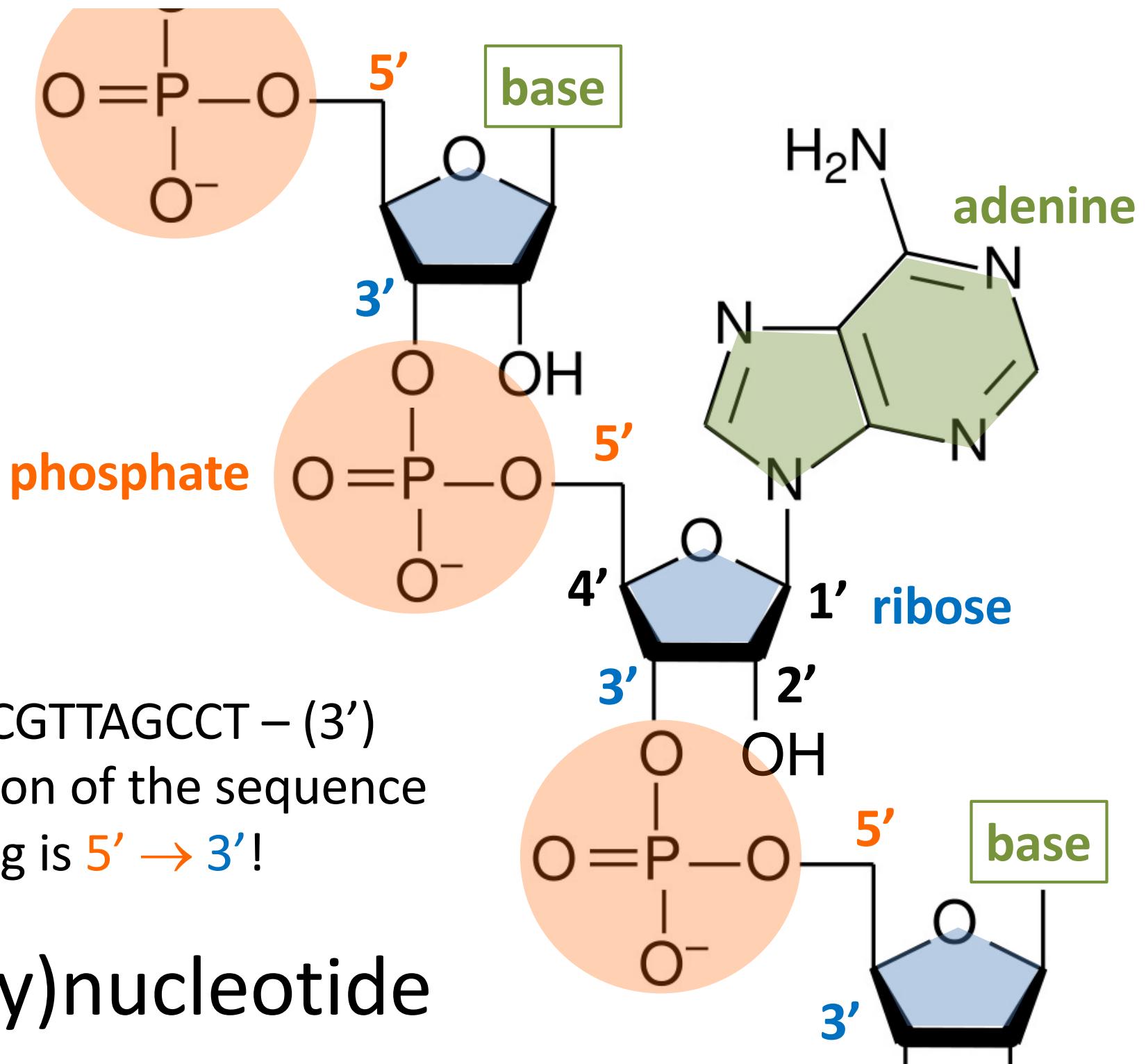


© 2010 Pearson Education, Inc.



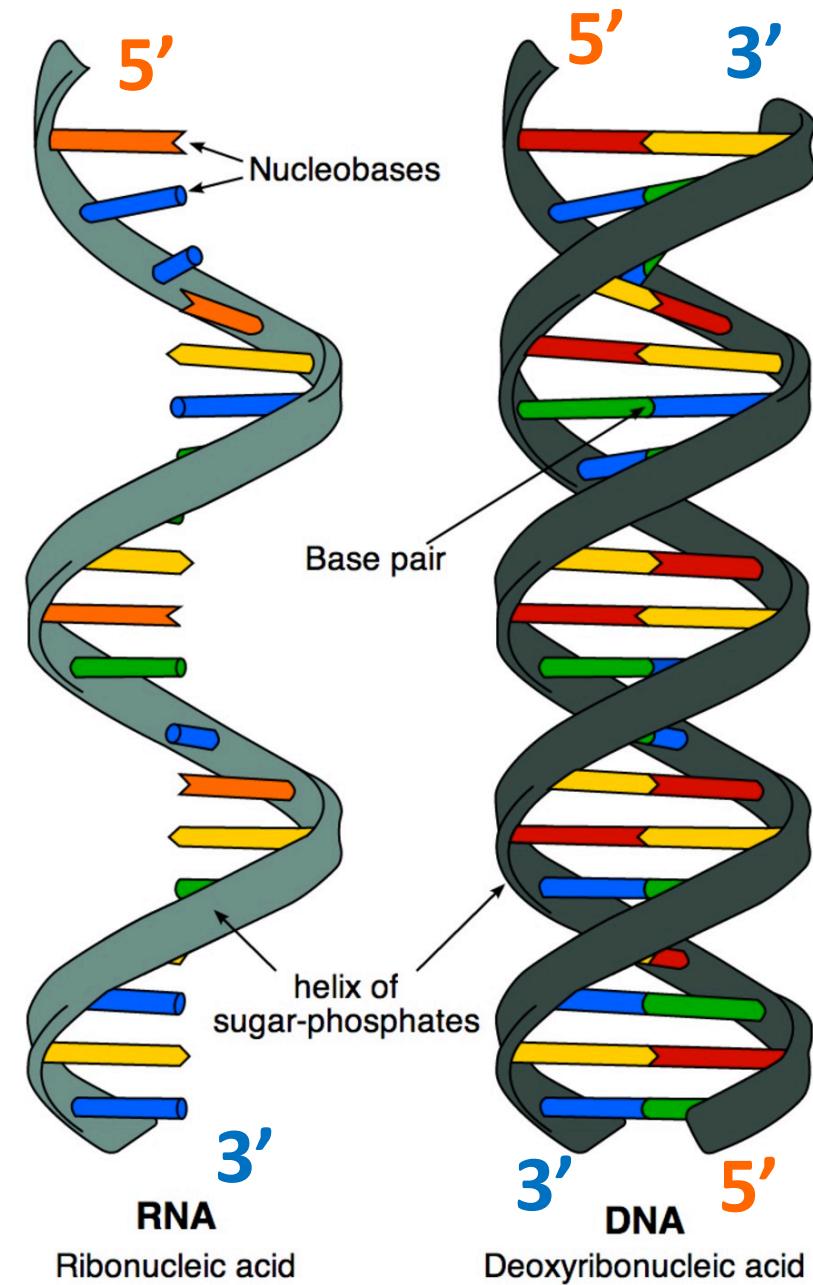
© 2010 Pearson Education, Inc.

pentose sugars



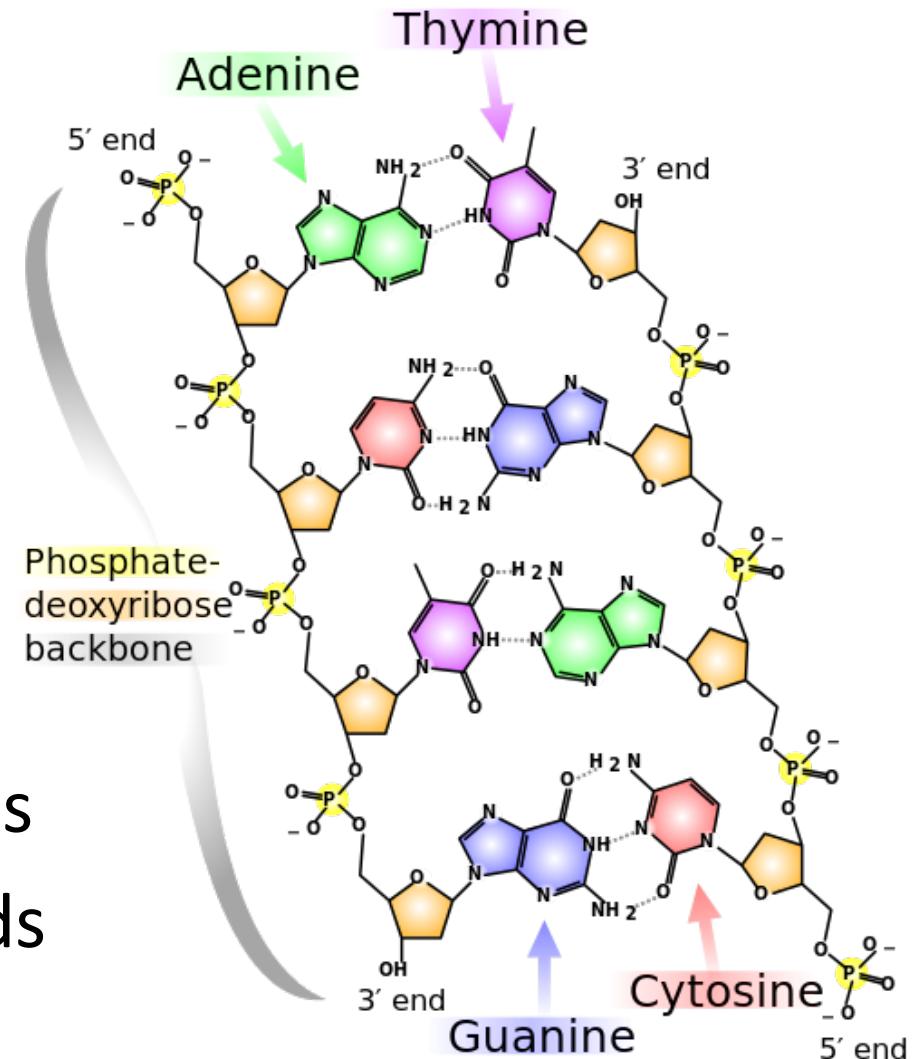
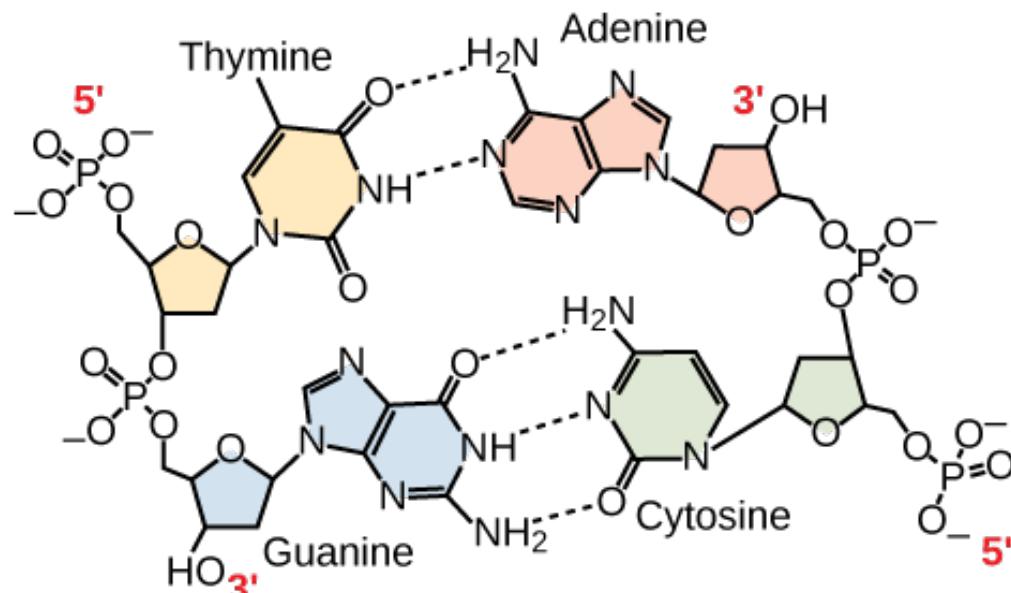
nucleotide polymers (strands)

- RNA is usually found as a single polynucleotide chain (strand)
- DNA forms double helix → 2 strands interacting through base pairing
- in DNA $5' \rightarrow 3'$ strand is called *sense* (coding strand), while the complementary chain is called *antisense* ($3' \rightarrow 5'$)
- only knowledge of sequence of one strand is required to retrieve the information



base pairing

- Watson-Crick base pairing in double helix (1953)



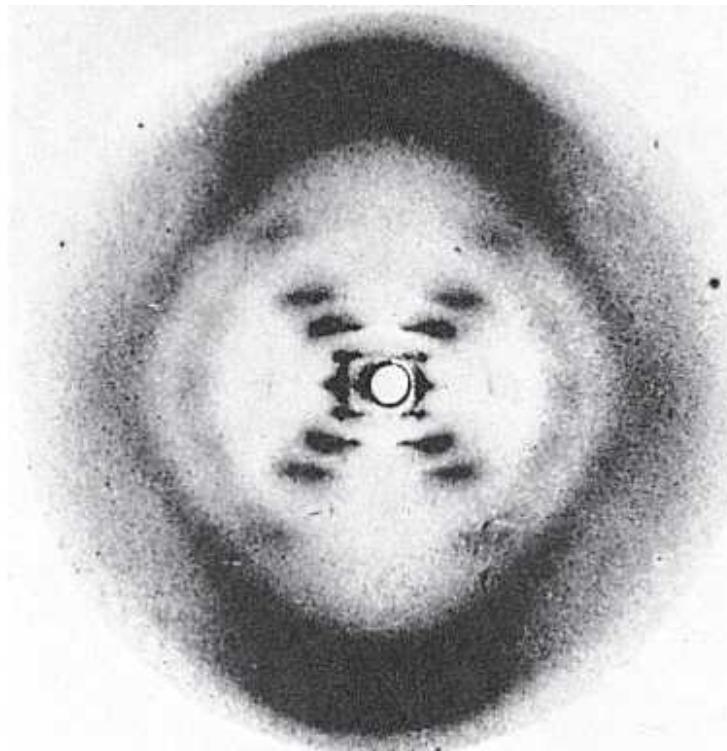
- AT pair → 2 hydrogen bonds
- CG pair → 3 hydrogen bonds

double helix history

- competition & controversy
- James Watson & Francis Crick 1953 → Nobel prize in 1962 together with Maurice Wilkins (experimental verification)



Rosalind Franklin
(1920-1958)



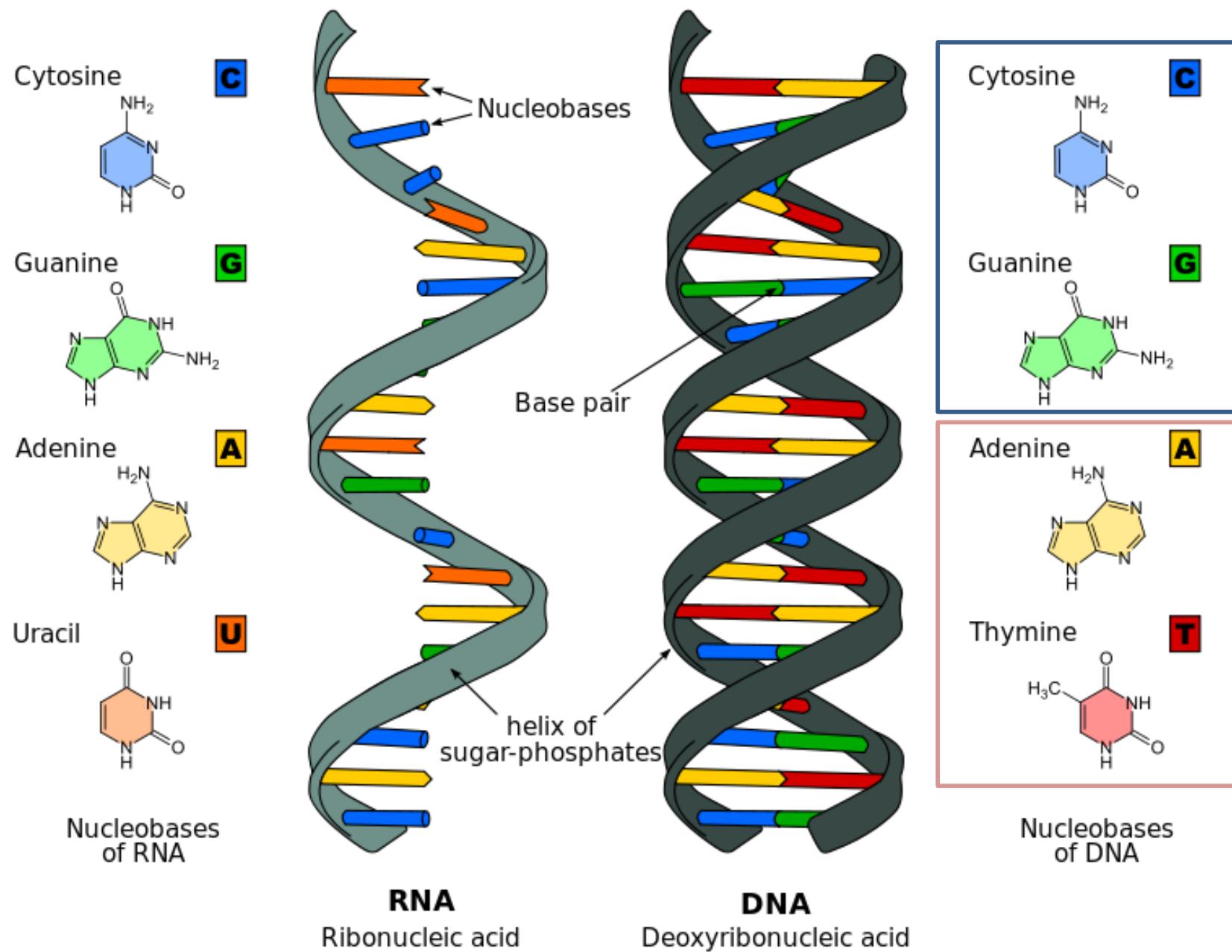
Famous Photo 51 of X-ray diffraction
image of DNA crystals

double helix history

HHMI

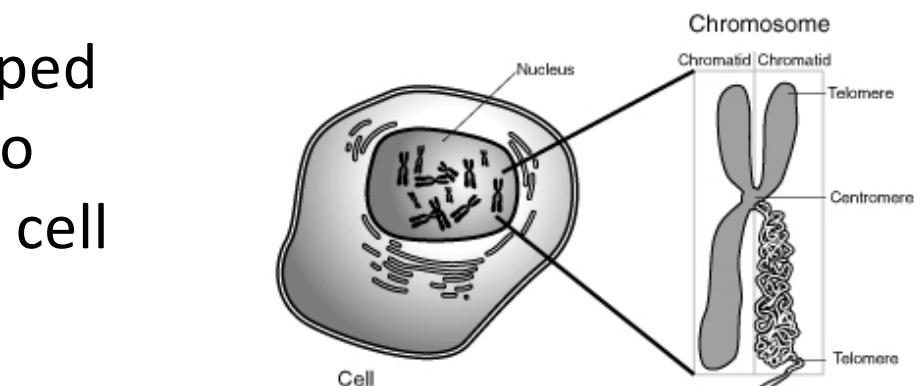
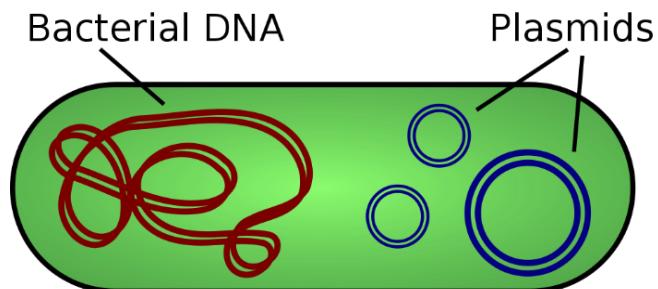
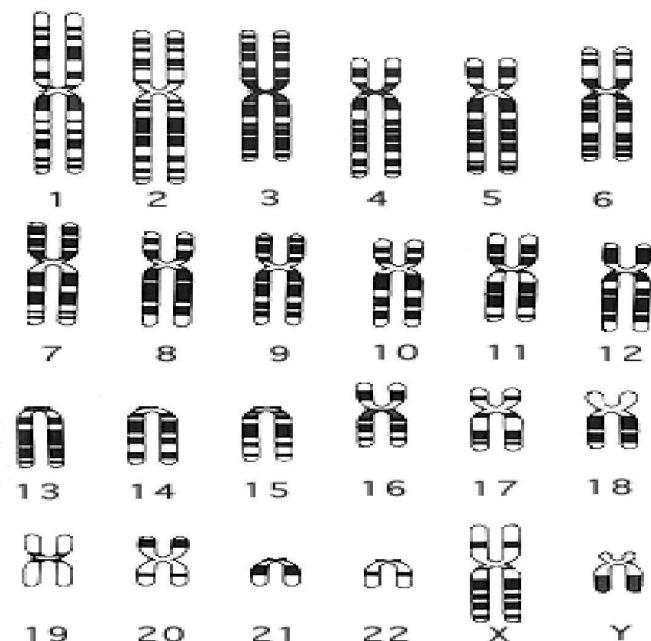
<http://www.hhmi.org/biointeractive/watson-constructing-base-pair-models>

in summary...



DNA in cells

- **prokaryotes** → circular DNA
- **eukaryotes** → linear DNA wrapped around proteins and packed into chromosomes (enclosed within cell nucleus) → ~1.8 m



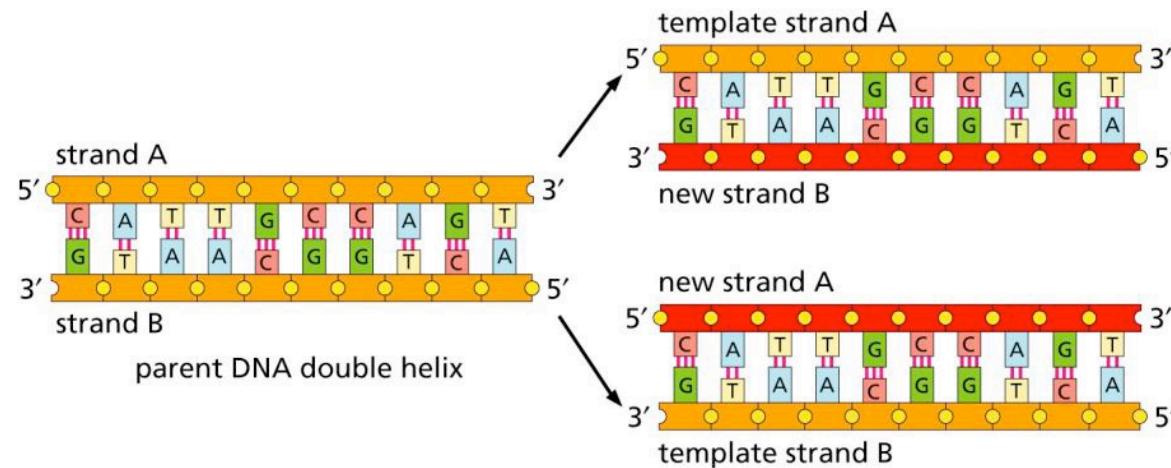
DNA wrapping into chromosomes

wehi.edu.au

Molecular visualizations of **DNA**

1. *DNA Wrapping*

http://www.wehi.edu.au/education/wehitv/molecular_visualisations_of_dna/

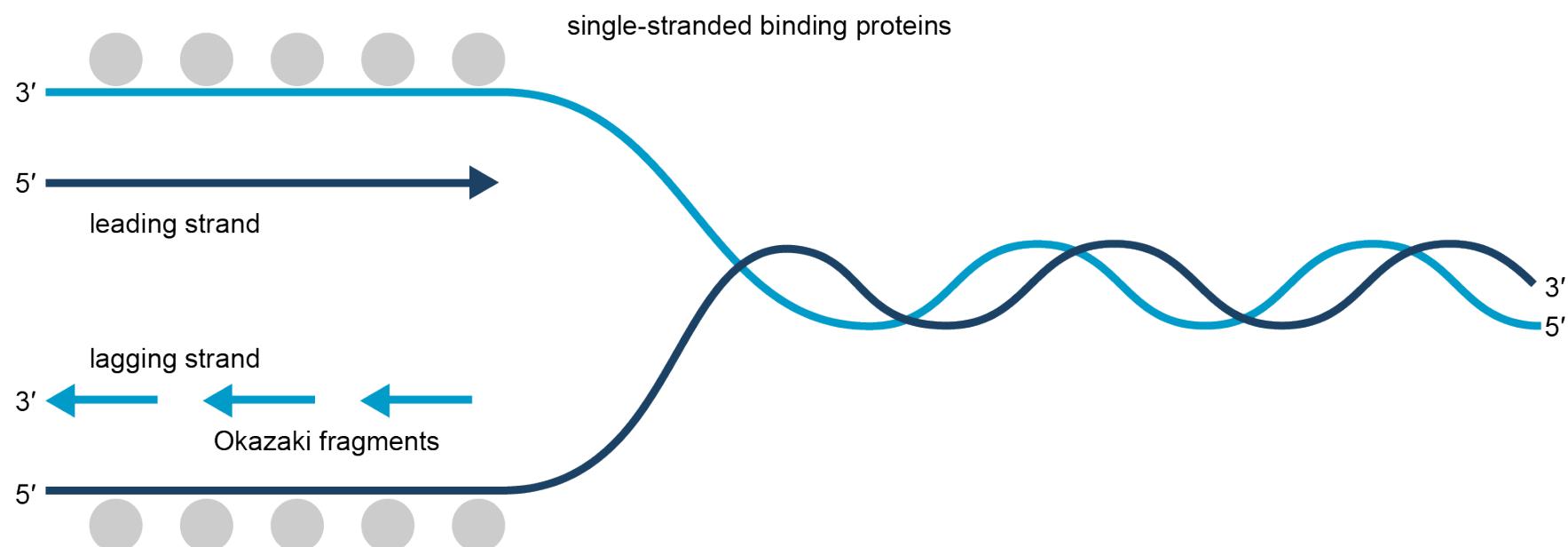


DNA replication

- cell division → duplication of genetic material → key to inheritance
- separation of the strands in double helix (helicase) → each strand serves as a template for a synthesis of the new complementary strand (DNA polymerase) → 2 new DNA chains
- proofreading → accuracy within 1 wrong base per billion!!

Basic steps of DNA replication

- initiation → unwinding of helix in AT rich areas – weaker H bonds
- elongation → DNA is always synthesized in the 5' to 3' direction (template is read in the 3' to 5' direction) → *leading* (5' → 3') and *lagging* strand (3' → 5')
- Okazaki fragments necessary as DNA always synthesized 5' to 3'



DNA replication

wehi.edu.au

Molecular visualizations of **DNA**

2. DNA Replication

http://www.wehi.edu.au/education/wehitv/molecular_visualisations_of_dna/

Key learning outcomes

- Appreciate the differences between prokaryote and eukaryote cells and genomes
- Consider the role of evolution when investigating genomes and genes
- Have a broad understanding of the central dogma of molecular biology
- Understand how the structure of DNA and RNA gives rise to the genetic code and the central dogma
- Recognise the process of DNA replication and packaging

Refer to Week 1-2 Study Guide document

Introduction to molecular biology: Gene structure and control



A/Prof Scott Beatson



s.beatson@uq.edu.au



BIOC6000/SCIE2100

DNA as a blueprint

- *genetic* information is stored in the sequence of nucleotides in each DNA molecule (ATCG)
→ single DNA molecule can contain millions of bases → each cell can have multiple DNA molecules
- DNA content of a cell is called *genome* (usually nuclear DNA is implied)
- in eukaryotic cells, DNA is also stored in mitochondria and chloroplasts (organelles)
- genome size ranges from a few million nucleotides in bacteria to billions in multicellular organisms
- human genome contains 3.2 billion bases divided between 23 pairs of DNA molecules of different length

genomes

- billions of nucleotides → how is this information flows to proteins?
- genes → DNA segments transcribed into RNA → *gene expression*
- genome consists of coding (*genes*) and non-coding parts (also known for a very long time as “junk” DNA)
- more of “junk” DNA found in complex organisms
- human genome contains 3.2 billion nucleotides, but only 5% is protein-coding
- great similarity between people and mice protein-coding regions (85%) → more differences in non-coding areas (up to 50%) → even closer resemblance to chimpanzees and bonobo monkeys (99%)

<https://genome.ucsc.edu/ENCODE/>

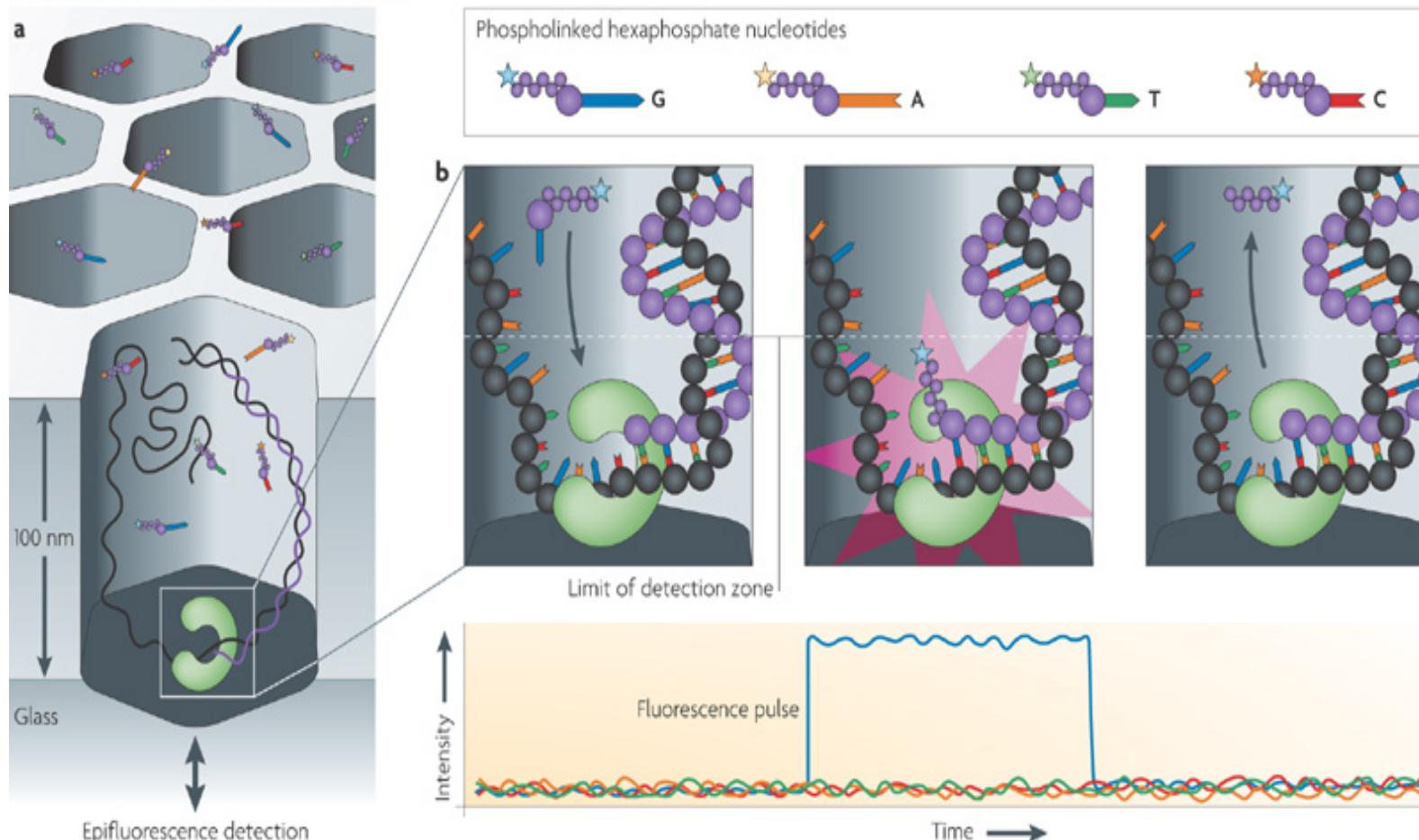
<https://www.genome.gov/>

DNA modifications

- damage as a result of hostile environment (UV radiation, aggressive chemicals, etc...)
- changes in the sequence → mutations, insertions, deletions of nucleotides
- cells have elaborate replication and repair machinery to preserve DNA
- over time, changes in the DNA bases accumulate (slowly, as in the case of evolution, or rapidly, as in the case of cancer)
- different mechanisms of changing DNA and increasing the versatility of sequence

RESEARCH UPDATE: determining the “methylome” with SMRT* sequencing

Pacific Biosciences — Real-time sequencing



Nature Reviews | Genetics

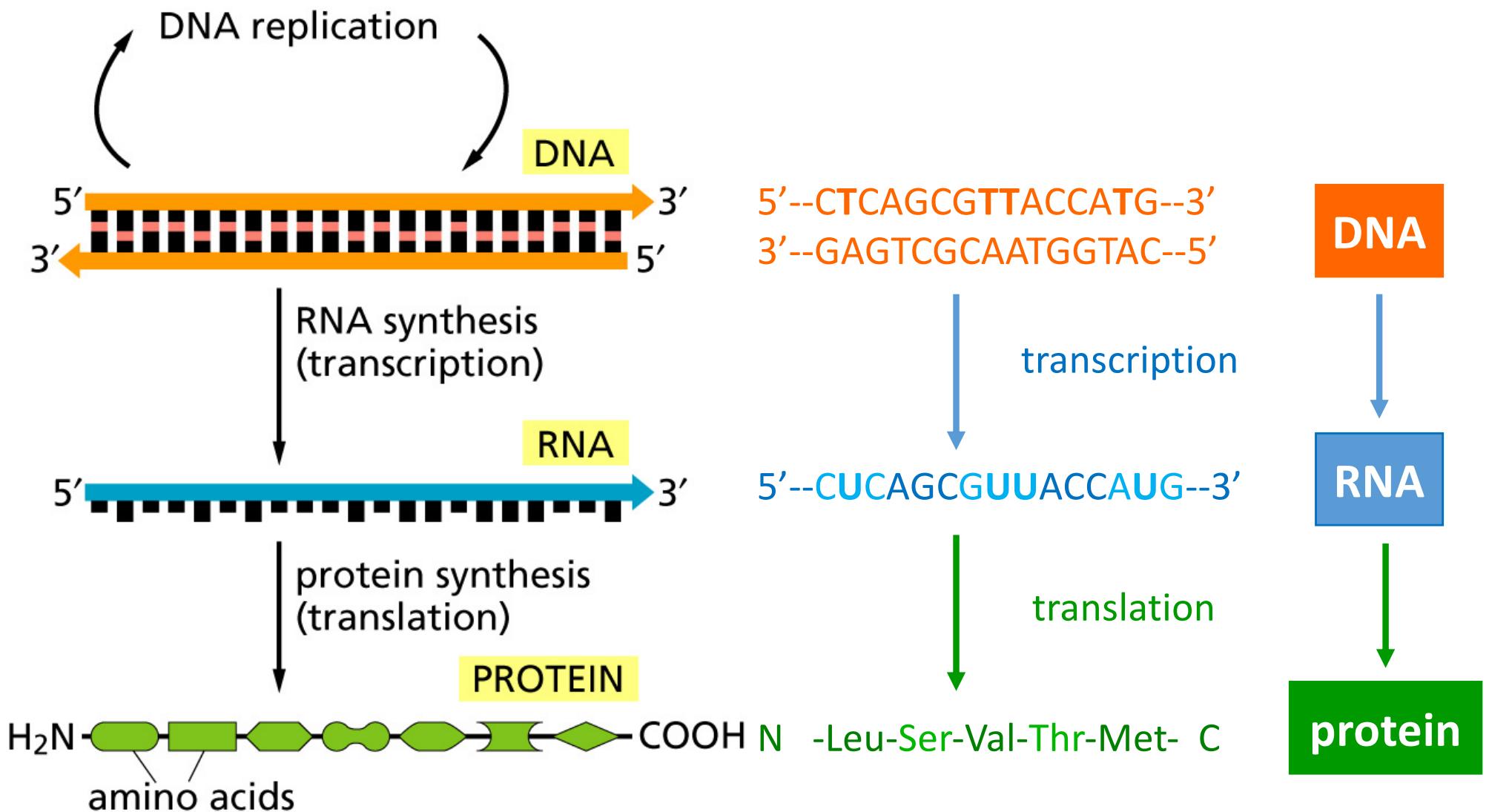
Metzker, *Nature Reviews Genetics* 11, 31-46 (2010), doi:10.1038/nrg2626

(*Single Molecule Realtime Sequencing with Pacific Biosciences “PacBio”)

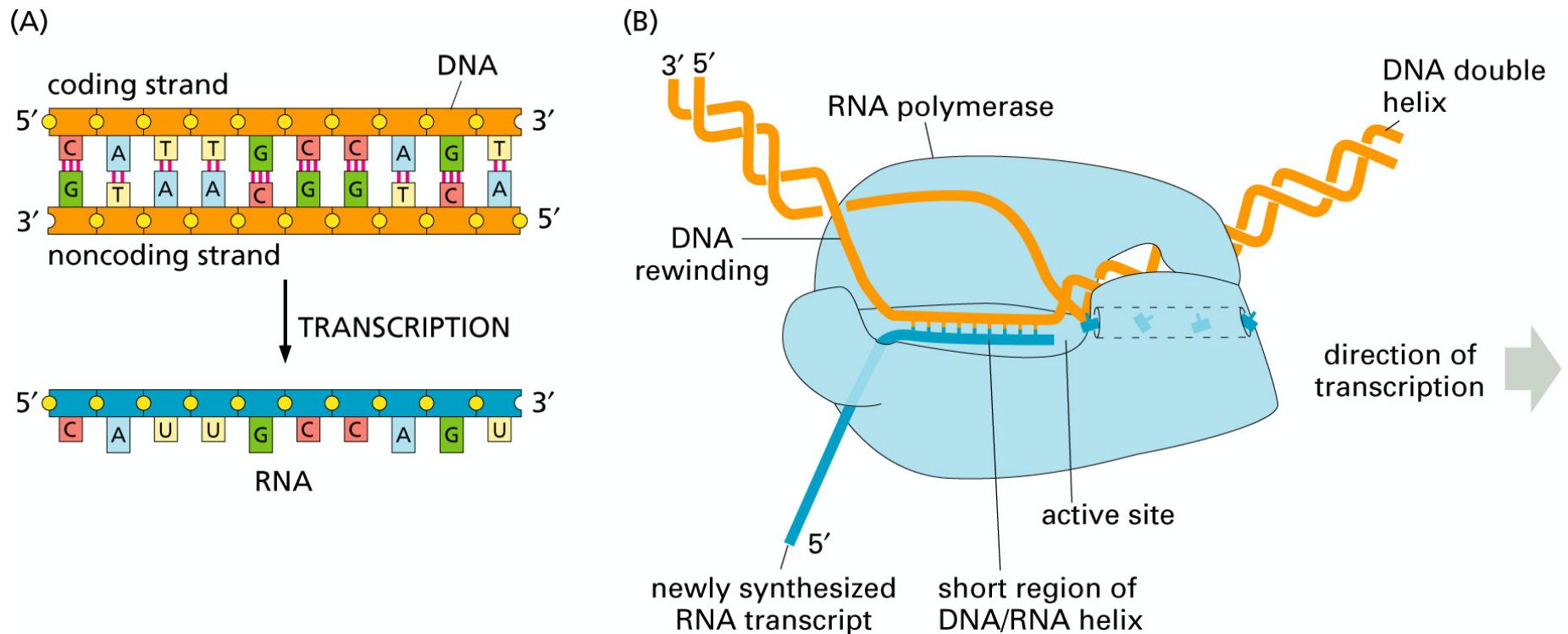
DNA is the information store

- DNA carries the blueprints how to build and maintain the living cell → directions for protein synthesis
- proteins are key ingredients of every cell playing important role in various aspects of cell functioning (*structure, transport, metabolism, catalysis, signalling, etc.*)
- translation of DNA to proteins follows the same basic scheme in ALL the cells → *central dogma of molecular biology*

central dogma of molecular biology



gene expression: transcription of genes to messenger RNA (mRNA)



- transcription → noncoding strand of DNA serves as a template for synthesis of so-called messenger RNA (mRNA) → process catalysed by RNA polymerase
- resulting mRNA has the same sequence as the coding DNA strand, except that thymine (T) is replaced with uracil (U)

Transcription video

<https://www.youtube.com/watch?v=5MfSYnItYvg>



what is a gene?

- definition of gene changed over the years as the knowledge increases:
 - an abstract concept where gene was defined as a unit of inheritance that ferried a characteristic from parent to child
 - development of biochemistry led to one gene – one protein relationship
 - advances in molecular biology turned genes into real, physical things – DNA sequences which can be converted into RNA, which in turn can lead to protein synthesis
 - a gene is a **heritable** string of nucleotides that can be transcribed, creating a molecule with biological activity
- protein is not necessarily the end product of gene transcriptions → sometimes it is RNA (i.e. “non-coding RNA” plays various roles in the cell)

<http://www.nature.com/nature/journal/v441/n7092/full/441398a.html>

<http://www.sciencemag.org/content/316/5831/1556.full>

<http://scienceblogs.com/digitalbio/2007/01/21/what-is-a-gene-my-definition-i>

overlapping genes

- overlapping genes → compact information packing → occurrence in viral and prokaryotic genomes, mitochondrial DNA, but surprisingly large number found in mammalian genomes



Veeramachaneni et al. Genome Res. 2004;14:280-286

- overlap between three human genes: MUTHYH, FLJ13949, and TESK2 (grey parts are not translated)
- complex mechanism of control and regulation of gene expression

<http://genome.cshlp.org/content/14/2/280.long>

<http://www.nature.com/nature/journal/v457/n7232/full/nature07728.html>



Coronavirus update: overlapping genes

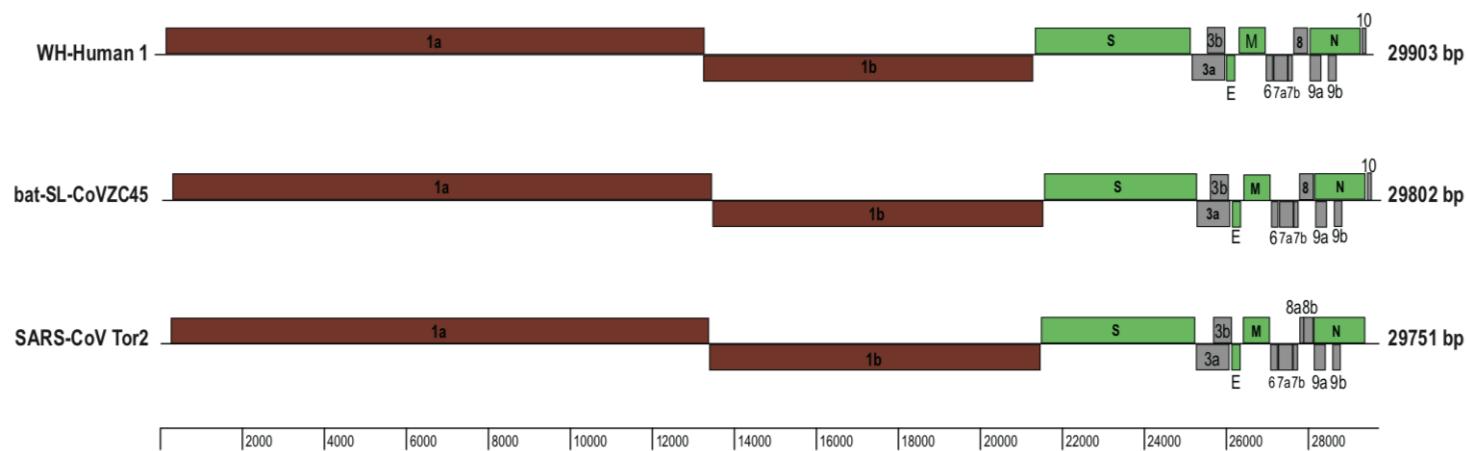


Figure 2. Genome organization of SARS and SARS-like CoVs including Tor2, CoVZC45 and WHCV determined here.

Wu et al. BioRxiv 2020 (<https://doi.org/10.1101/2020.01.24.919183>)

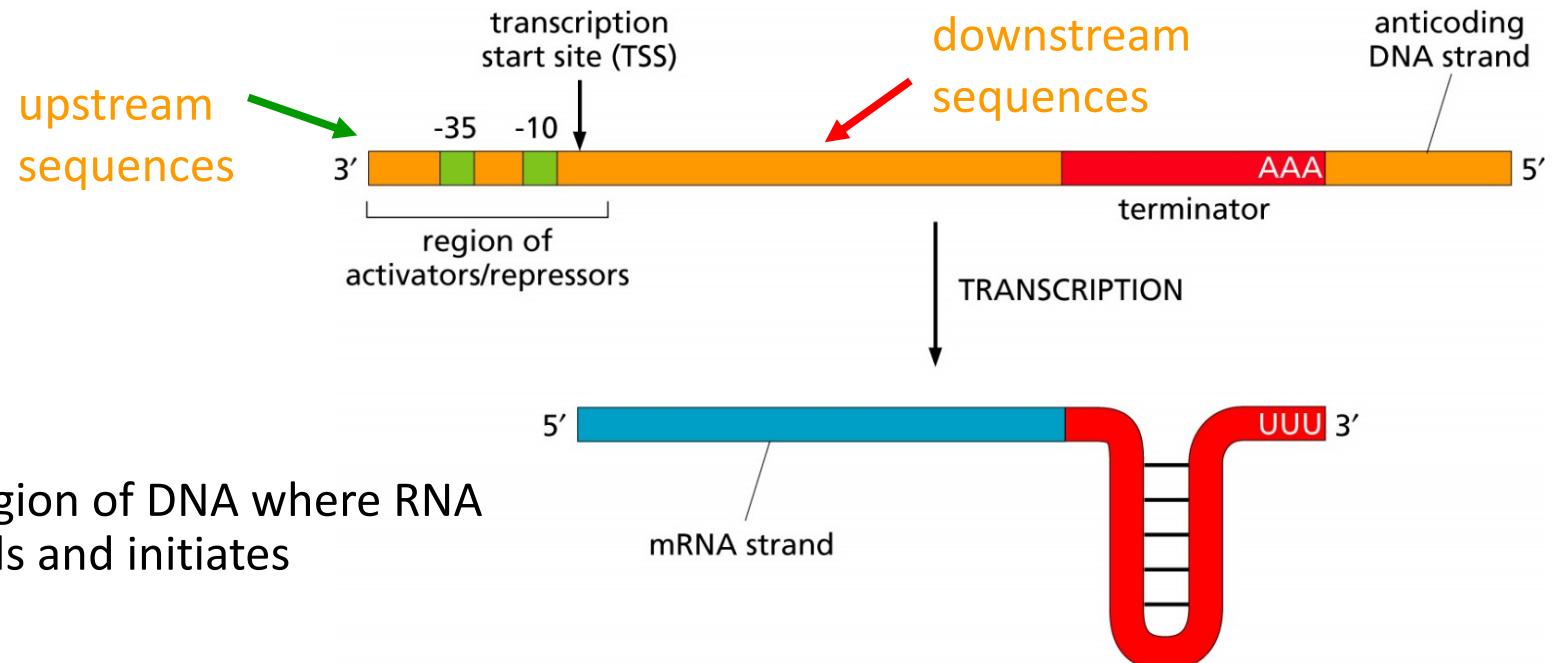
there is
more to
DNA than
just
genes...

- genomic DNA contains more than just the protein coding sequences:
 - control sequences (promoter sequences, stop codons, etc);
 - coding sequences (actual genes);
 - regulatory sequences (how much of the gene to express)
- ... and we still don't understand the huge part of it
- in humans, only 5% of protein-coding genes and for a lot of them, gene-protein-function relationship is unknown
→ bioinformatics

control and regulation

- gene expression is carefully regulated within a cell → expression dictates the cellular function
- control of transcription and translation rules the amount of every protein in the cell
- difference between prokaryotic and eukaryotic genomes → more complex mechanisms present in the eukaryotic cells
- eukaryotes have more complicated gene structure, transcription takes place within a nucleus physically separated from ribosomes (not the case in prokaryotes), multiple regulatory points and more regulatory proteins involved

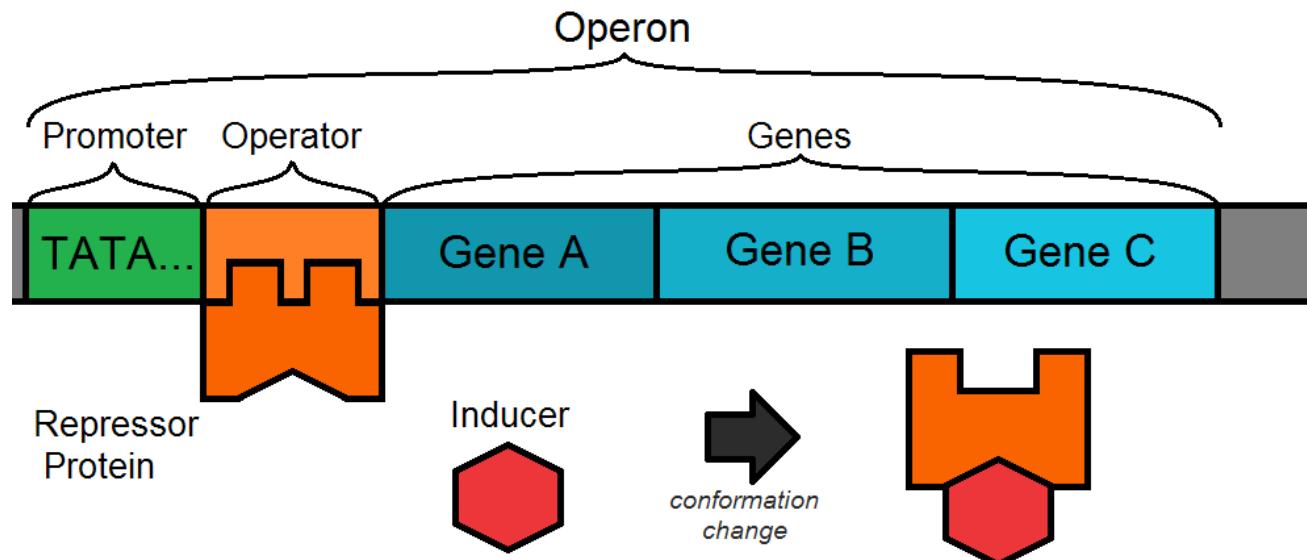
transcription control in bacteria



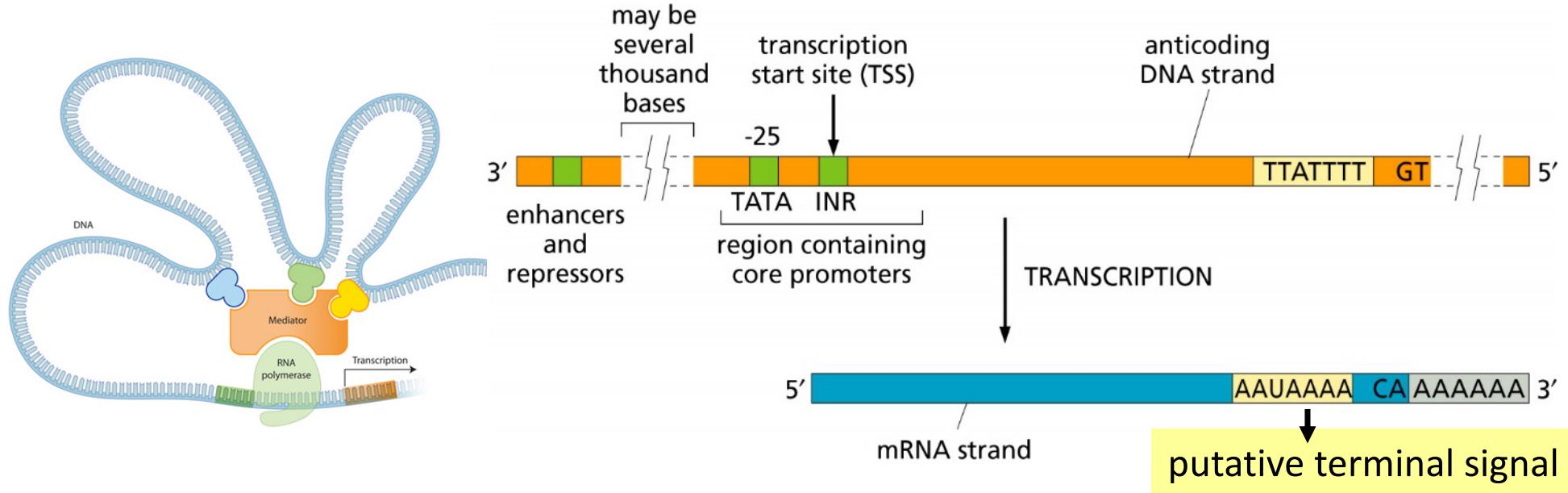
- *promoters* → region of DNA where RNA polymerase binds and initiates transcription
- problem of finding a promoter in DNA sequences due to variations → *consensus sequences* (most frequent bases at these positions) in *E. coli* at -10 TATAAT and at -35 TTGACA → tighter binding of polymerase means more frequently the region is being transcribed
- terminator sequences: two short stretches of complementary bases to form double helix + min. 3 consecutive U nucleotides
- additional controls → *activator* (improve efficiency) and *repressor* proteins (bind to operator site to block promoter sites) → crucial role in regulation
- DNA modifications (e.g. methylation) can prevent activator/repressor binding

Bacterial operons and gene expression

- Shine-Dalgarno sequence (consensus AGGAGGU) is located a few bases upstream from starting codon (AUG) → indicates ribosome binding site
- operons → clusters of functionally related protein-coding sequences transcribed as a single mRNA → specific for prokaryotes, rare in eukaryotes
- specific proteins translated separately → only one control region required to activate expression of several genes (i.e. for metabolic pathways)



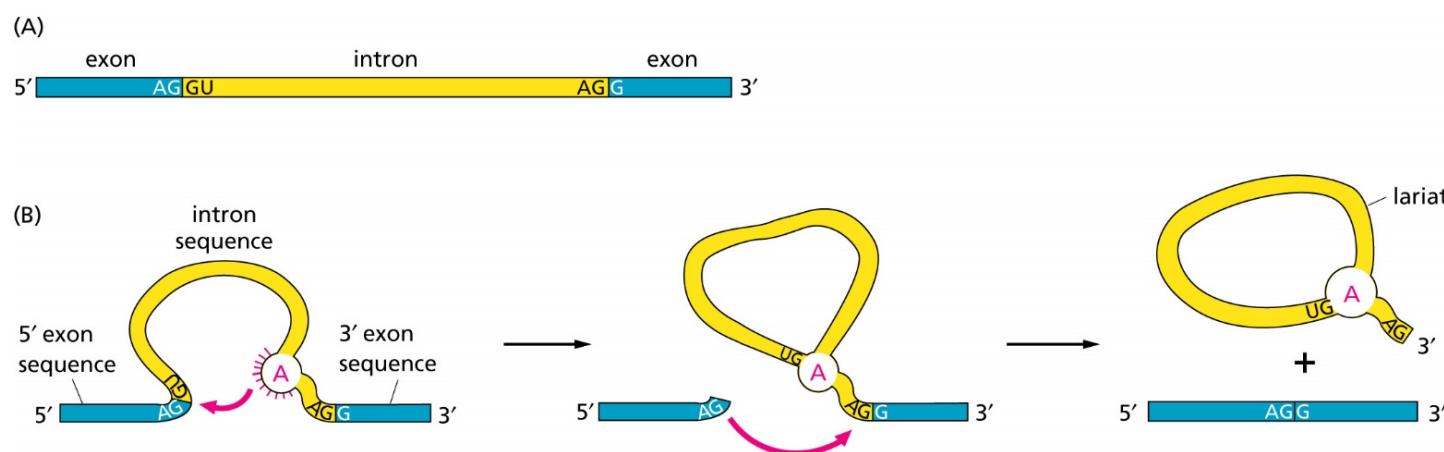
transcription control in eukaryotes



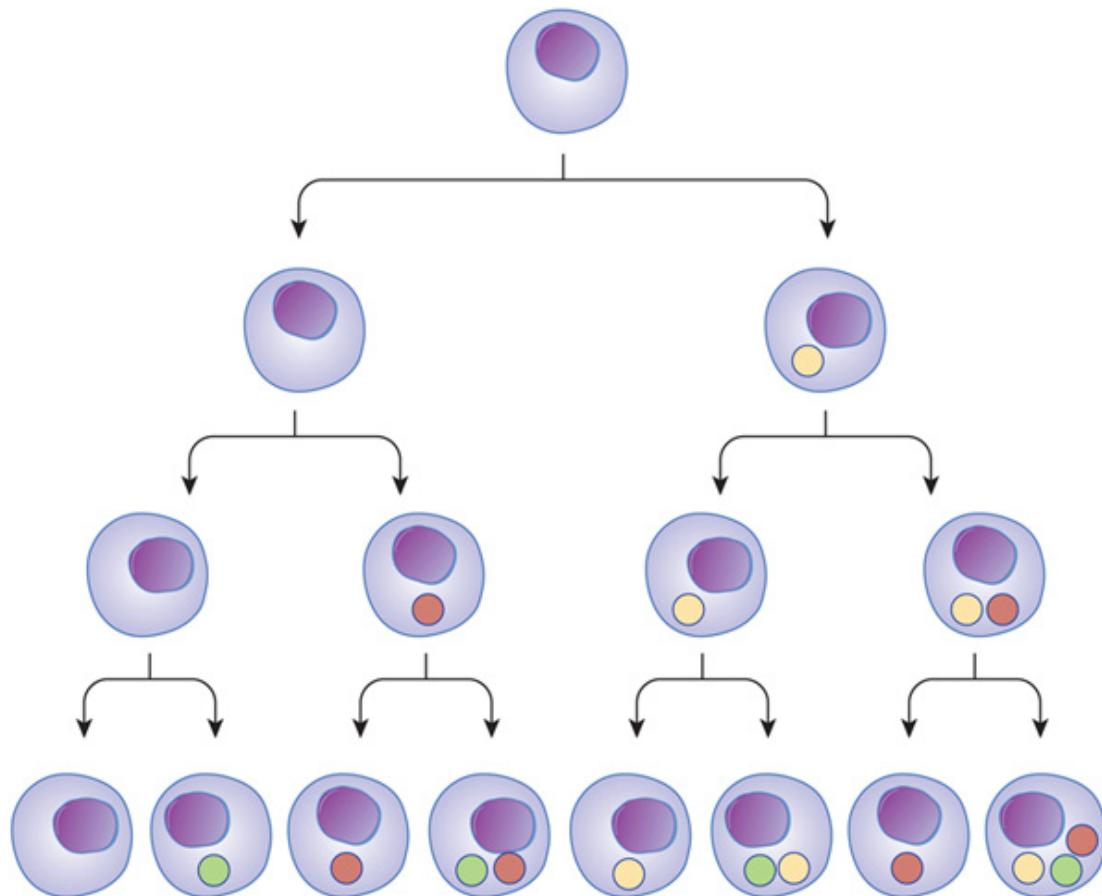
- three types of RNA polymerases → RNA polymerase II catalyses mRNA synthesis (types I and III are related to transcription of genes for tRNA, rRNA and other RNA molecules, regulated with different promoters)
- a set of core promoter signals in TSS region → binding of general transcription initiation factors → TATA box (TATA binding protein, TBP)
- regulatory regions (enhancers and silencers) controlling transcription can be far away from TSS in eukaryotes and both upstream and downstream → DNA loops to bring all the important regulatory proteins together into complex that controls transcription

mRNA modification in eukaryotes

- original mRNA transcript undergoes several modifications before translation in eukaryotes:
 - RNA capping – addition of modified G nucleotide to 5' end (role in ribosome binding and translation)
 - polyadenylation of 3' end after cleaving mRNA triggered by AAUAAA signal (~ 200 polyA chain)
 - RNA splicing – excision of *introns* (noncoding stretches of DNA) and merging *exons* (protein-coding sequences) → process carried out by spliceosome (*small nuclear RNA + proteins*)
- alternative splicing → different ways of merging the same exons resulting in greater protein variability from smaller number of genes

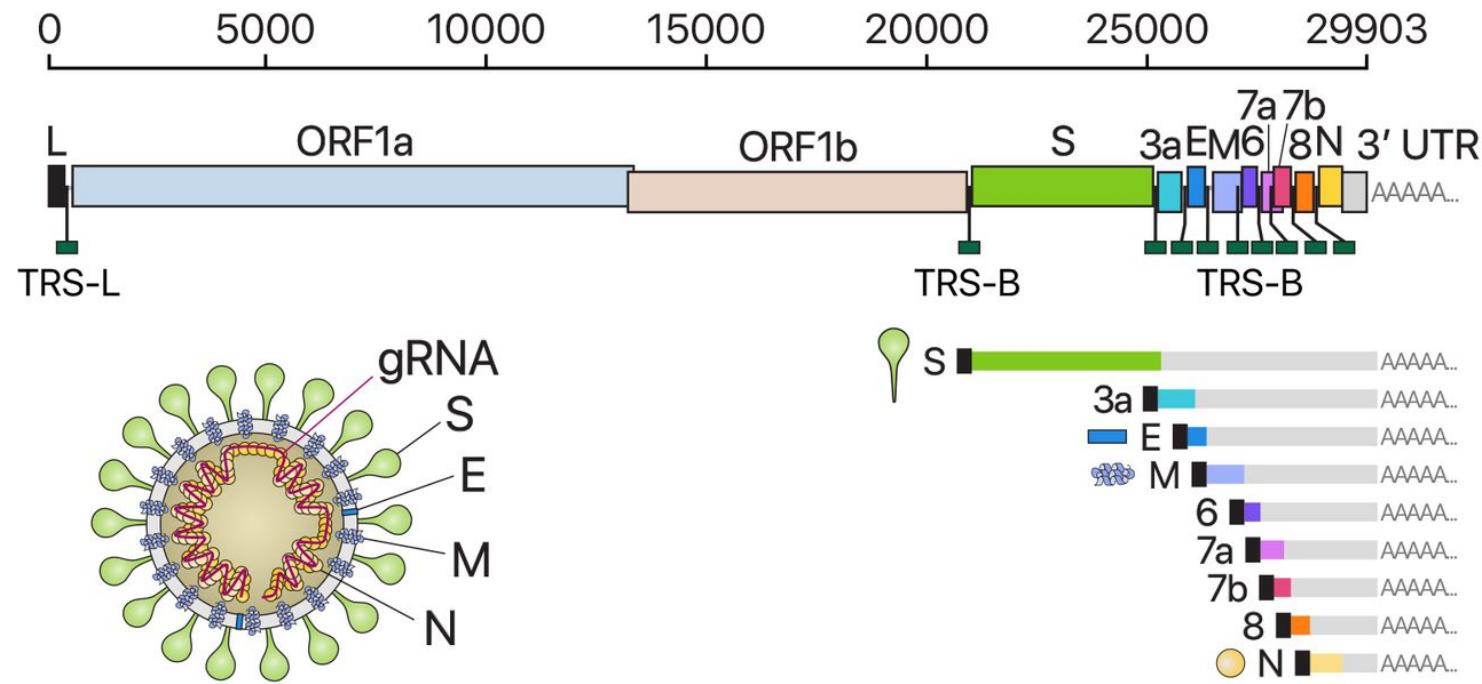


gene expression determines cell type



- the wide variety of cell types in a single organism can depend on different transcription factor activity in each cell type
- different transcription factors can turn on at different times during successive generations of cells resulting ultimately with different cell types

Coronavirus update: gene regulation



Kim et al., Cell 2020, DOI:<https://doi.org/10.1016/j.cell.2020.04.011>

some useful links:

ONLINE TEXTBOOK ON CELL BIOLOGY:

<http://www.nature.com/scitable/ebooks/essentials-of-cell-biology-14749010/contents>

CENTRAL DOGMA ANIMATION VIDEOS:

<http://www.youtube.com/watch?v=J3HVVi2k2No>

<http://www.youtube.com/watch?v=ZNcFTRX9i0Y>

TRANSCRIPTION VIDEO:

<https://www.youtube.com/watch?v=5MfSYnItYvg>

Key learning outcomes

- Understand the basic steps in transcription
- Recognize how genes and genomes are structured.
- Understand the key features of an operon and how it functions in bacterial gene control.
- Appreciate how functional mRNA molecules produced in eukaryote organisms.

Refer to Week 1-2 Study Guide document

Introduction to molecular biology: Protein synthesis and structure



A/Prof Scott Beatson

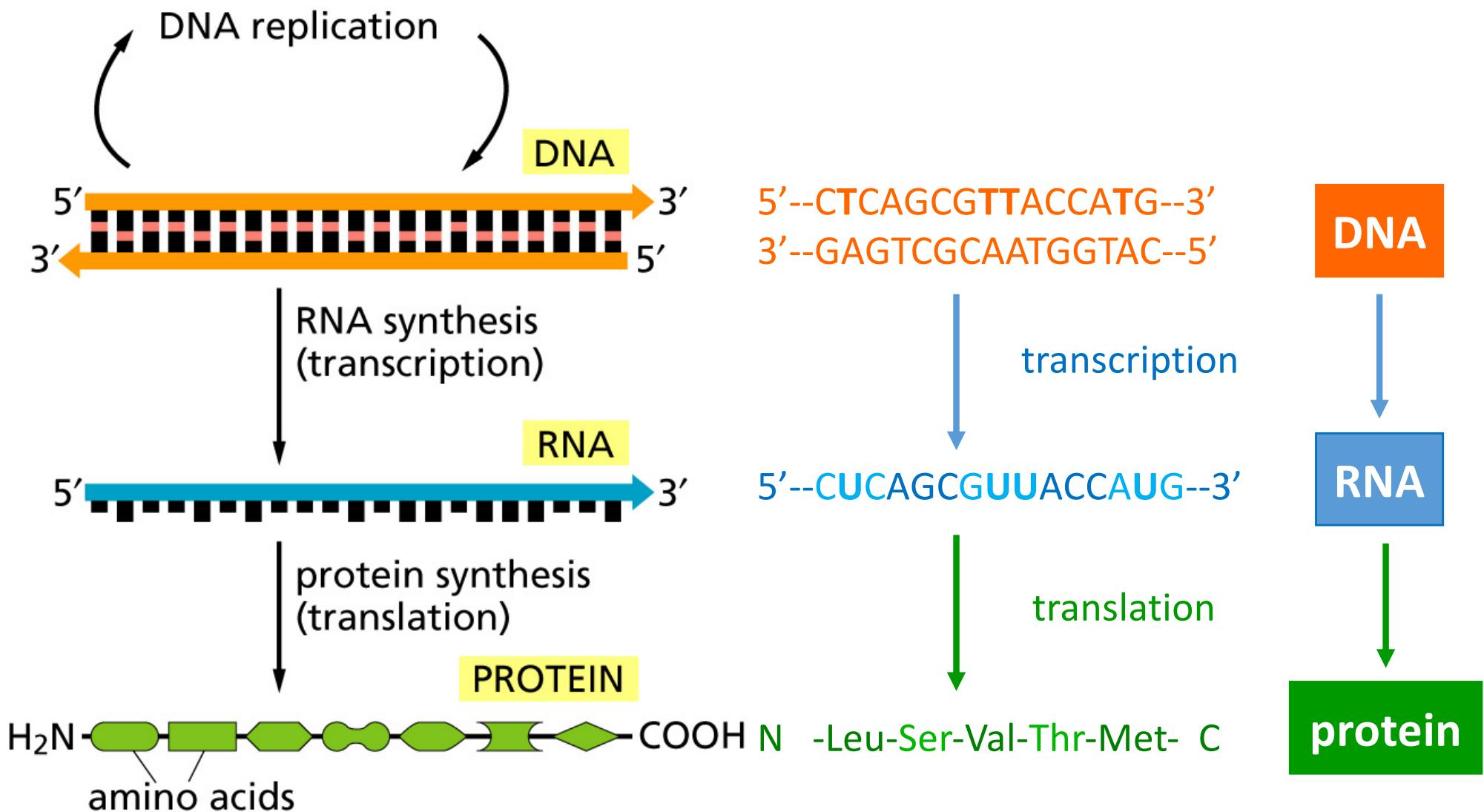


s.beatson@uq.edu.au



BIOC6000/SCIE2100

central dogma of molecular biology



translation: genetic code

- translation of mRNA into protein by following according to the genetic code (*which base for which amino acid?*)
- 20 different amino acids build proteins
- each amino acid is encoded by set of 3 consecutive bases (ATCG) → *codons*
- 4^3 combinations → degeneracy of genetic code → most amino acid can be specified by more than one codon
- implication → you can deduce protein sequence from DNA or RNA, but not vice versa

standard genetic code

		Second letter of the codon							
		U	C	A	G			3' end	
First letter of the codon	5' end	U	C	A	G			3' end	
	U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U			
	U	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C			
	U	UUA Leu	UCA Ser	UAA Stop	UGA Stop	A			
	U	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G			
C	C	CUU Leu	CCU Pro	CAU His	CGU Arg	U			
	C	CUC Leu	CCC Pro	CAC His	CGC Arg	C			
	C	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A			
	C	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G			
A	A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U			
	A	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C			
	A	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A			
	A	AUG Met	ACG Thr	AAG Lys	AGG Arg	G			
G	G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U			
	G	GUC Val	GCC Ala	GAC Asp	GGC Gly	C			
	G	GUA Val	GCA Ala	GAA Glu	GGA Gly	A			
	G	GUG Val	GCG Ala	GAG Glu	GGG Gly	G			

5' — C U C A G C G U U A C C A U — 3'

— Leu — Ser — Val — Thr —

translation: reading frames

- translation of non-overlapping sets of three bases (codons) → 3 possible ways to read the code → *reading frames*

#1 5' — C U C A G C G U U A C C A U — 3'
— Leu — Ser — Val — Thr —

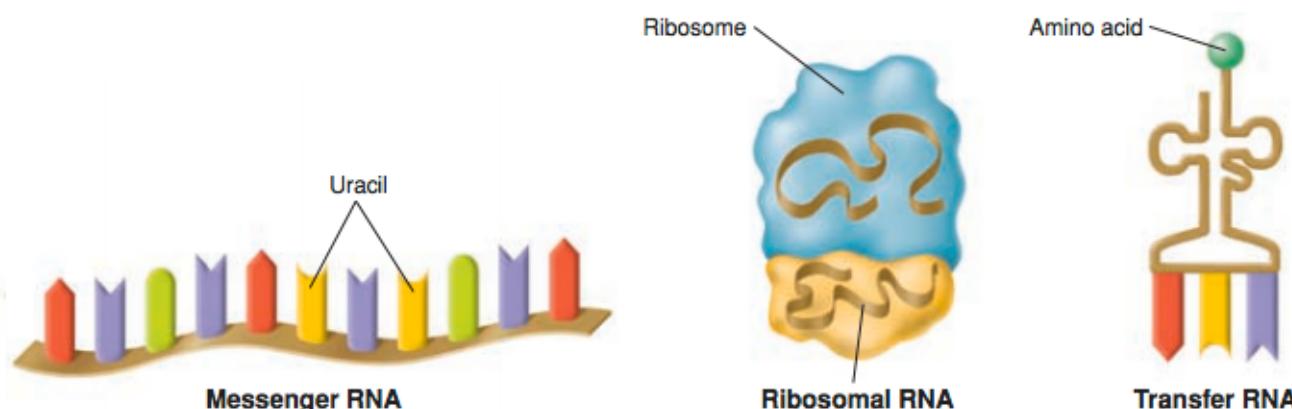
#2 5' — C U C A G C G U U A C C A U — 3'
— Ser — Ala — Leu — Pro —

#3 5' — C U C A G C G U U A C C A U — 3'
— Gln — Arg — Tyr — His —

- prediction of protein-coding sequence → *open reading frame* (ORF)
→ a segment of DNA that when translated to amino acids contains no stop codons (UAA, UAG, UGA)

different RNAs are involved in translation

- three main classes of RNA in all the cells (part of the translation process):
 - *messenger RNA (mRNA)*: formation during transcription from DNA (information carrier)
 - *transfer RNA (tRNA)*: mediator in the recognition between the codons and amino acids
 - *ribosomal RNA (rRNA)*: together with proteins forms *ribosomes*
→ molecular machines for protein synthesis
- numerous smaller RNAs with various roles (i.e. small nuclear RNA, snRNA)



Types of RNA The three main types of RNA are messenger RNA, ribosomal RNA, and transfer RNA. Ribosomal RNA is combined with proteins to form ribosomes.

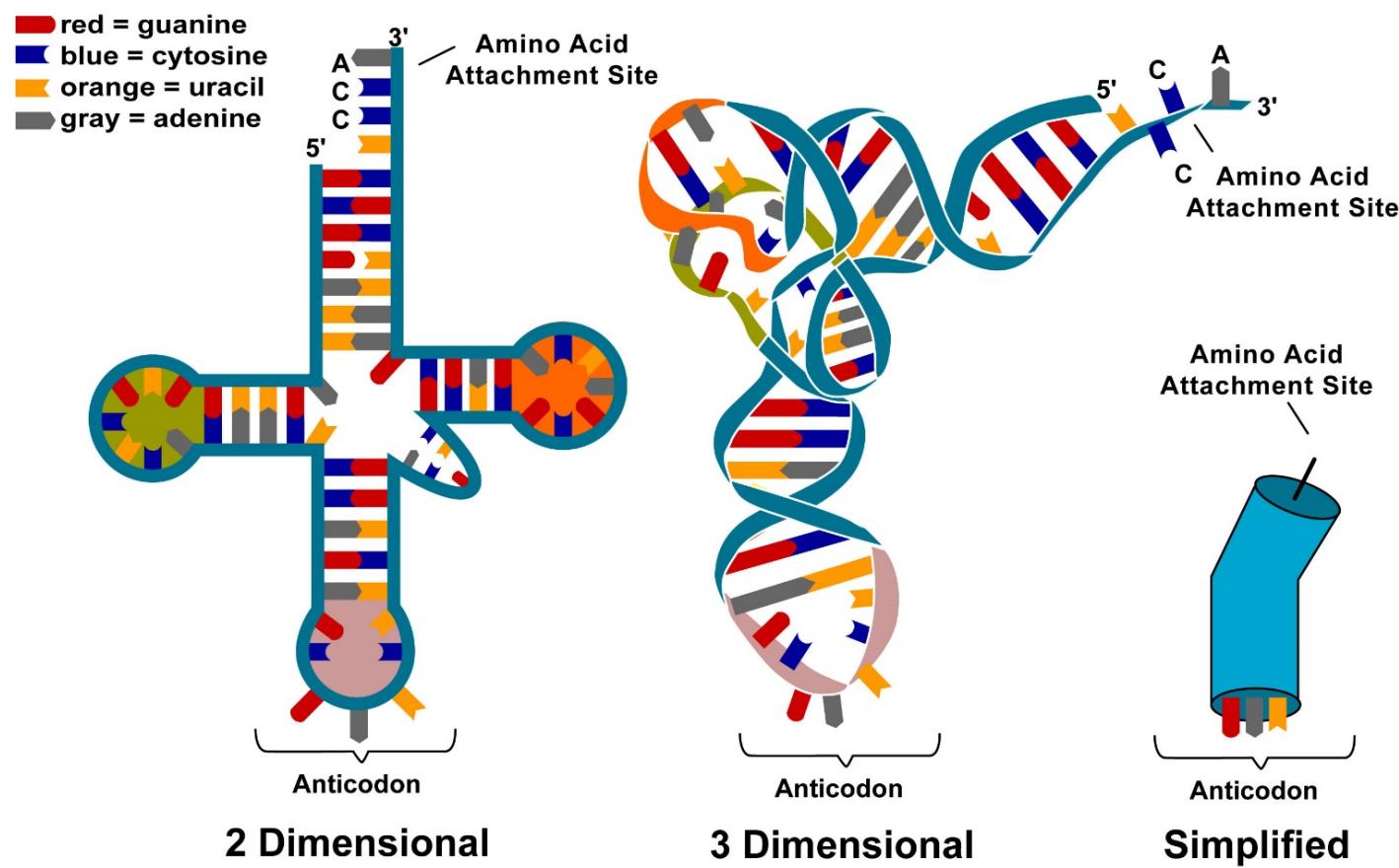
Video: protein synthesis



<http://www.youtube.com/watch?v=lpb5s2F1pyM>

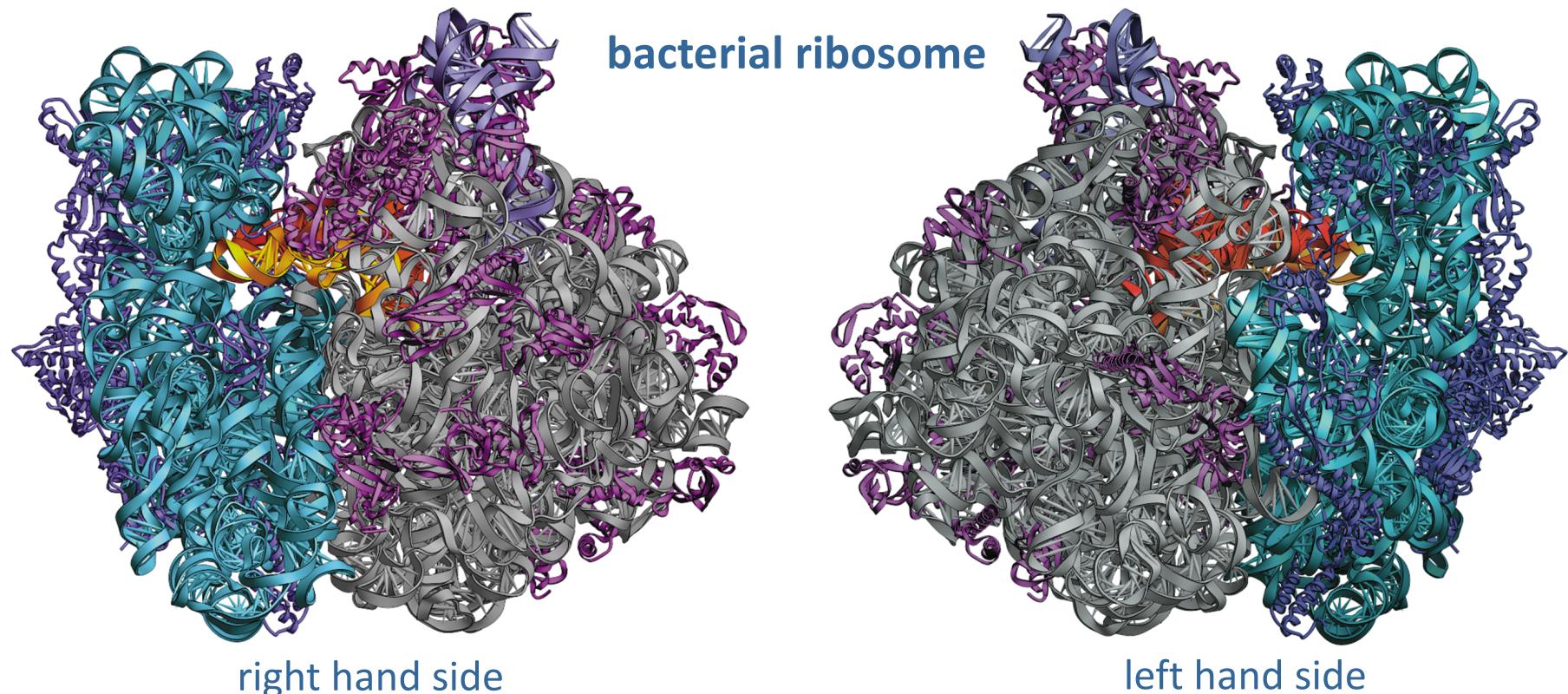
transfer RNA (tRNA)

- amino acids do not recognize codons directly → tRNA as a mediator
- tRNA contains a *binding site for the specific amino acid* at one end and an *anticodon* at the other → amino acid attached by aminoacyl-tRNA synthetase
- codon (mRNA) – anticodon (tRNA) recognition (via complementary bases)

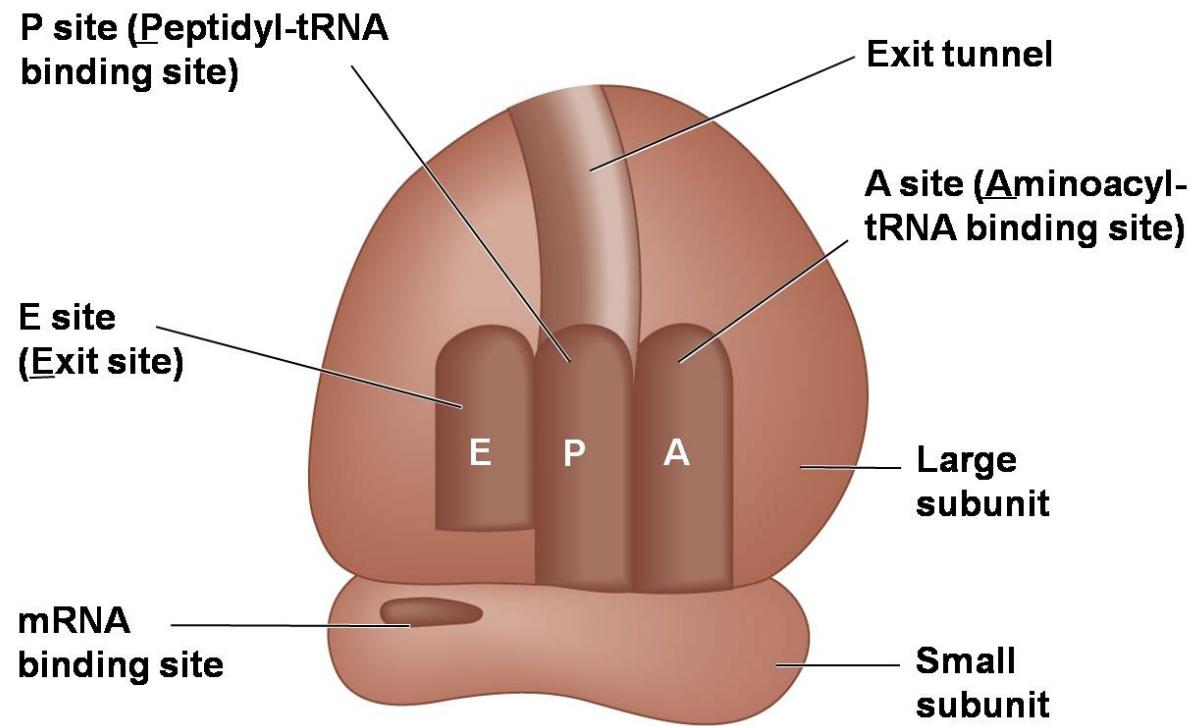


Ribosomes: protein synthesis machines

- ribosomes → large complexes of rRNA and proteins → two main subunits: *small* (mRNA recognition) and *large* (peptidyl transfer)
 - first high resolutions structures in 2000
 - Nobel prize in 2009 awarded “*for studies of the structure and function of ribosome*” to V. Ramakrishnan, T. A. Steitz and A. E. Yonath

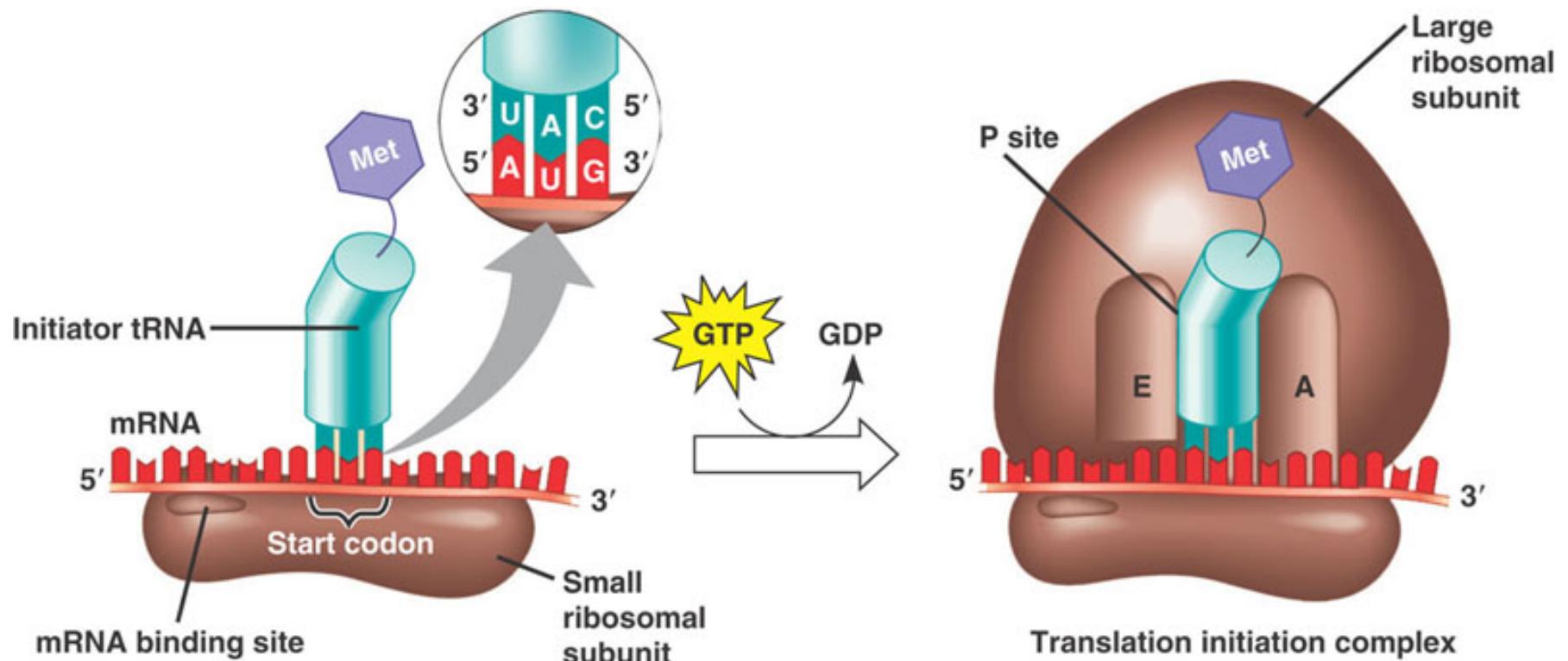


ribosome



- ribosome catalyses formation of polypeptide chain (protein) by translation the genetic code on mRNA in three steps:
 1. ***initiation*** – complexation of ribosome, mRNA and tRNA
 2. ***elongation*** of the protein chain by adding new amino acids by following instruction encoded in mRNA (large subunit)
 3. ***termination*** – decomposition of the ribosomal machinery

1. translation: initiation



- assembly of the necessary components for translation around the start codon on mRNA (AUG)
 - Requires protein initiation factors (e.g. IF1, IF2 and IF3)
- the initiator tRNA carries the amino acid that corresponds to the starting codon – methionine

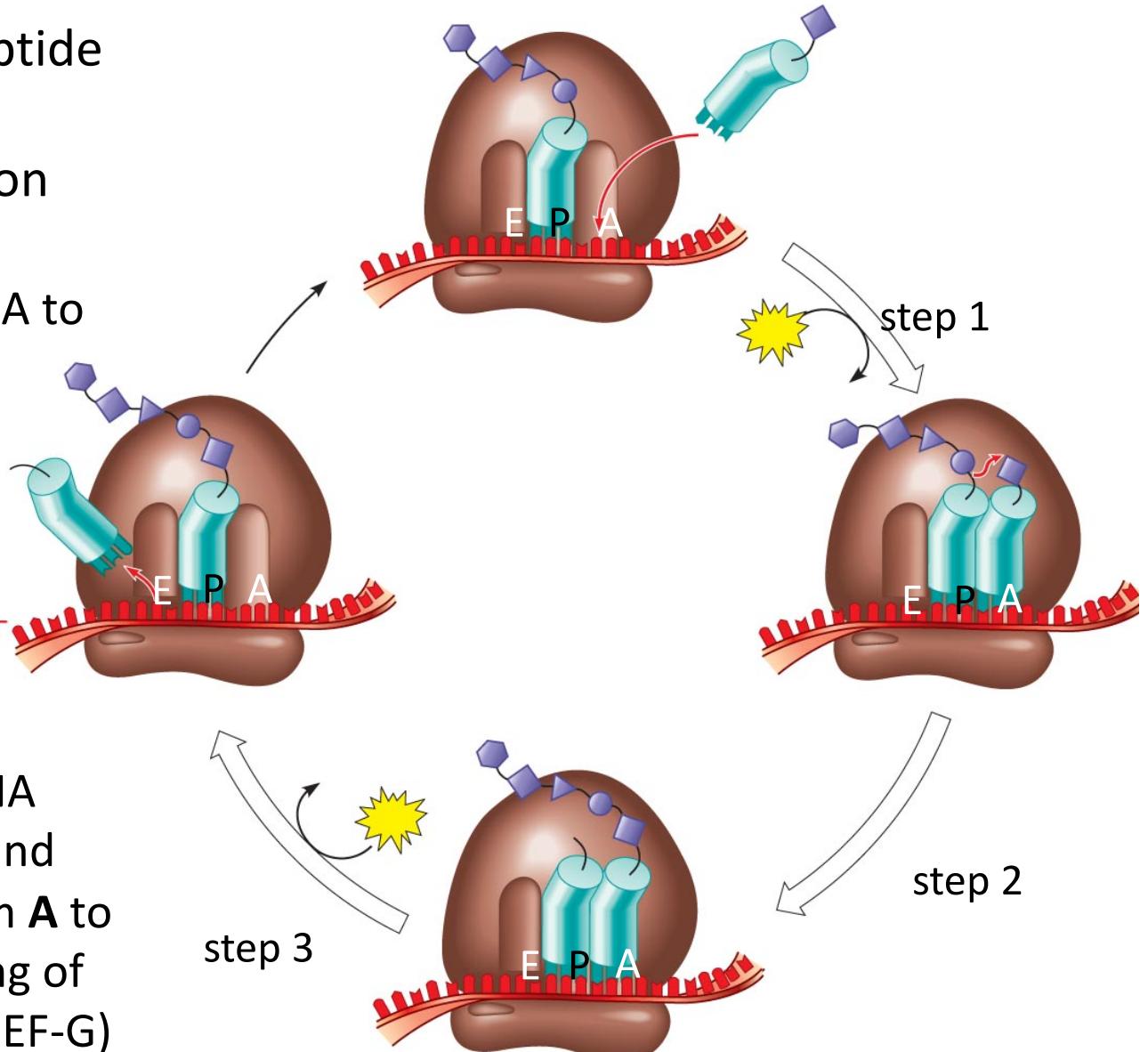
2. translation: elongation

- elongation of the polypeptide chain proceeds in 3 steps (participation of elongation factors – EF):

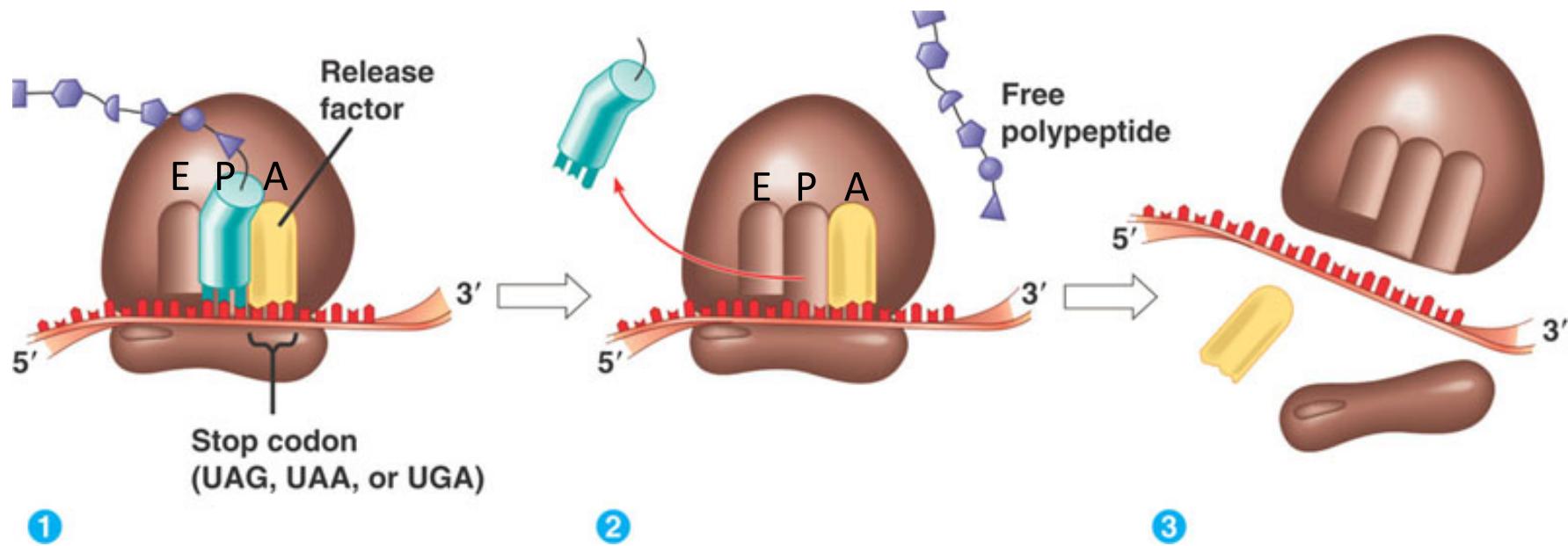
1. ***binding*** of aminoacid-tRNA to **A site (EF-Tu)**

2. ***peptidyl transfer*** involves migration of the existing peptide chain from **P site** upon binding the new amino acid

3. ***translocation*** is step during which the lone tRNA is kicked from **P** to **E** site and peptidyl-tRNA moves from **A** to **P** site, resulting with sliding of mRNA to the next codon (EF-G)

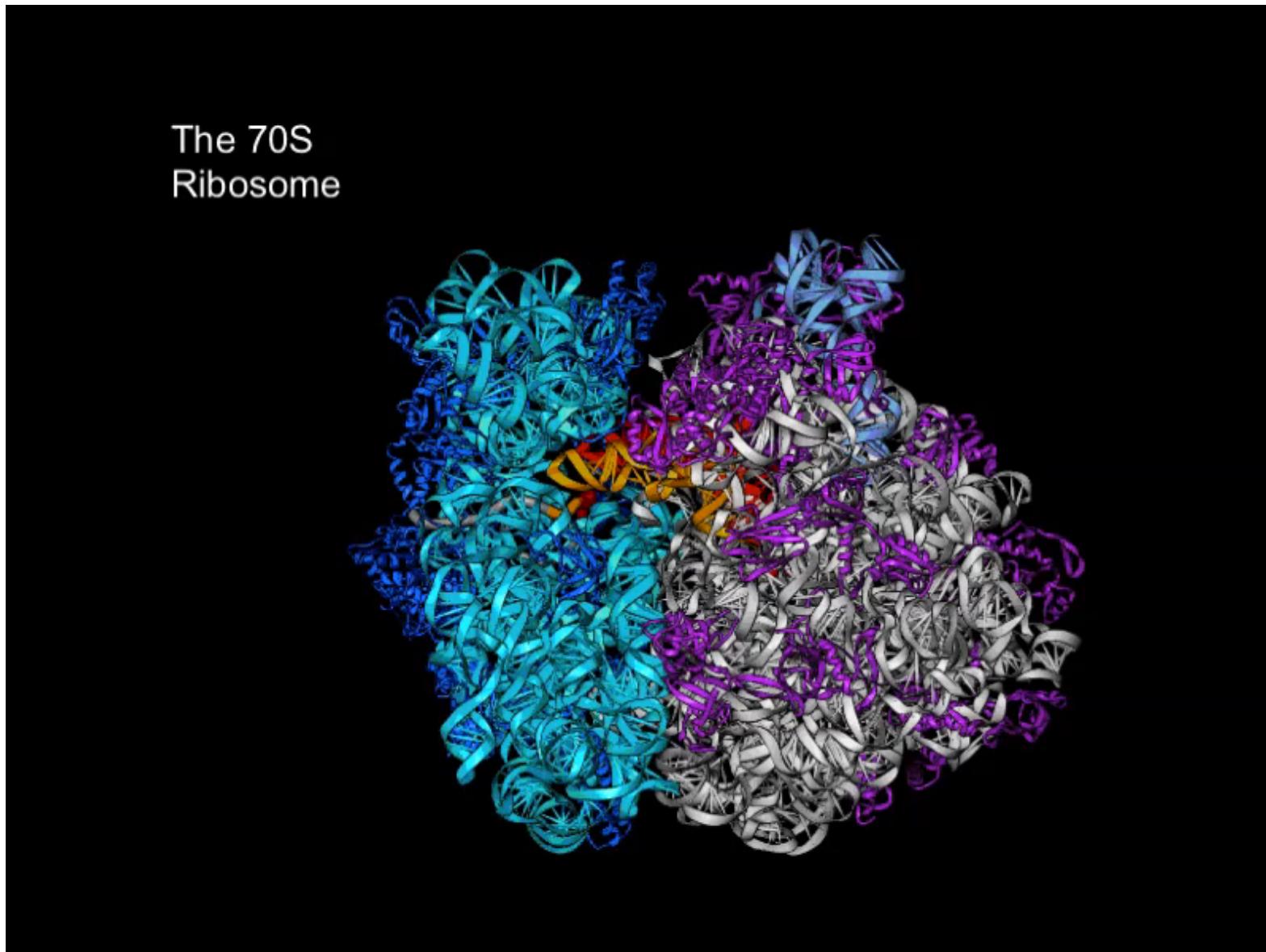


3. translation: termination

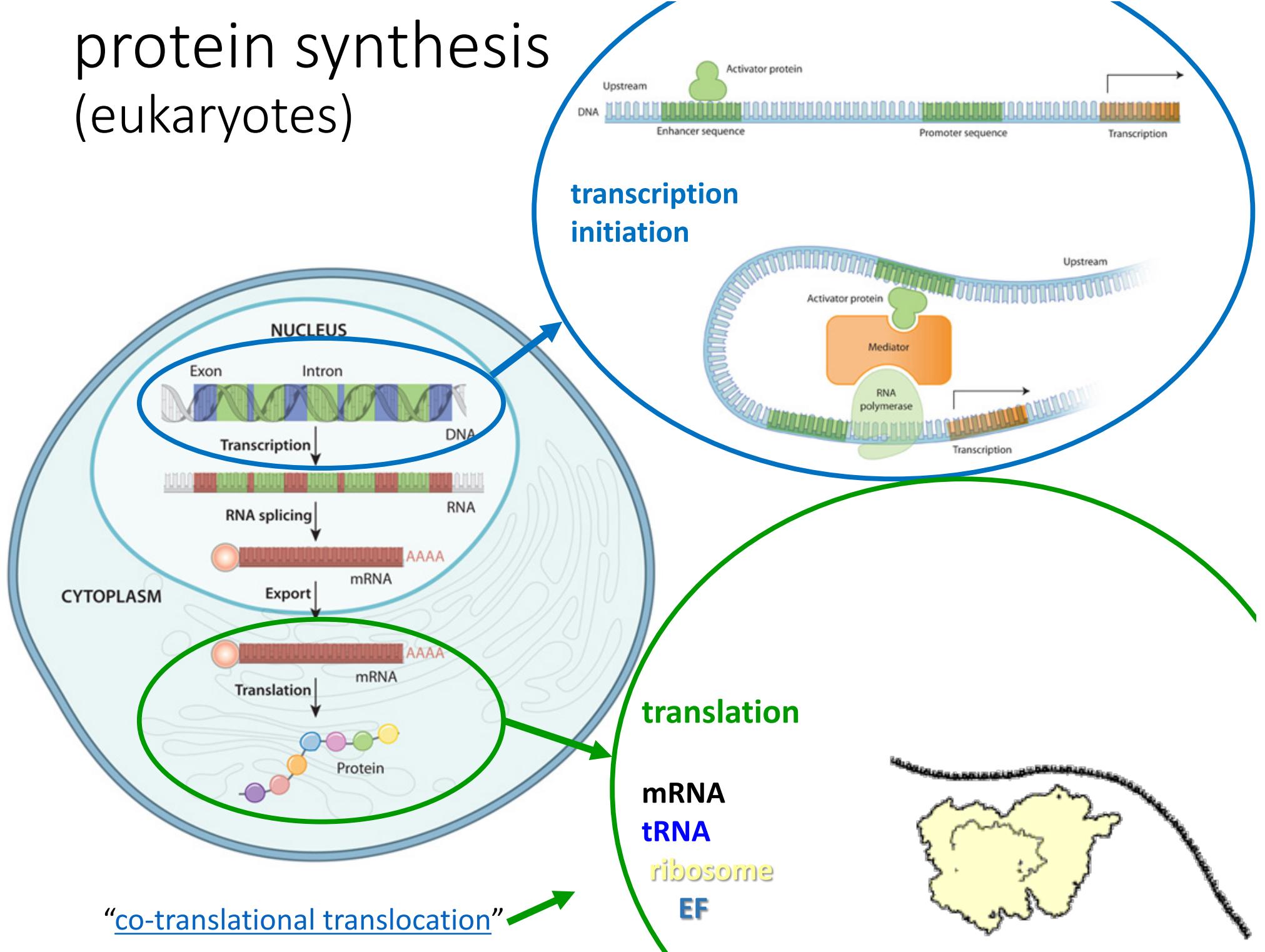


- termination of translation process after finding a stop codon on mRNA → binding of the release factor (RF) at the site A
- cleavage and release of the new polypeptide (protein) from the P site, followed by disassembling of the ribosome complex

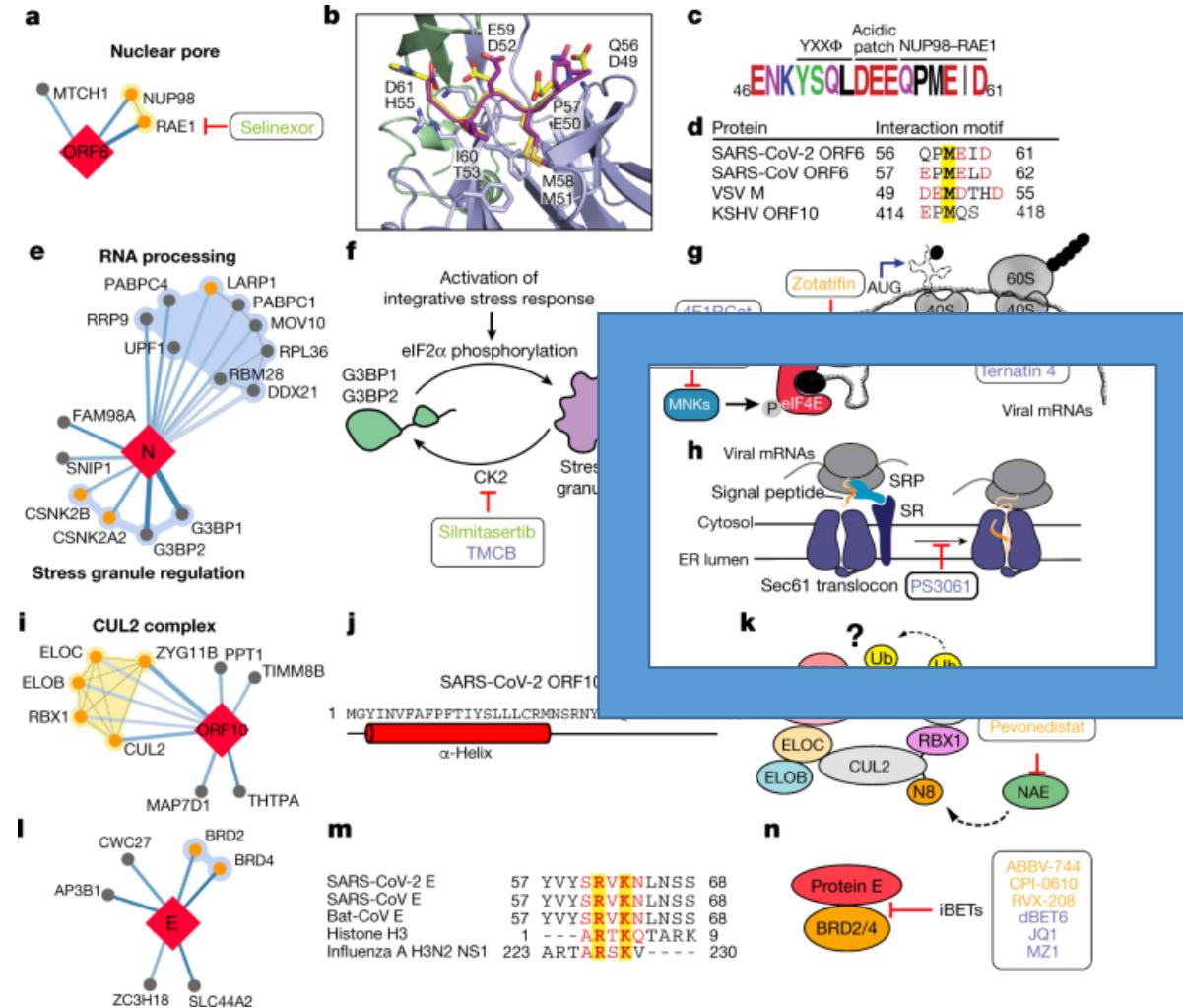
elongation & termination



protein synthesis (eukaryotes)



Coronavirus update: SARS-CoV-2 hijacks co-translational translocation



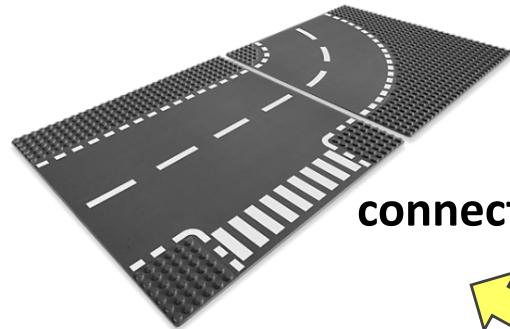
Gordon, D.E., Jang, G.M., Bouhaddou, M. et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 583, 459–468 (2020).
<https://doi.org/10.1038/s41586-020-2286-9>

if a cell was a lego city...



- ... then **DNA** would be the building instructions for various structures written in Latin, for example;
- **mRNA** would be a copy of instruction;
- **tRNA** would be a dictionary;
- **amino acids** would be lego pieces...

proteins: infrastructure



connections



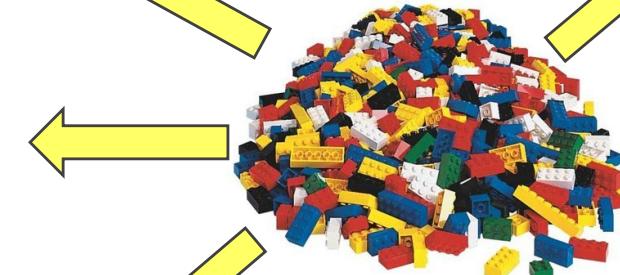
transport



**breakdown of fuel
(sugars, fats, proteins)**



**synthesis of new
compounds**



signalling



motion



assembly

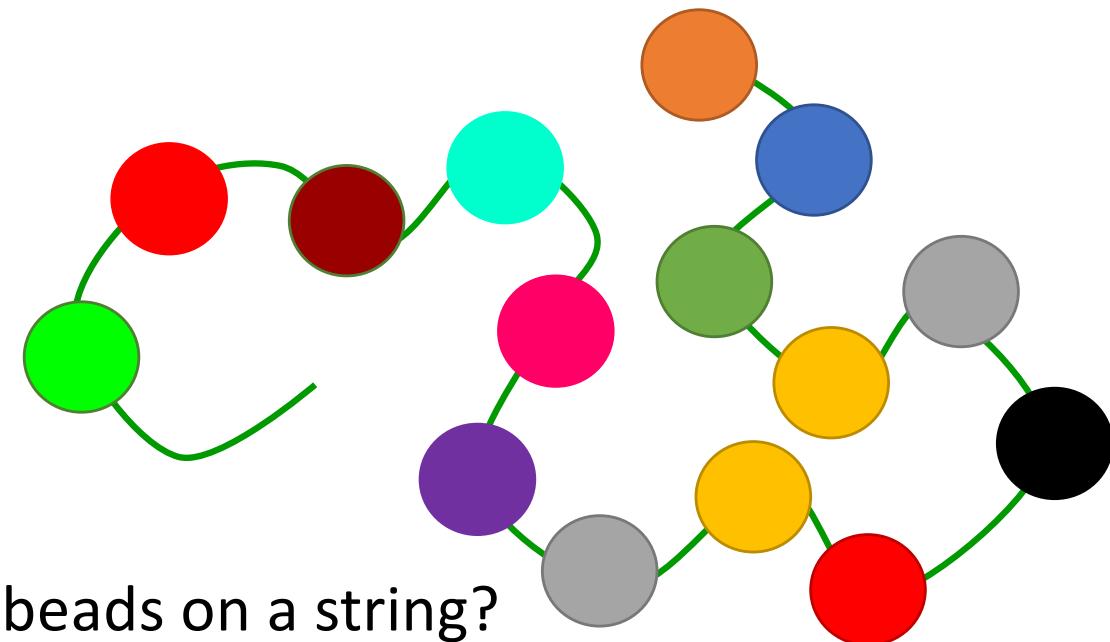


storage

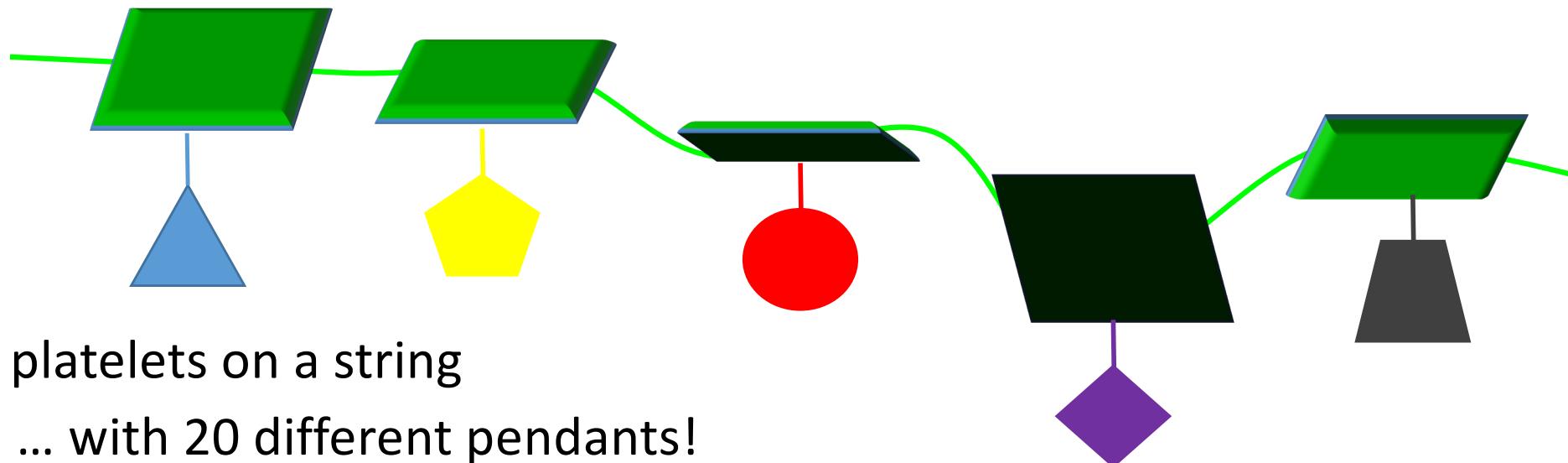
proteins



lego pieces?



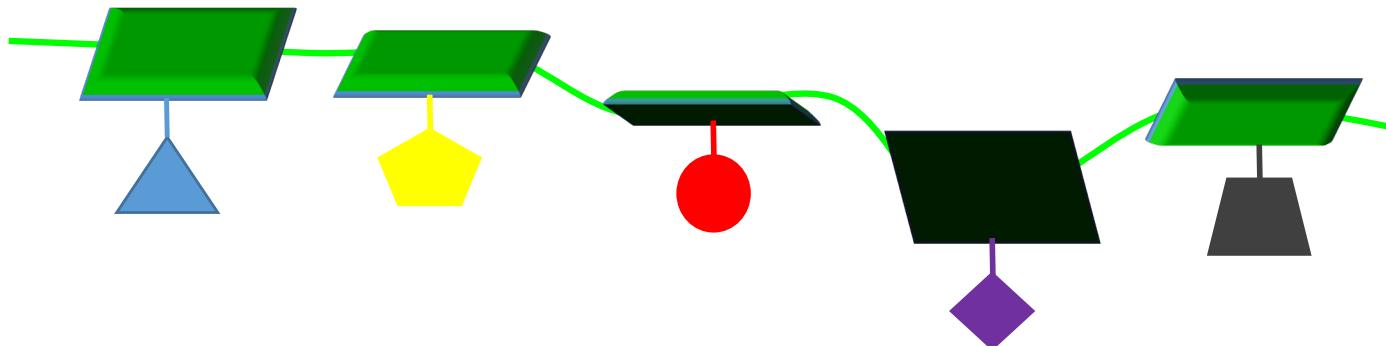
beads on a string?



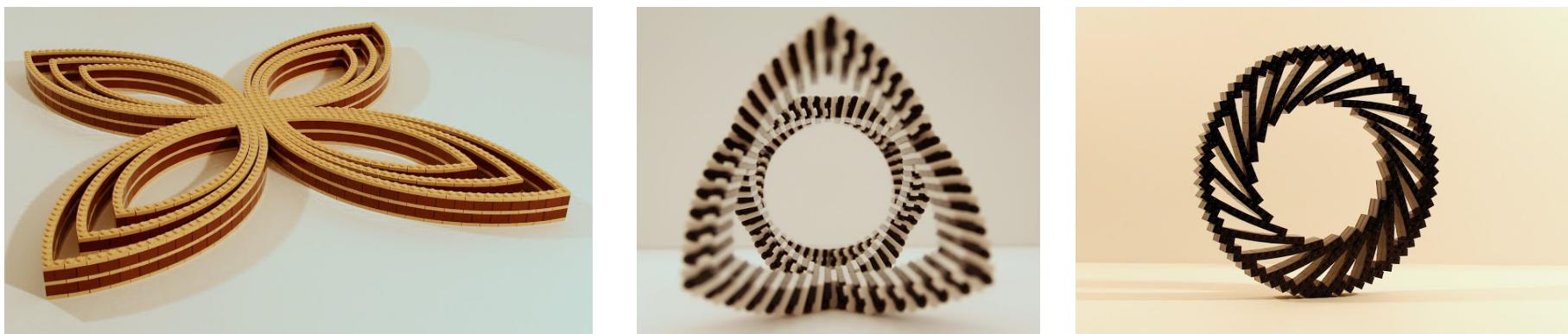
platelets on a string

... with 20 different pendants!

protein structure



- working definition: proteins are long chains of platelets with 20 different pendants → ability to create various 3D shapes and forms



- correct structure of crucial importance for protein functionality
→ structure-activity relationship (SAR)

protein structure

- long molecules can have a very large number of possible conformations → only a few provide biological activity
- *Levinthal paradox* → astronomical number of conformation available even to the small proteins → exploring all of them would take an eternity → folding into a native structure on micro- and millisecond scale

unfolded

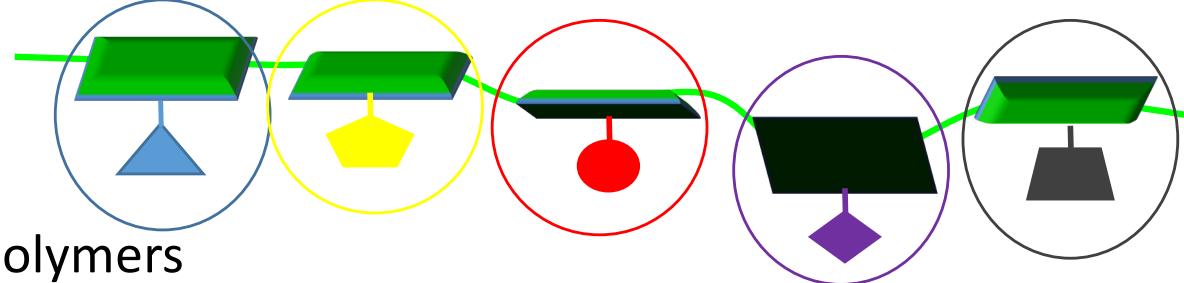


folded
(native)



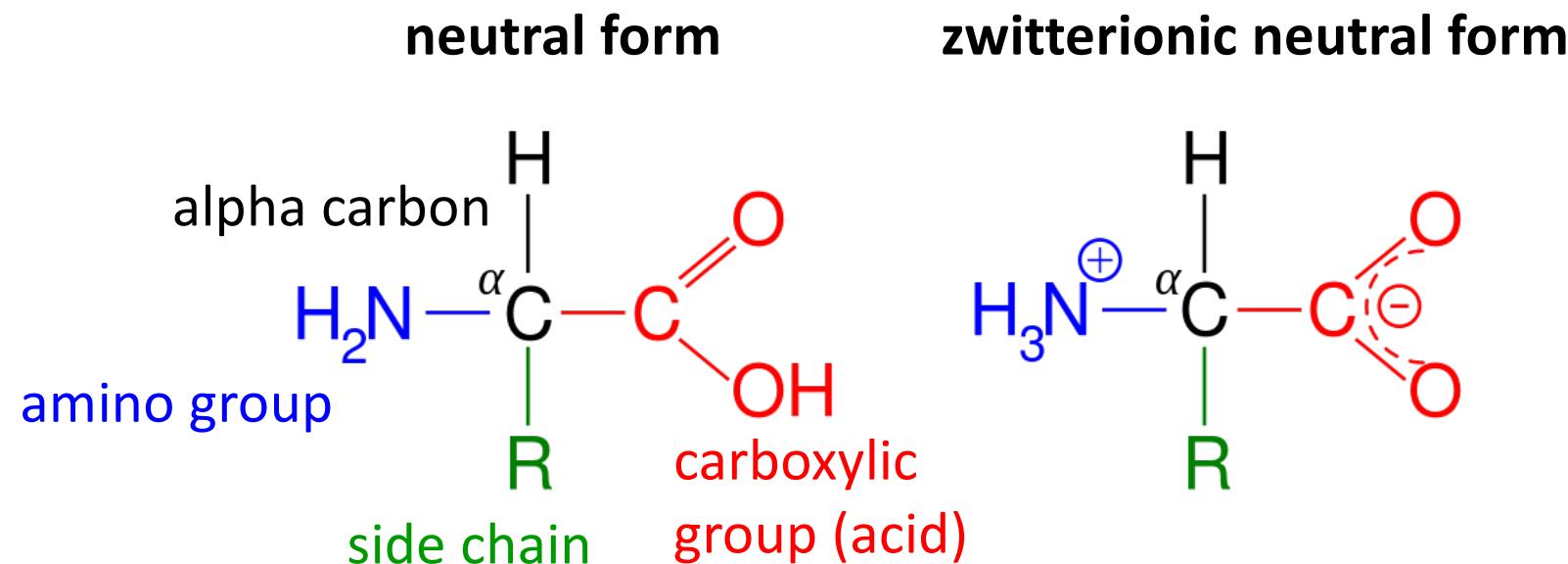
- interactions within protein and with its environment → the pendants on our long chain have some preferences for other pendants and general surrounding, platelets also prefer some angles over other...

proteins = amino acid polymers



amino acids

- 20 different amino acids build all the proteins



- 20 different side chains (pendants)
- zwitter = hybrid in german

amino acids

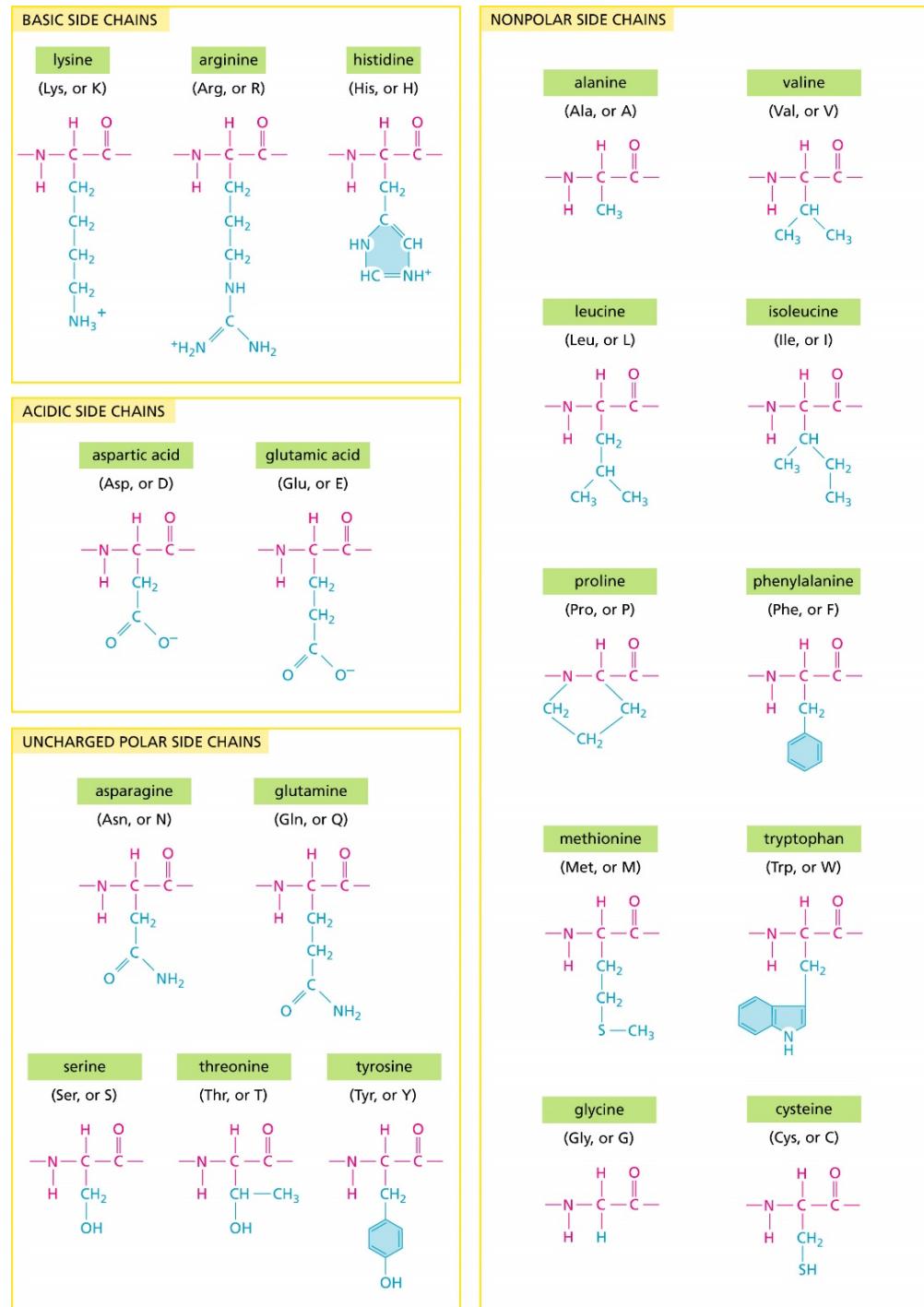
- amino acids can be classified in a few different ways
- E.g. classification into 4 groups based on the chemical properties of their side chain:
 - basic (positively charged);
 - acidic (negatively charged);
 - polar (uncharged);
 - nonpolar (uncharged)

base: proton acceptor

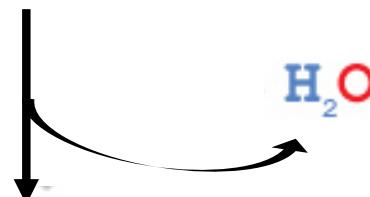
acid: proton donor

polar residues: contain a dipole (neutral molecules with charge separation)

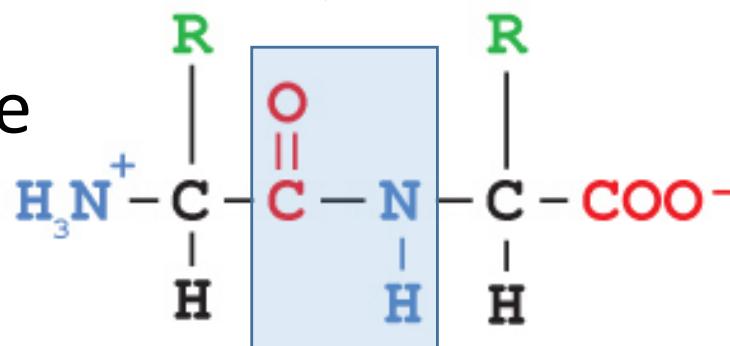
nonpolar residues: neutral molecules without a dipole (weak interactions, hydrophobic)



peptide bond

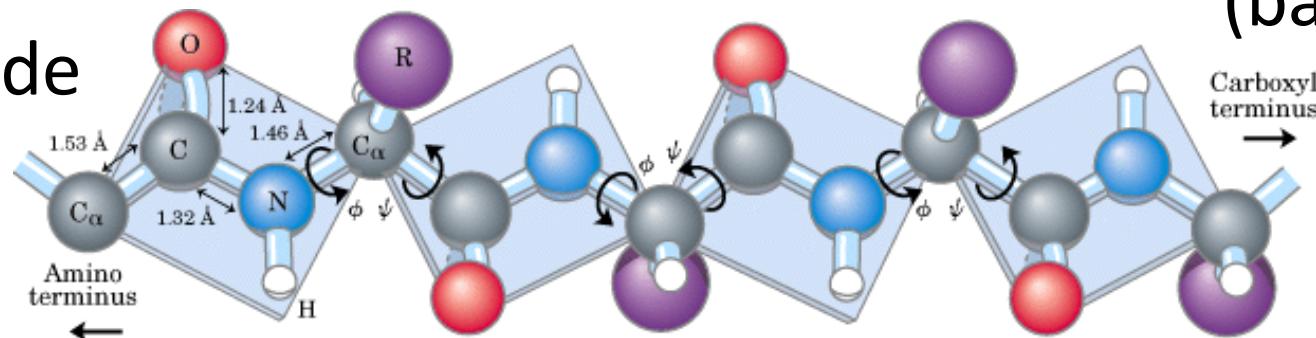


dipeptide

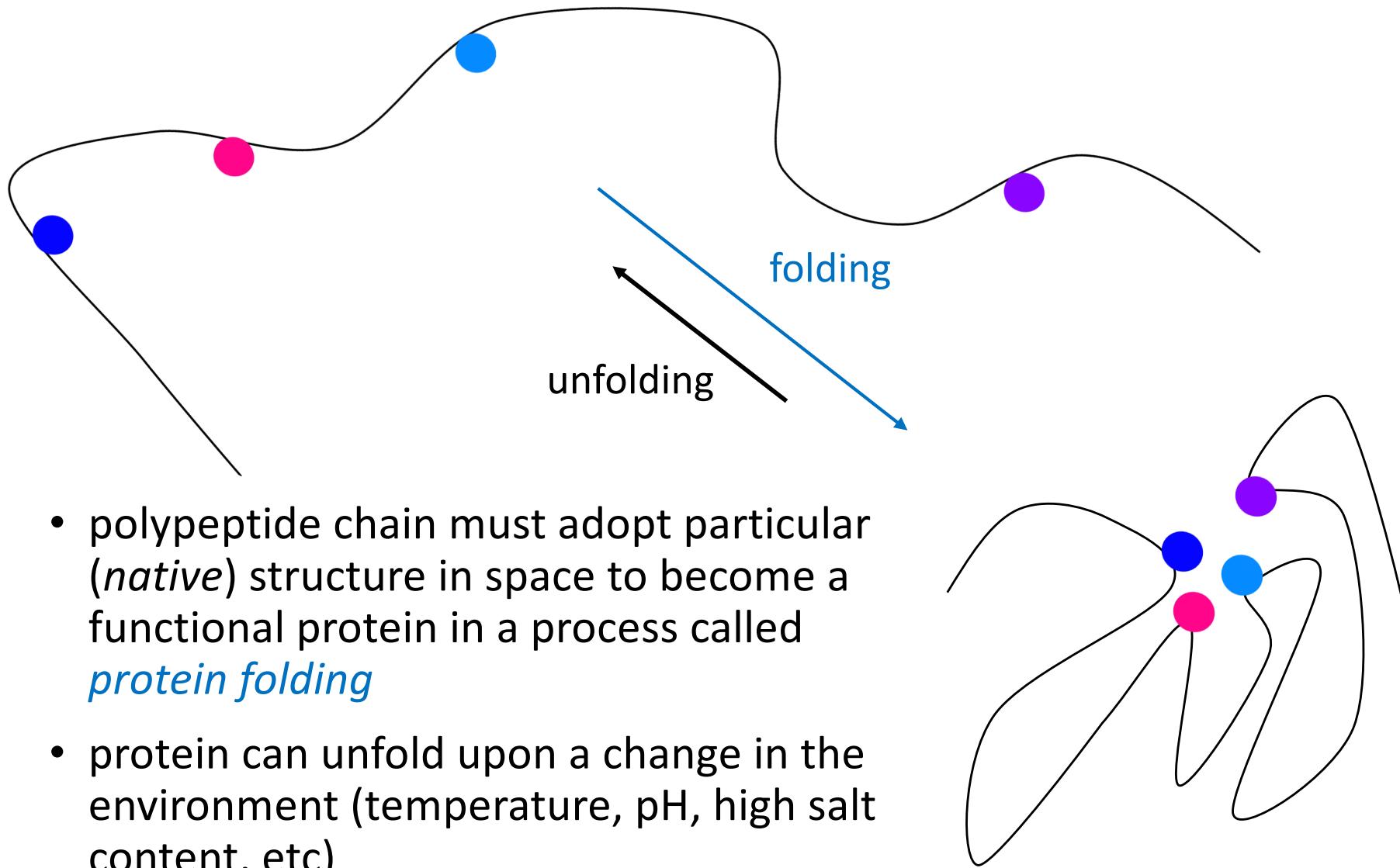


main chain
(backbone)

polypeptide

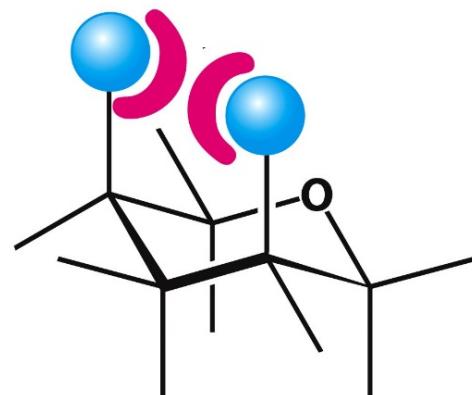


protein vs. polypeptide

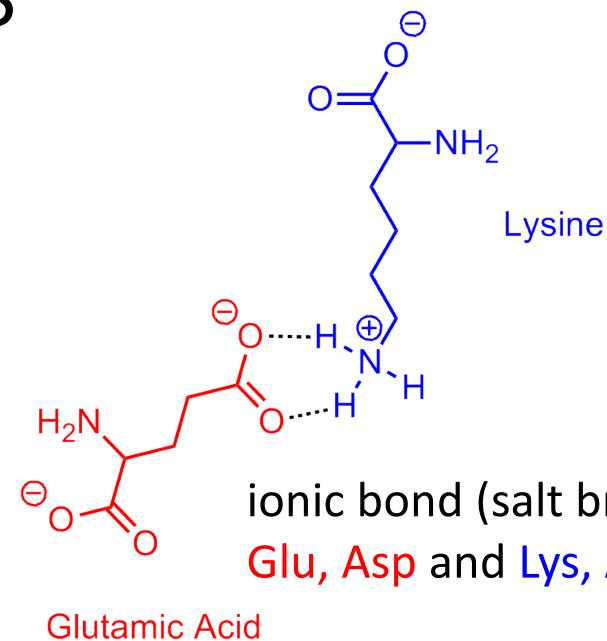
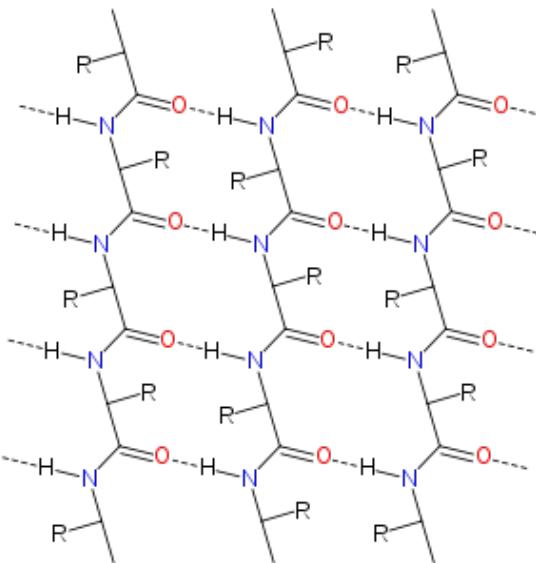


interactions & motions

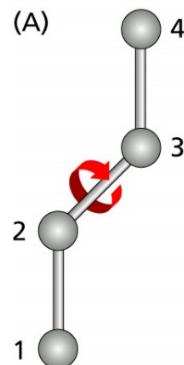
steric hindrance



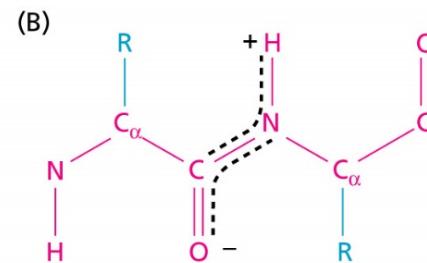
Unnumbered 11 p324
Biochemistry, Seventh Edition
© 2012 W. H. Freeman and Company



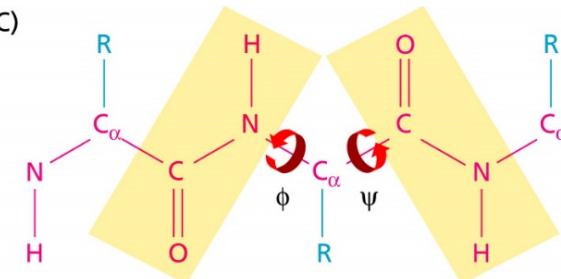
hydrogen bonding



rotation around
single bond

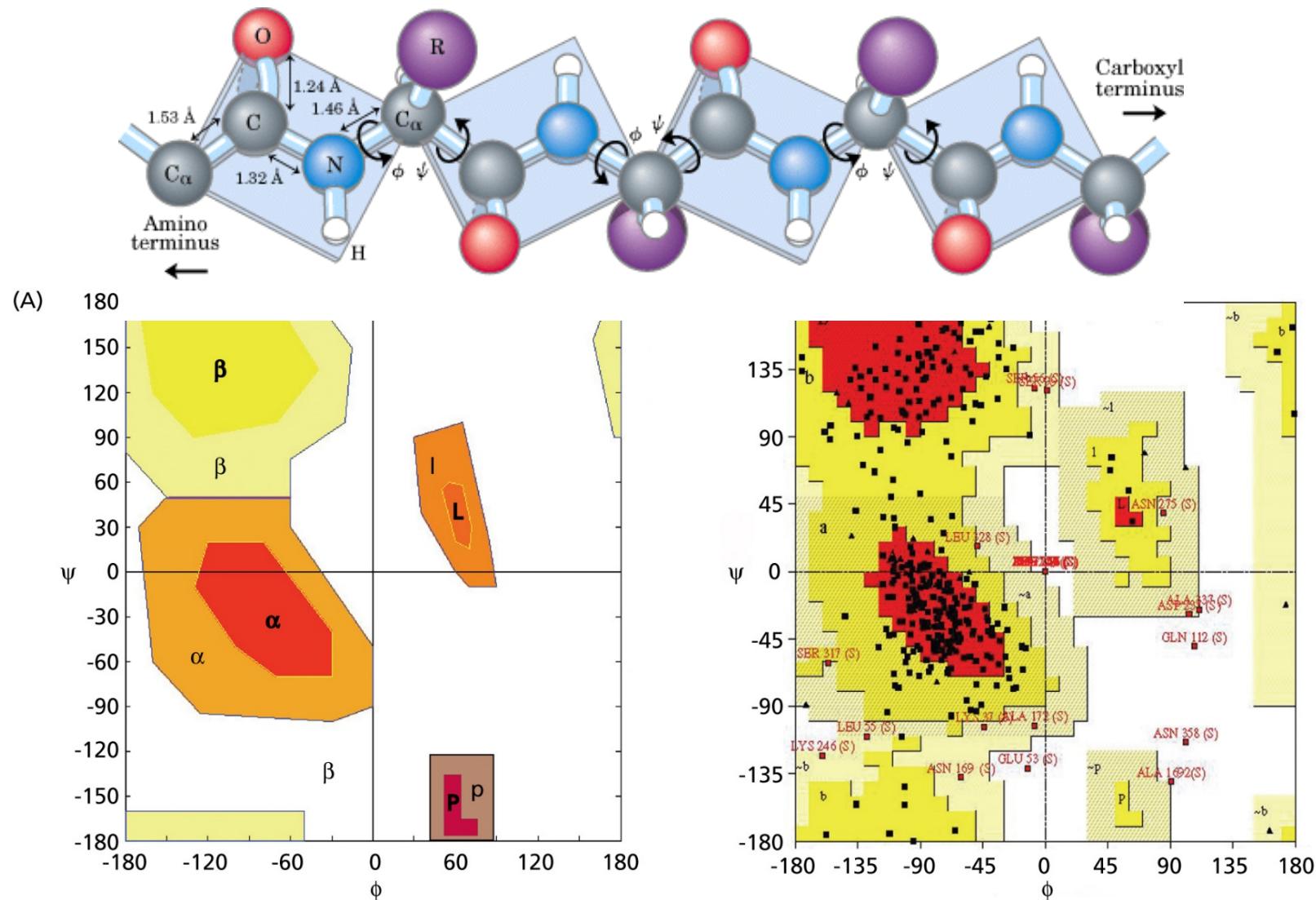


peptide bond is planar due to
the electron resonance (partial
double bond character)



two possible rotations around Cα-N
bond (ϕ) and C α -C bond (ψ)

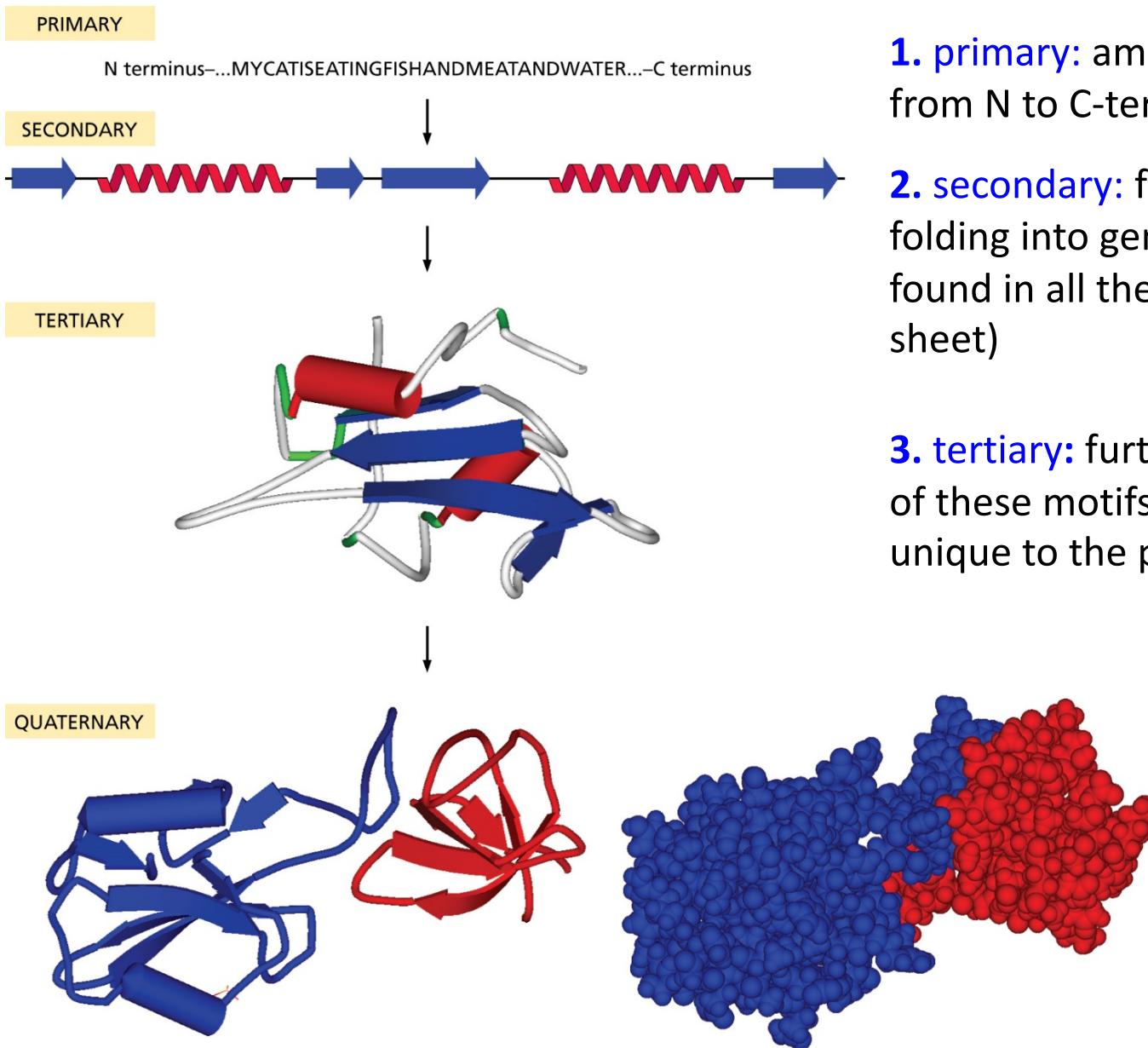
Ramachandran plot



existence of the preferred orientations between the peptide bonds

protein structure

- there are four levels of protein structure:



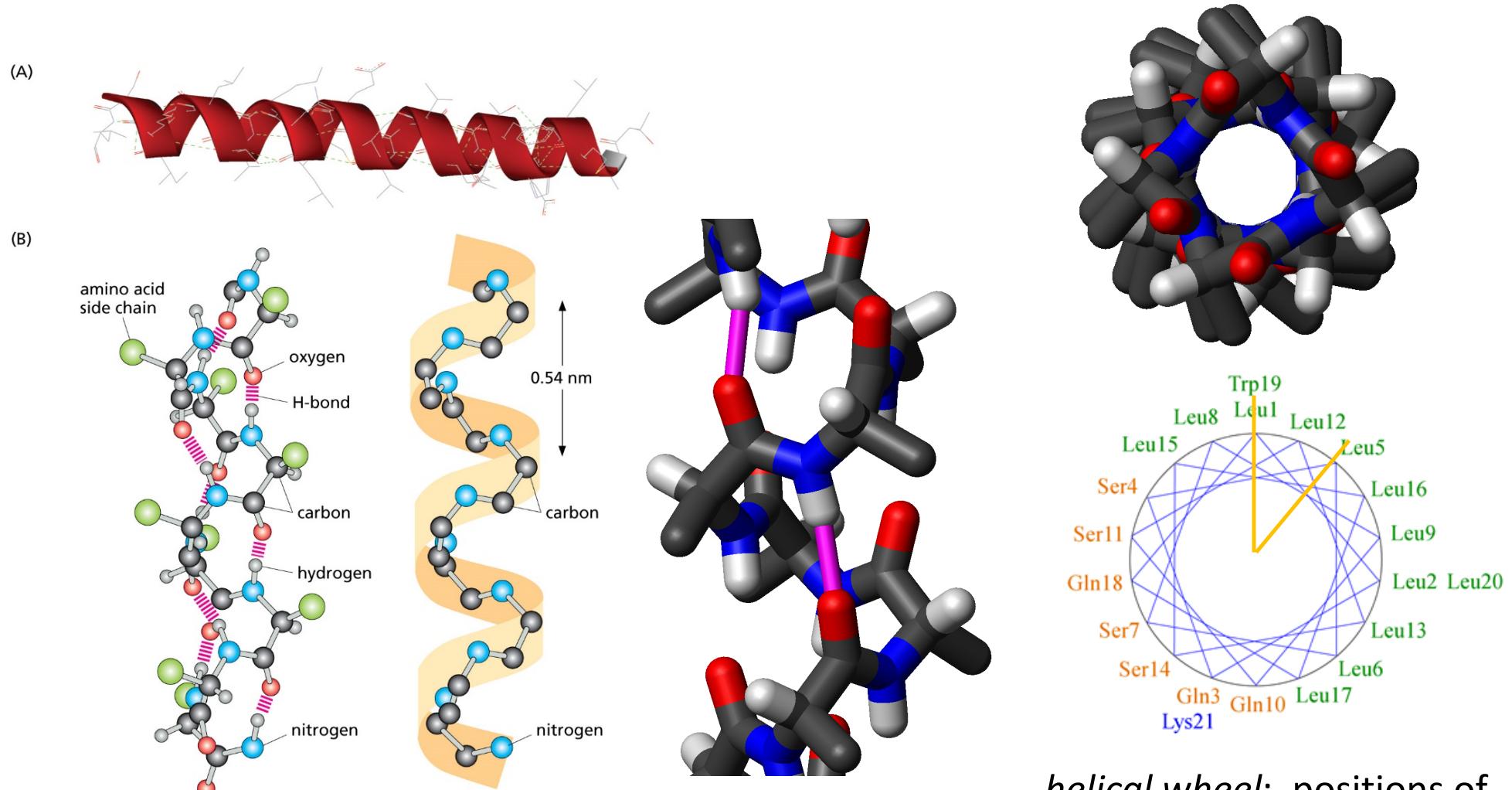
1. primary: amino acid sequence read from N to C-terminus

2. secondary: first level of protein folding into generic structural motifs found in all the proteins (α -helix, β -sheet)

3. tertiary: further folding and packing of these motifs into a 3D conformation unique to the protein

4. quaternary: assembly of multimeric proteins consisting of more than one (folded) protein chain – oligomers

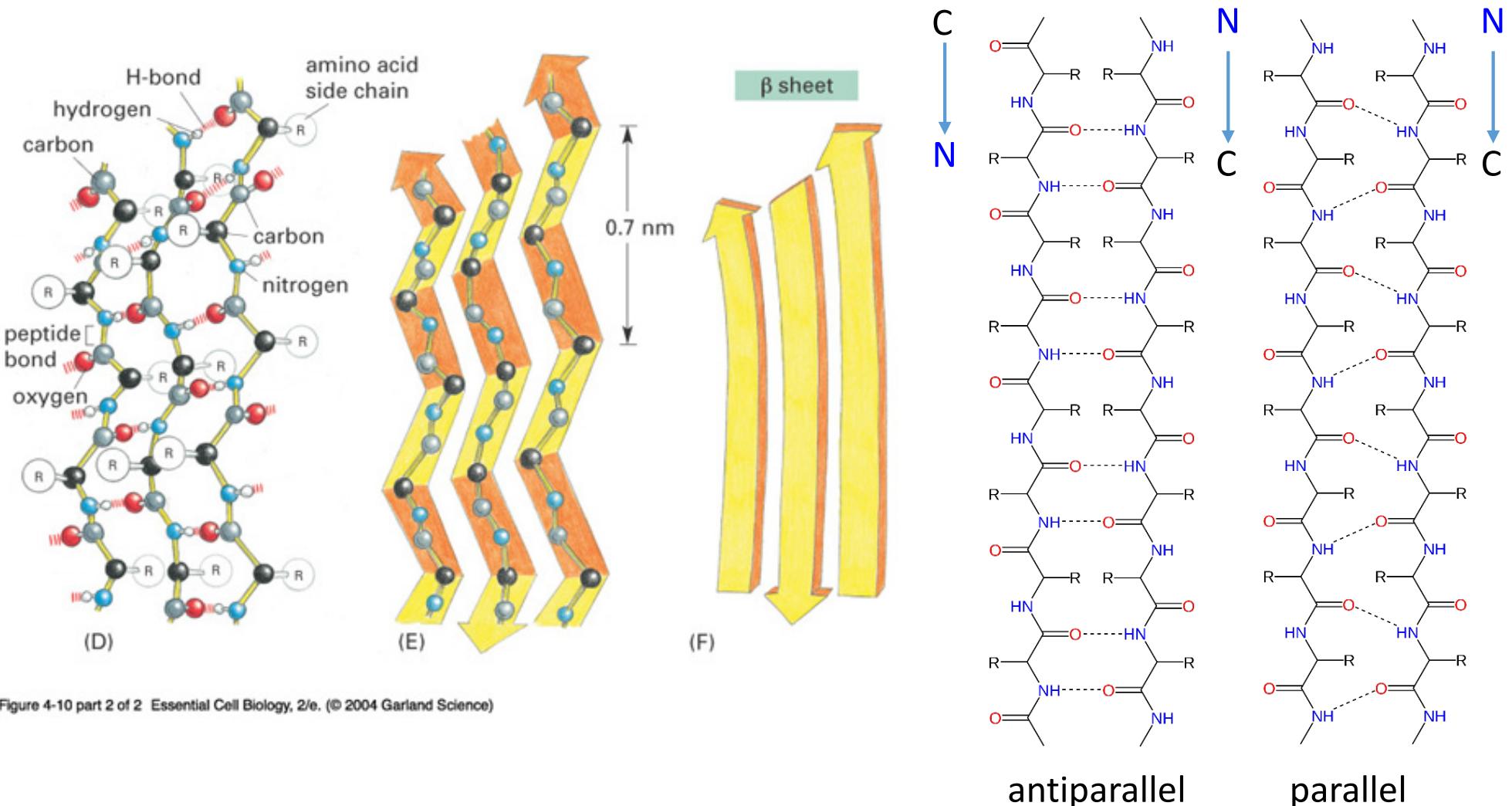
secondary structure: α -helix



C=O and N-H in the backbone make H-bonds 3.6 amino acid (residues) per turn \rightarrow O(i) hydrogen bonds to N(i+4)

helical wheel: positions of the side chains along regular helix (40° angle between i th and $i+4$ residues – orange)

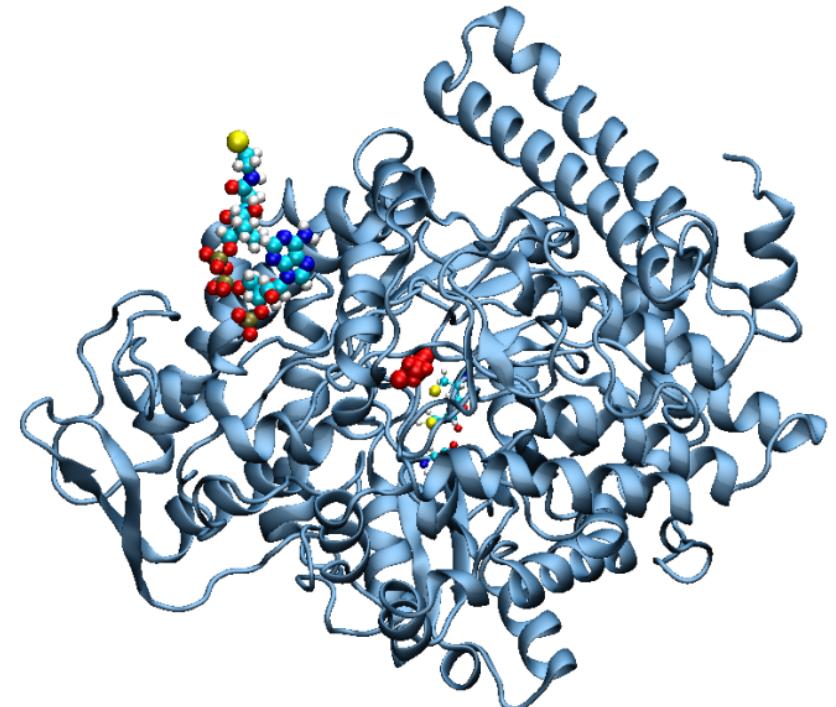
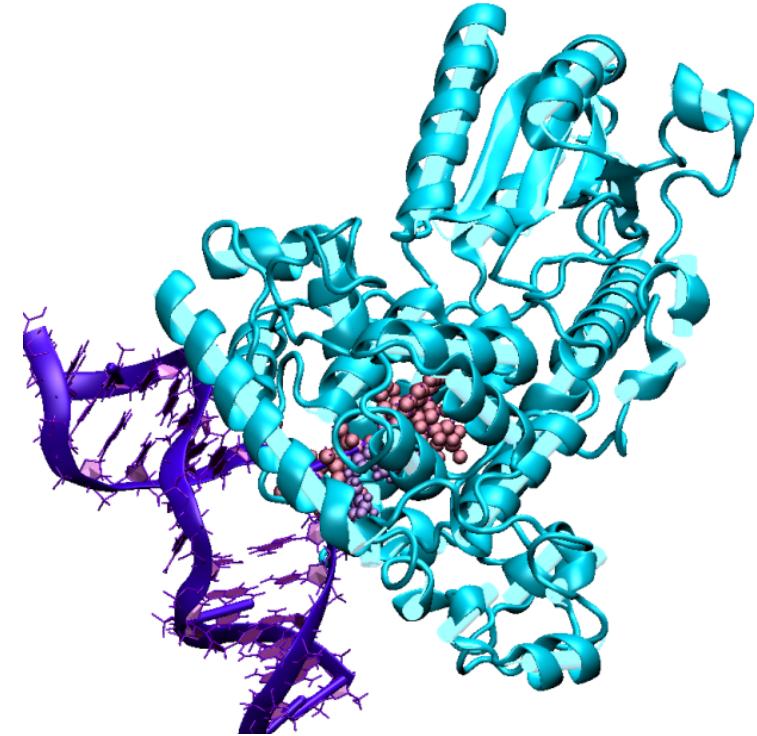
secondary structure: β -sheet



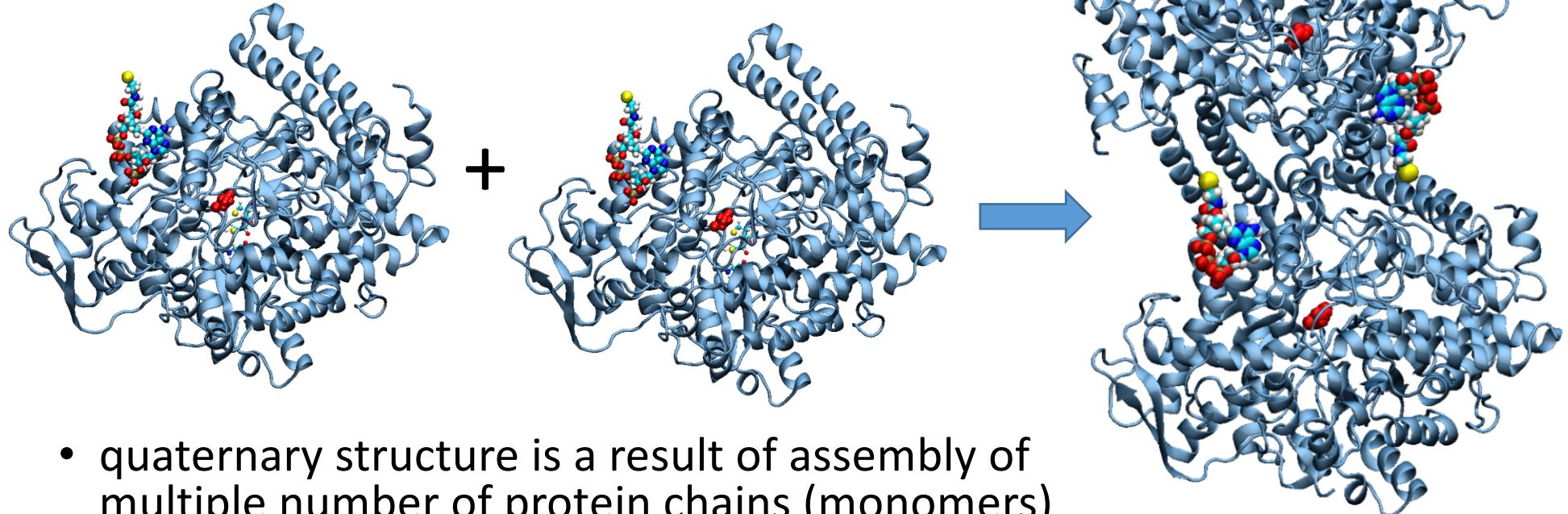
- perfect β -sheets contain all-parallel or all-antiparallel strands, but in nature they come as mixture of both

tertiary structure

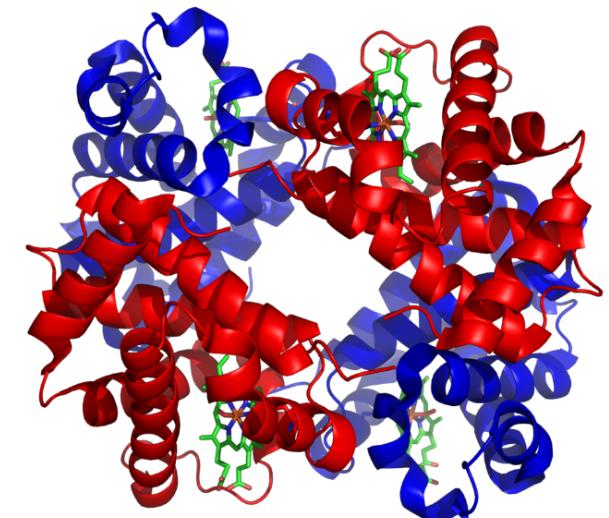
- tertiary structure is a product of packing various combinations of secondary structures (domains)
- we are still unable to predict the tertiary structure of the protein based only on its sequence → *homology modelling*
- *homologous* proteins have a common ancestor → similarity in structure and function (not always)
- sometimes different sequences result with similar structural motifs (convergent evolution)



quaternary structure



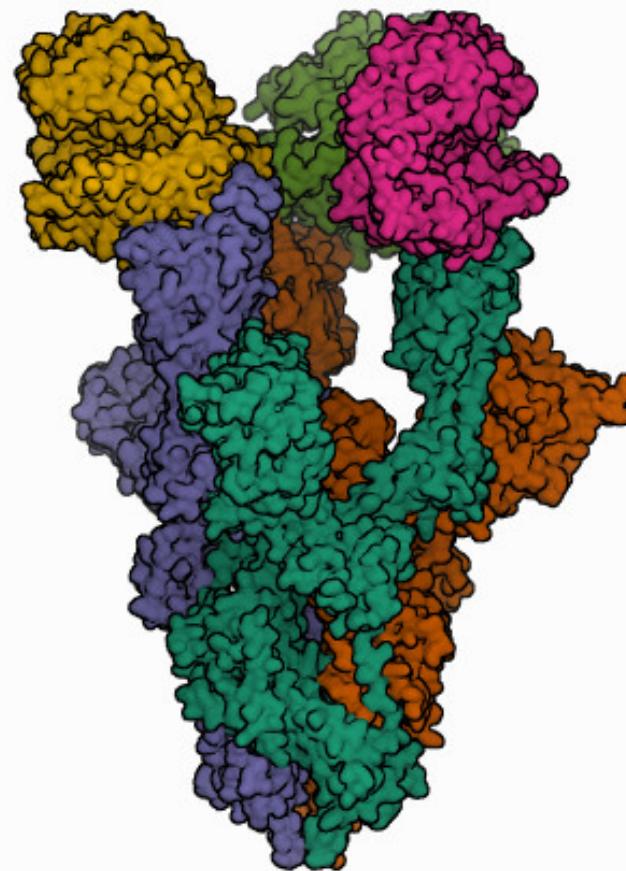
- quaternary structure is a result of assembly of multiple number of protein chains (monomers) into large supramolecular complexes → *oligomers*
- the subunits can be identical (for example, homodimers) or they can be different (heterodimers)
- tetrameric haemoglobin (right) with two α and two β subunits



Coronavirus update: 3D structures

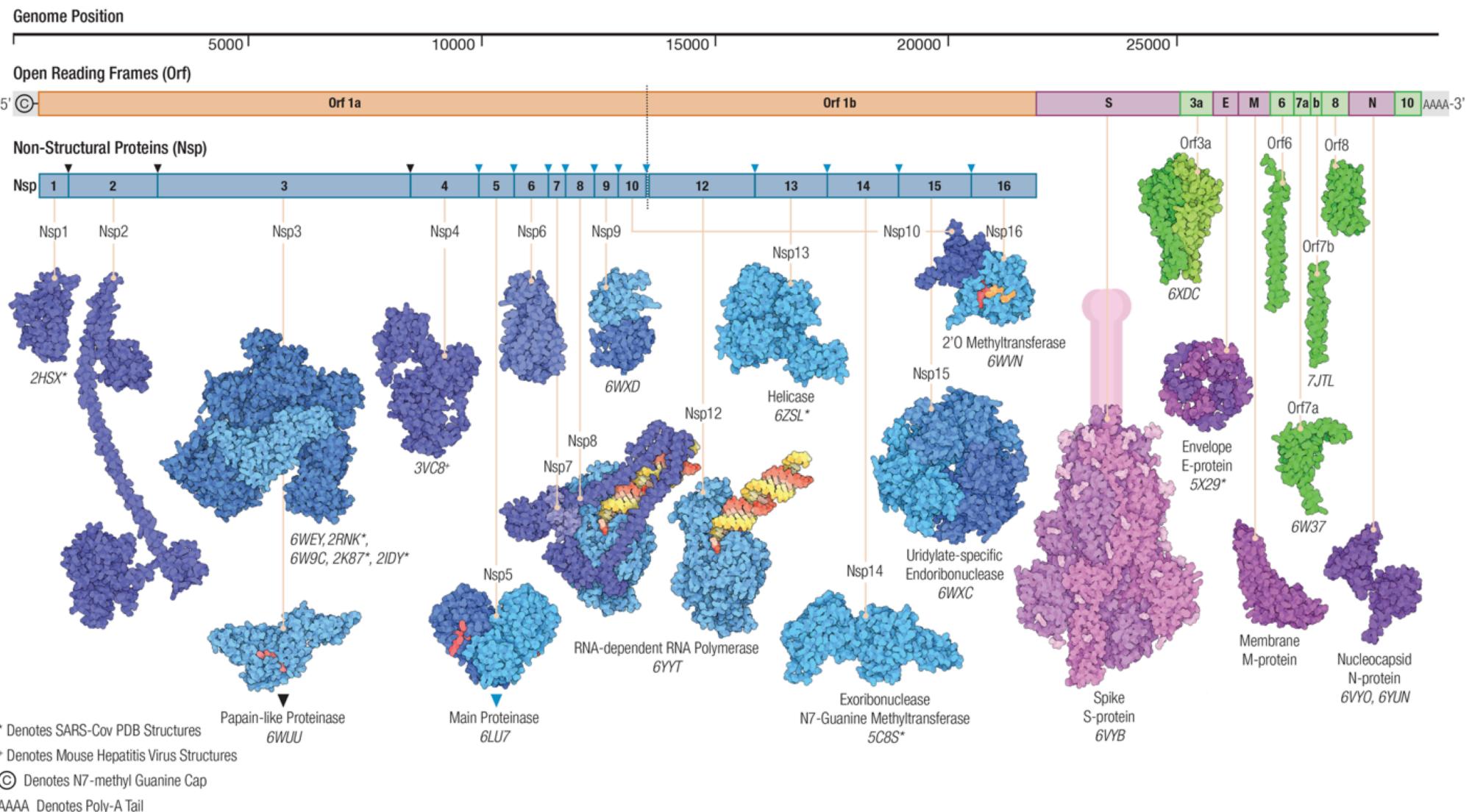


ACE2 bound to spike
protein (binding domain)



3 x ACE2 bound to spike
protein (full protein)

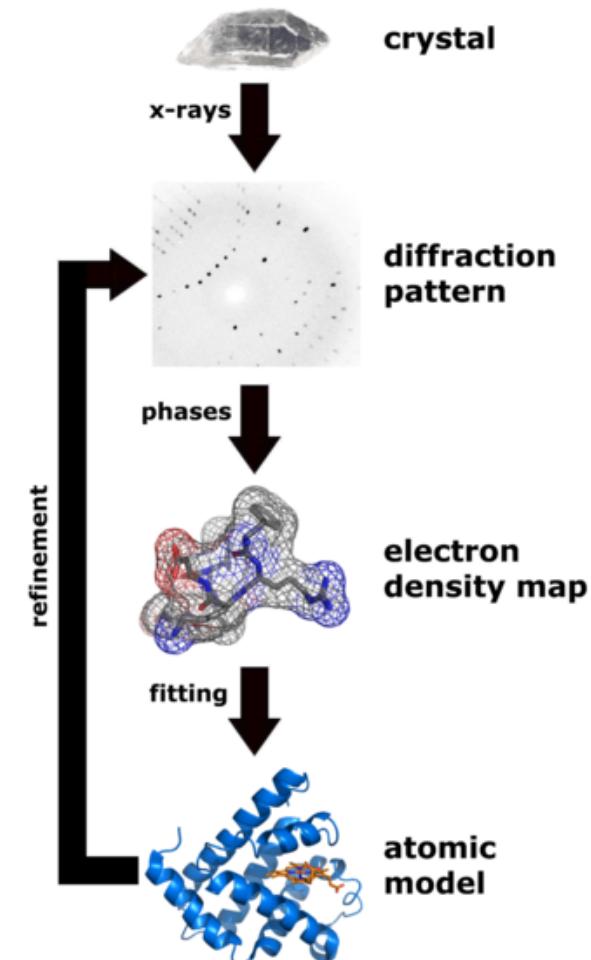
Coronavirus update: 3D structures



Lubin et al., BioRxiv. Evolution of the SARS-CoV-2 proteome in three dimensions (3D) during the first six months of the COVID-19 pandemic
doi: <https://doi.org/10.1101/2020.12.01.406637>

experimental methods for structure resolving

- X-ray crystallography → diffraction of X-rays on the electron densities of the protein in the crystal form → positions of heavy atoms (no hydrogens)
 - difficulties: obtaining protein crystals (especially for hydrophobic proteins), expensive
- nuclear magnetic resonance (NMR) spectroscopy → nuclei with spin (magnetic moment) can absorb radiowaves when placed in a magnetic field
 - NMR can be done in a solution (no crystals required), but the resolution of the structure is lower
- **State-of-the-art 2020:** Cryo-Electron Microscopy
 - Can resolve large multimeric structures



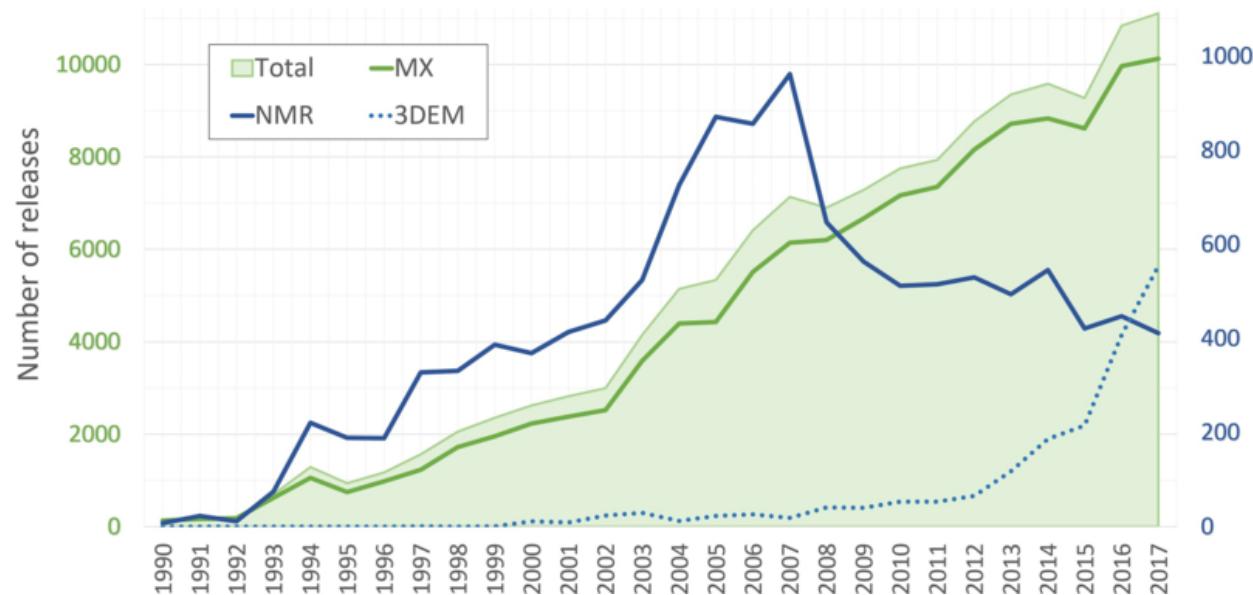
protein databases

- UniProt → resource of protein sequences

<http://www.uniprot.org/>

- Protein Data Bank (PDB) → deposition of the solved protein structures

<http://www.rcsb.org/pdb/home/home.do>

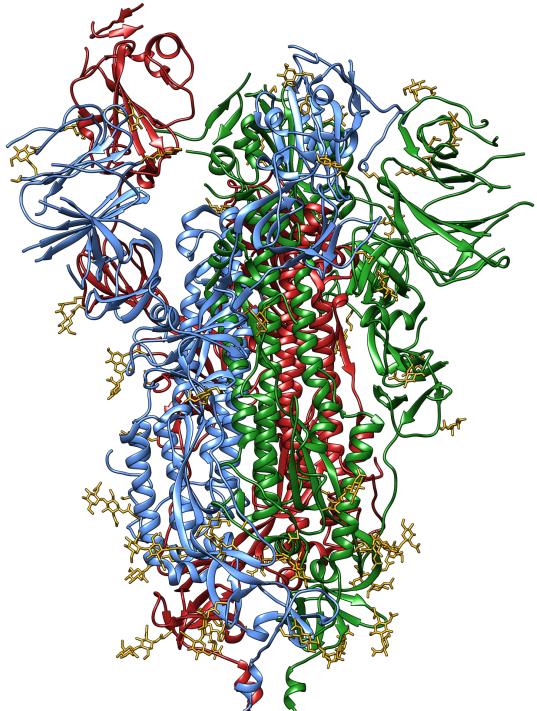


174,000 structures available in 2021 (77,000 in 2011)

Key learning outcomes

- Understand the genetic code and how it works in translation.
- Recognize the key RNA and protein components of the translation process.
- Appreciate how ribosomes work to synthesize polypeptides.
- Recognize the physio-chemical properties of amino acids that are important for protein function.
- Understand the difference between secondary, tertiary and quaternary protein structure.

Refer to Week 1-2 Study Guide document



Thanks for your
attention!

Scott Beatson
(s.beatson@uq.edu.au)
<http://beatsonlab.com/>

Molecular clamp vaccine design for Covid-19:
<https://www.uq.edu.au/news/article/2020/02/significant-step'-covid-19-vaccine-quest>

Top: Cryoelectron microscopy structure of 2019-nCoV (COVID-19) spike glycoprotein (PDB ID [6vsb](#)) Bottom: David S Goodsell, Coronavirus illustration ([doi: 10.2210/rcsb_pdb/goodsell-gallery-019](https://doi.org/10.2210/rcsb_pdb/goodsell-gallery-019)).