

Introduction to molecular biology: Protein synthesis and structure



A/Prof Scott Beatson

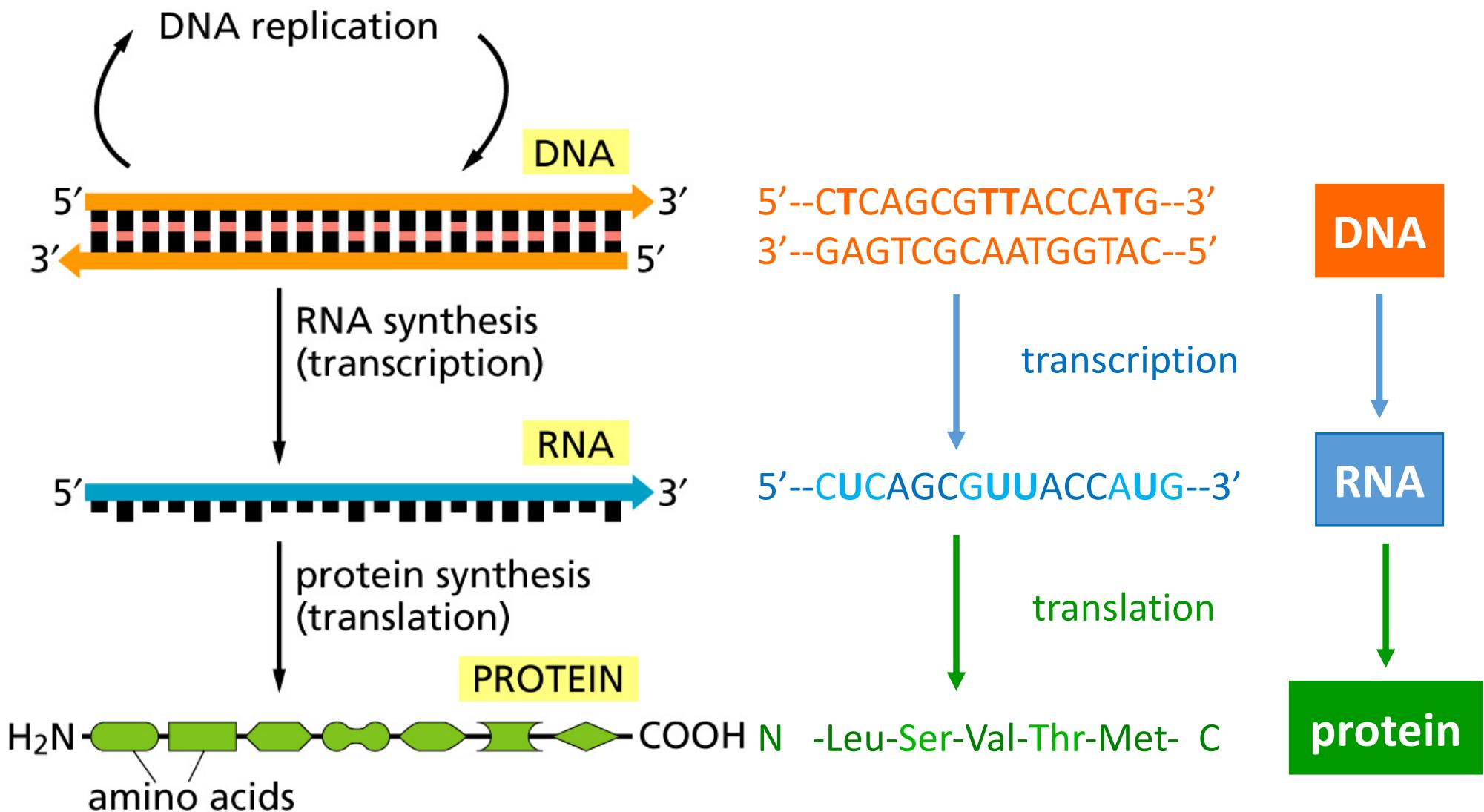


s.beatson@uq.edu.au



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 —

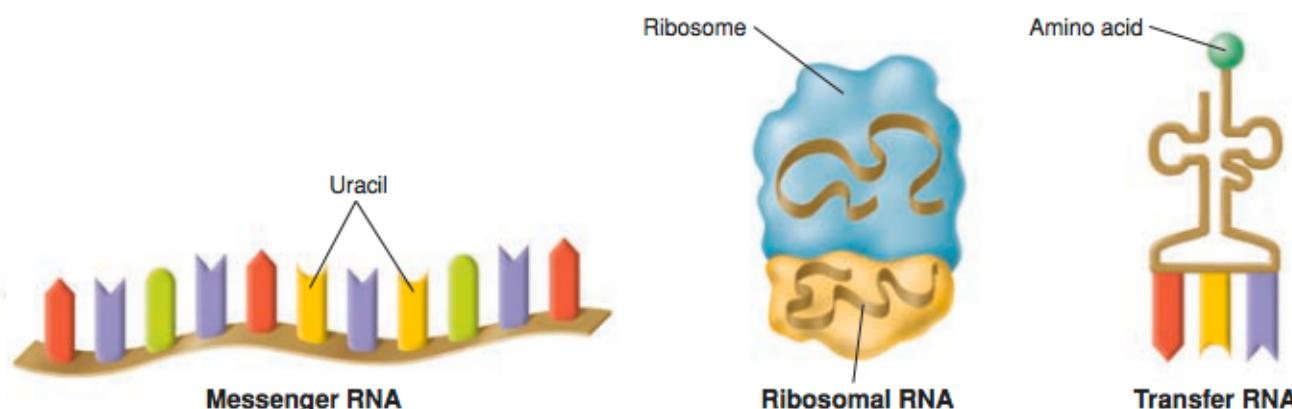
#2 5' — C U C A G C G U U A C C A U — 3'
— Ser — Ala — Leu — Pro —

#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.

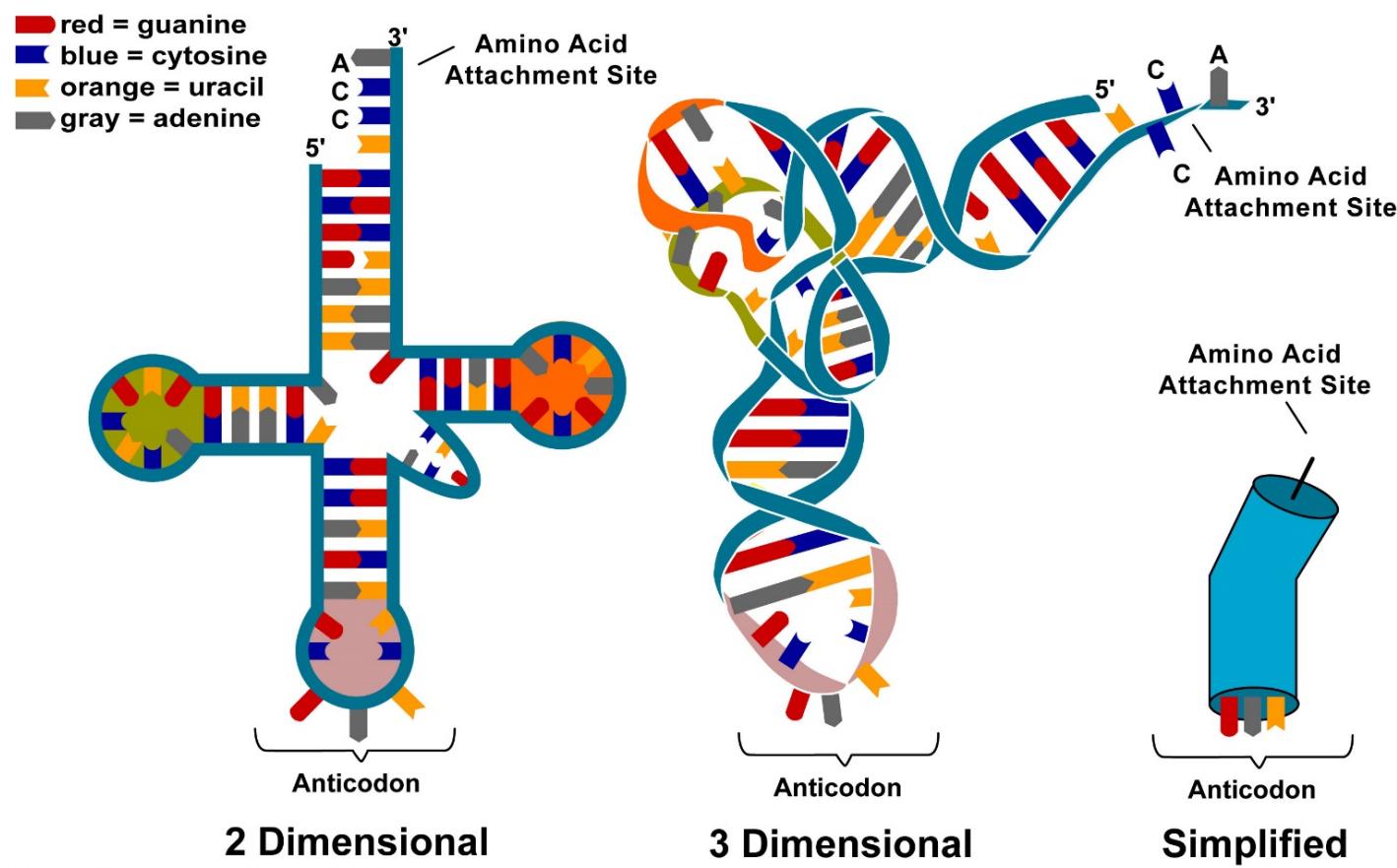
Video: protein synthesis



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

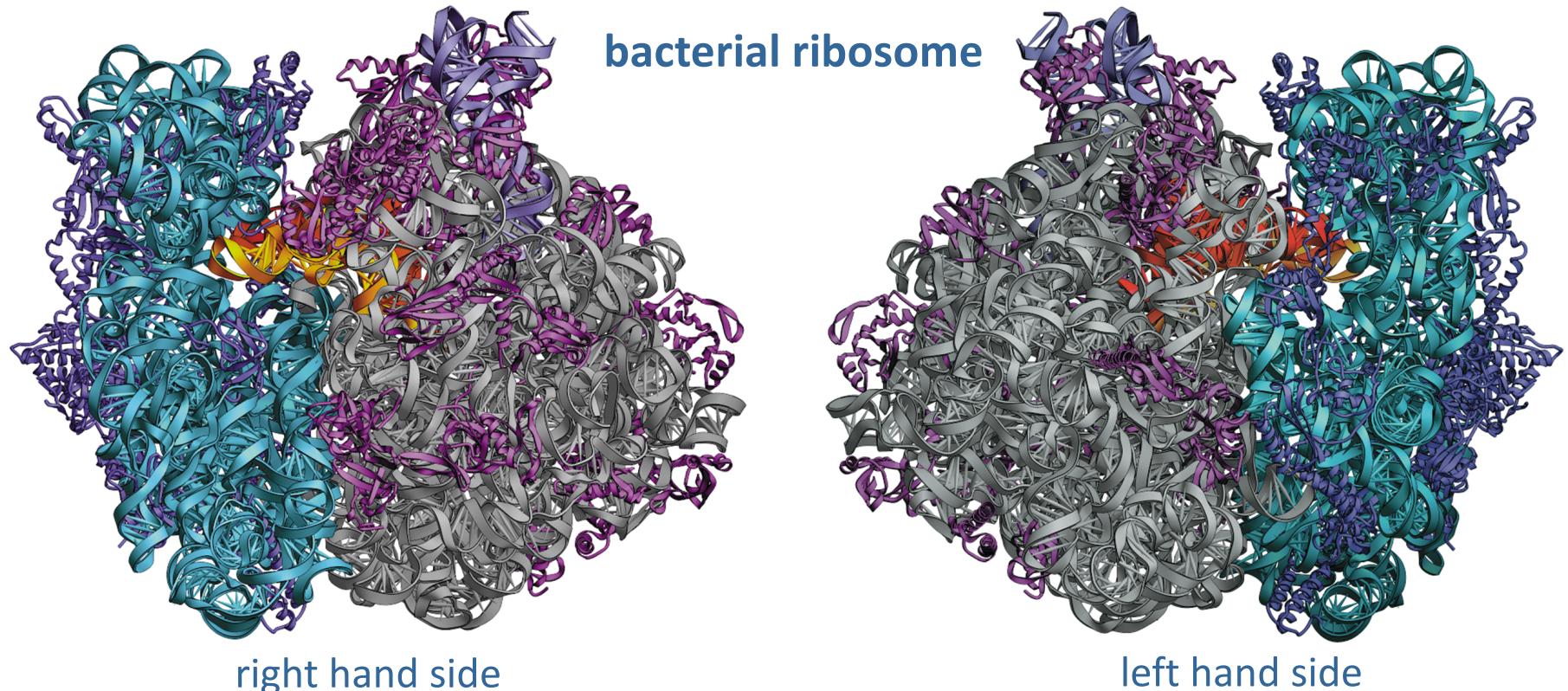
transfer RNA (tRNA)

- 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)

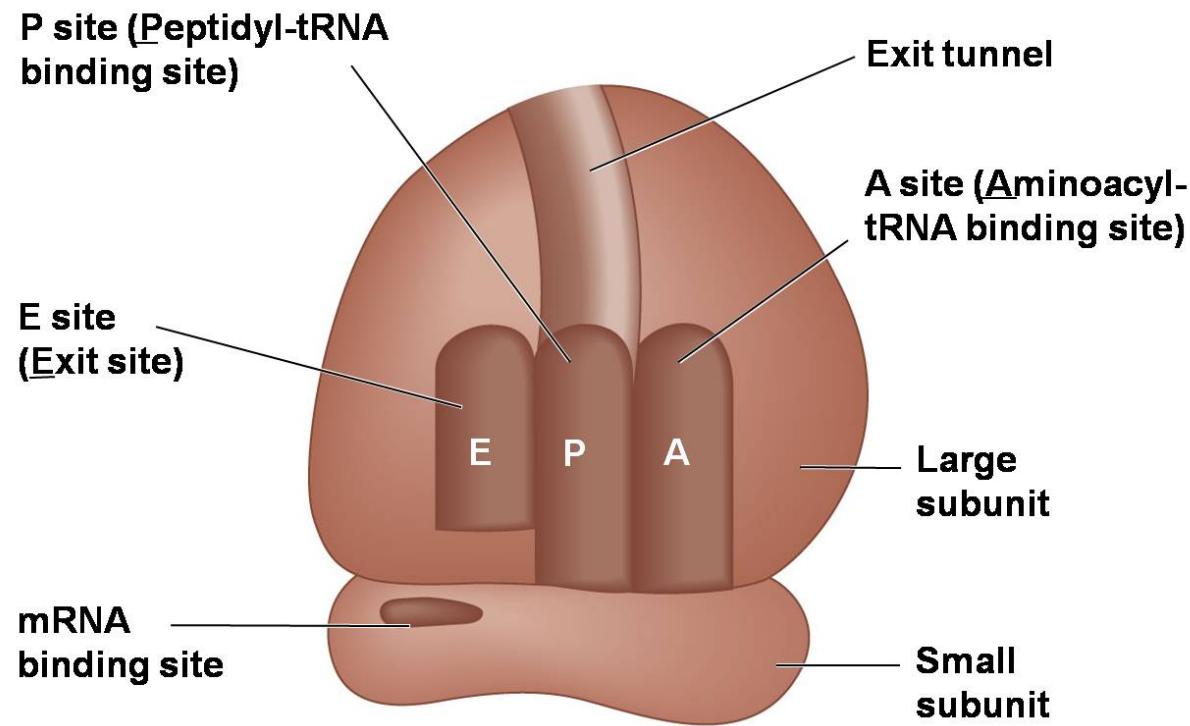


Ribosomes: protein synthesis machines

- 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

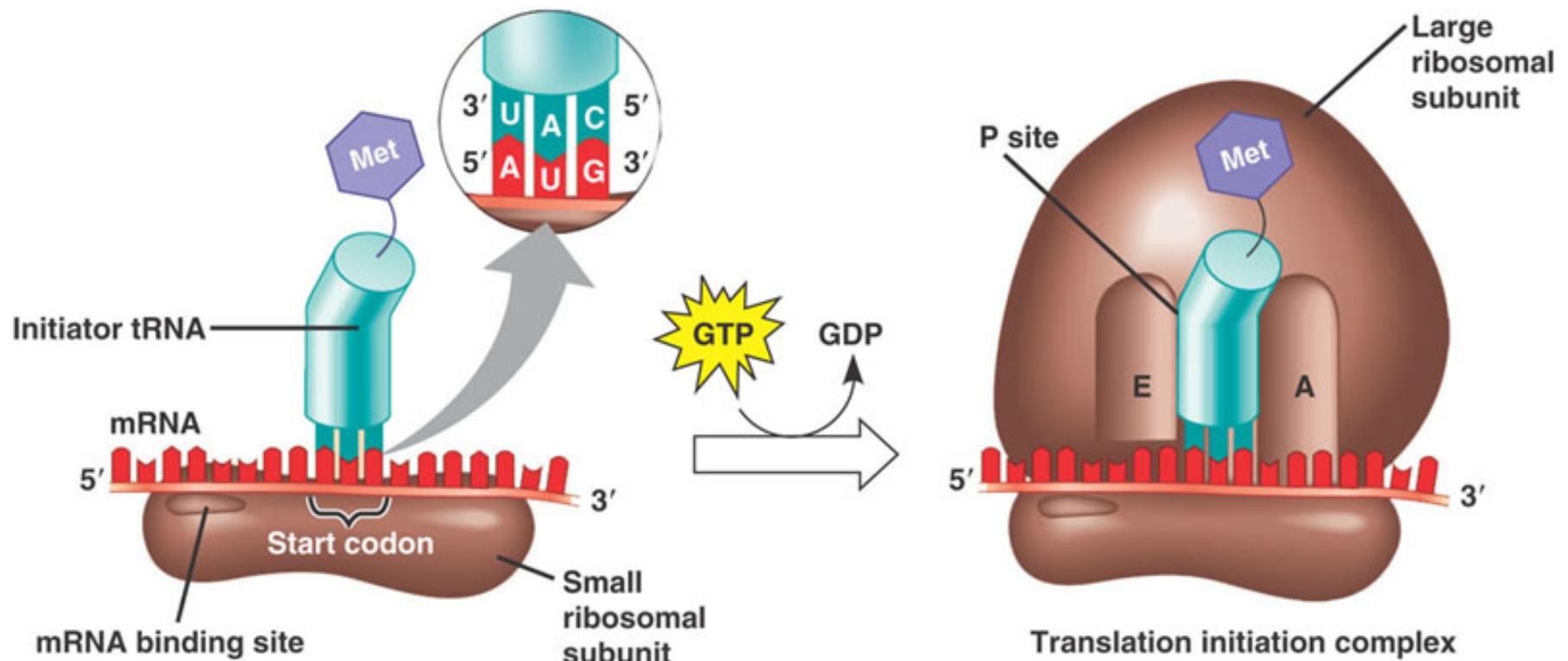


ribosome



- 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

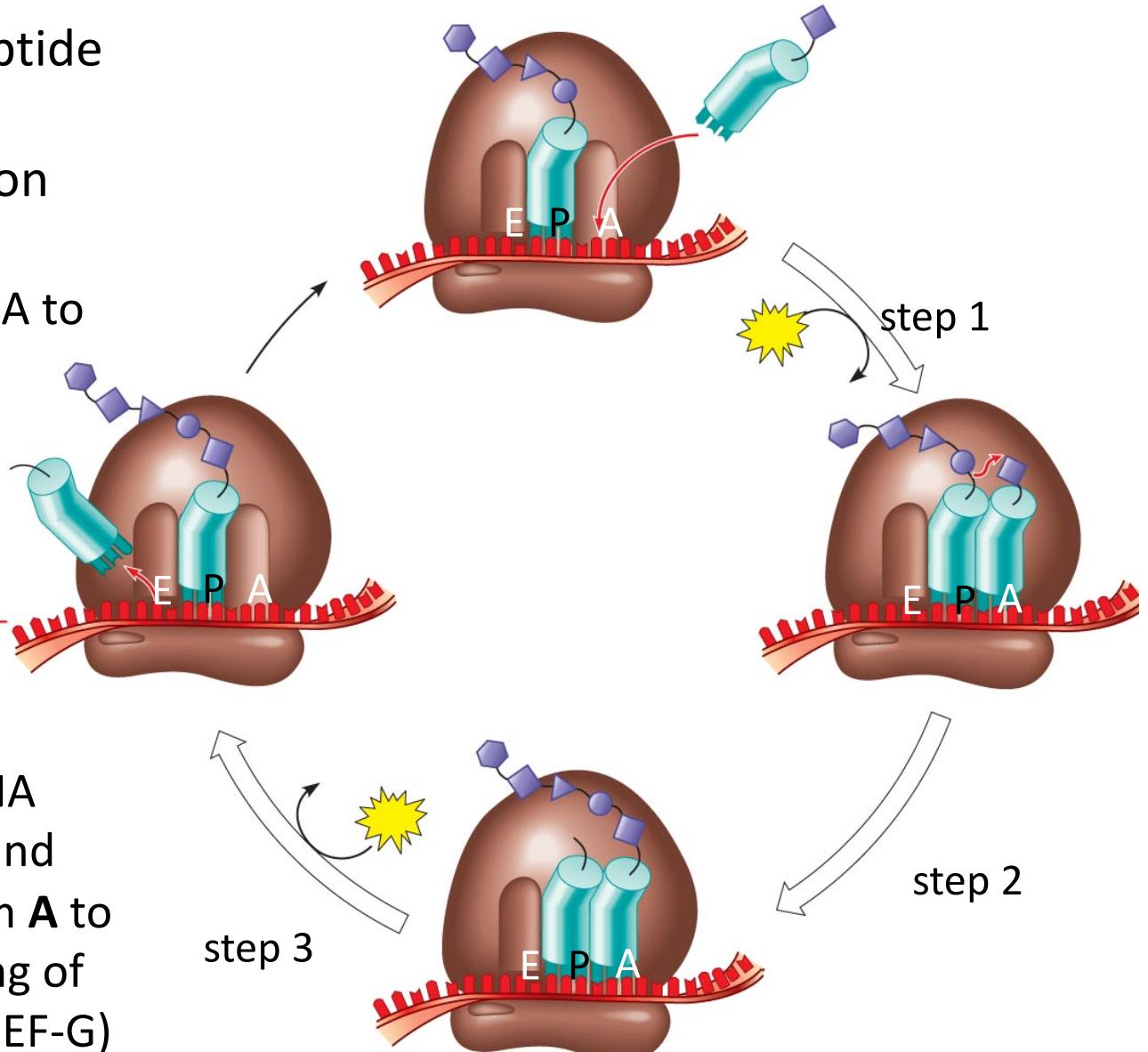
2. translation: elongation

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

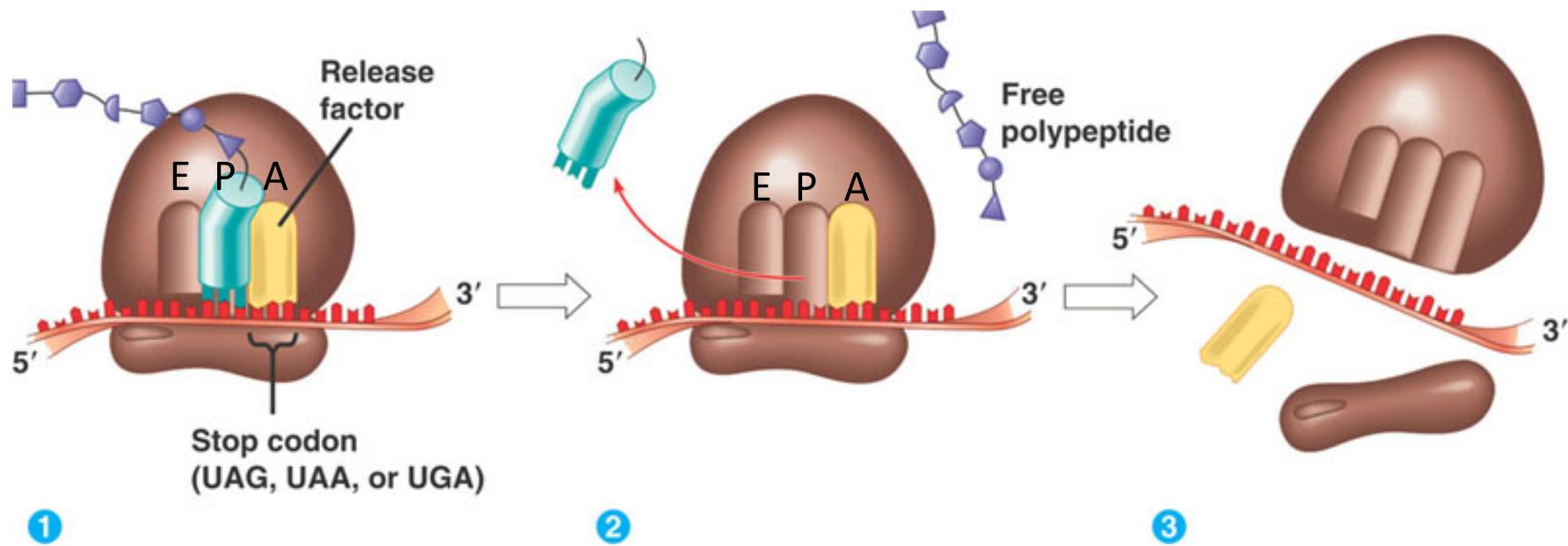
1. ***binding*** of aminoacid-tRNA to **A site (EF-Tu)**

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)

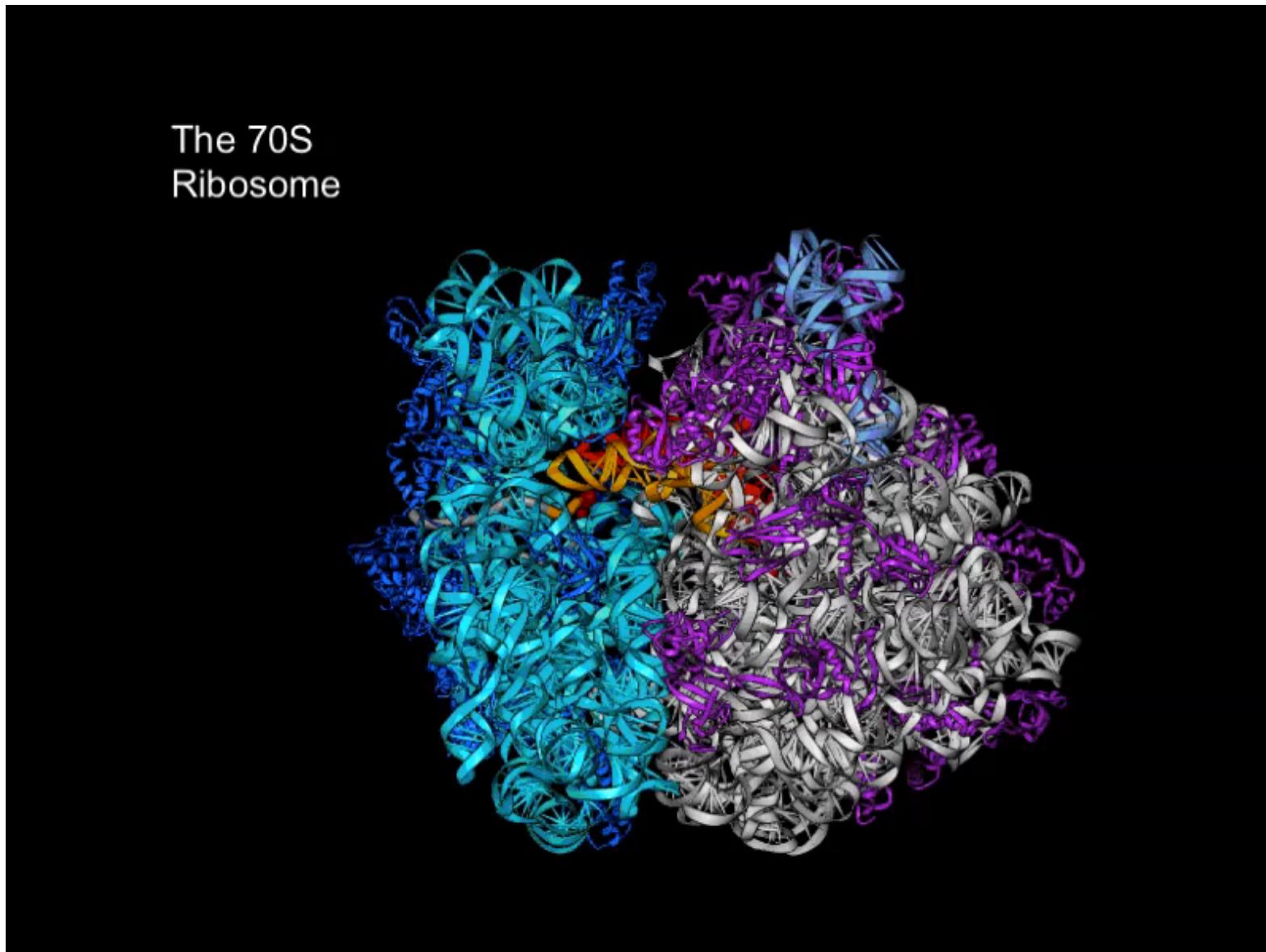


3. translation: termination

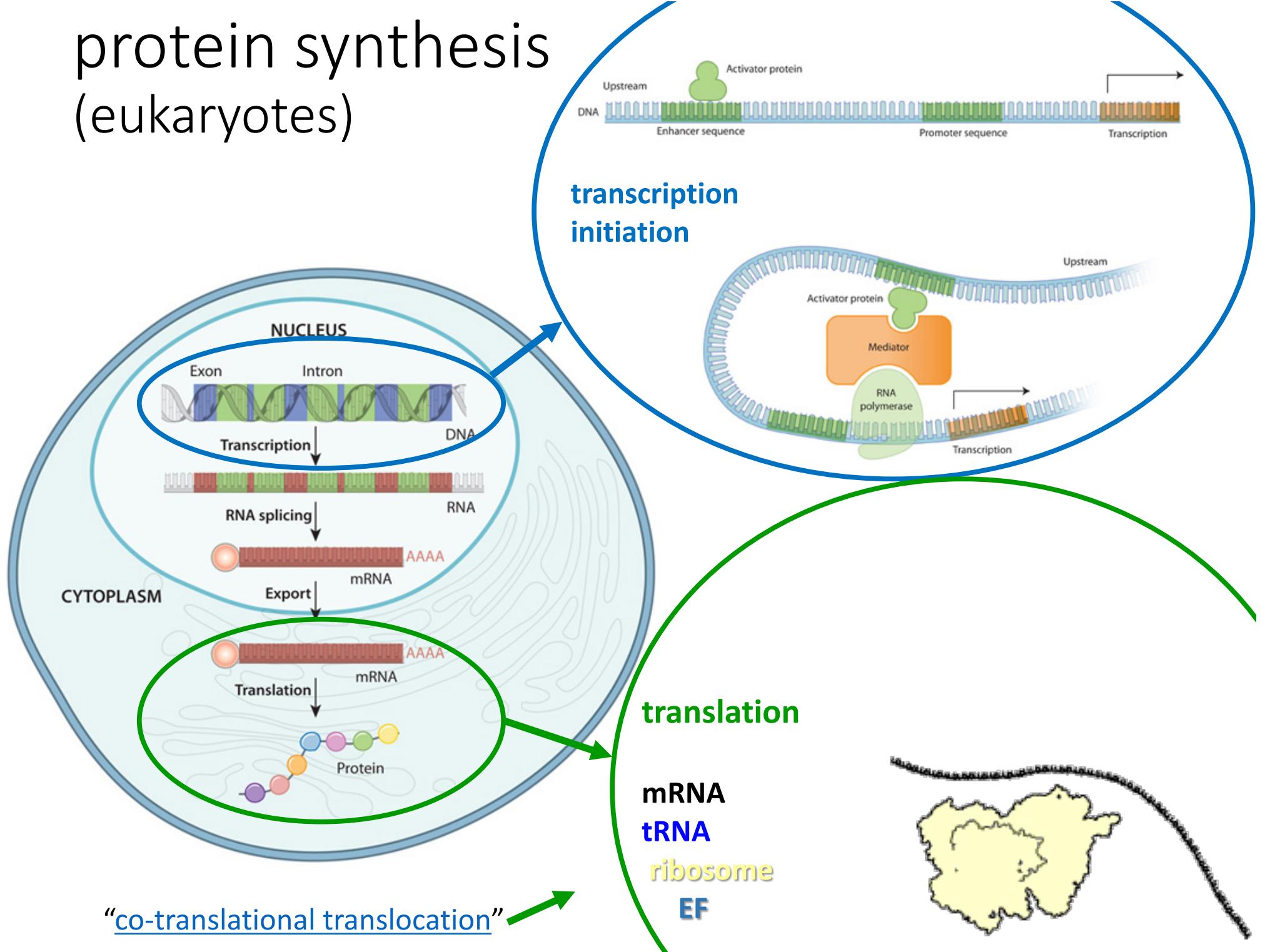


- termination of translation process after finding a stop codon on mRNA → binding of the release factor (RF) at the site A
- cleavage and release of the new polypeptide (protein) from the P site, followed by disassembling of the ribosome complex

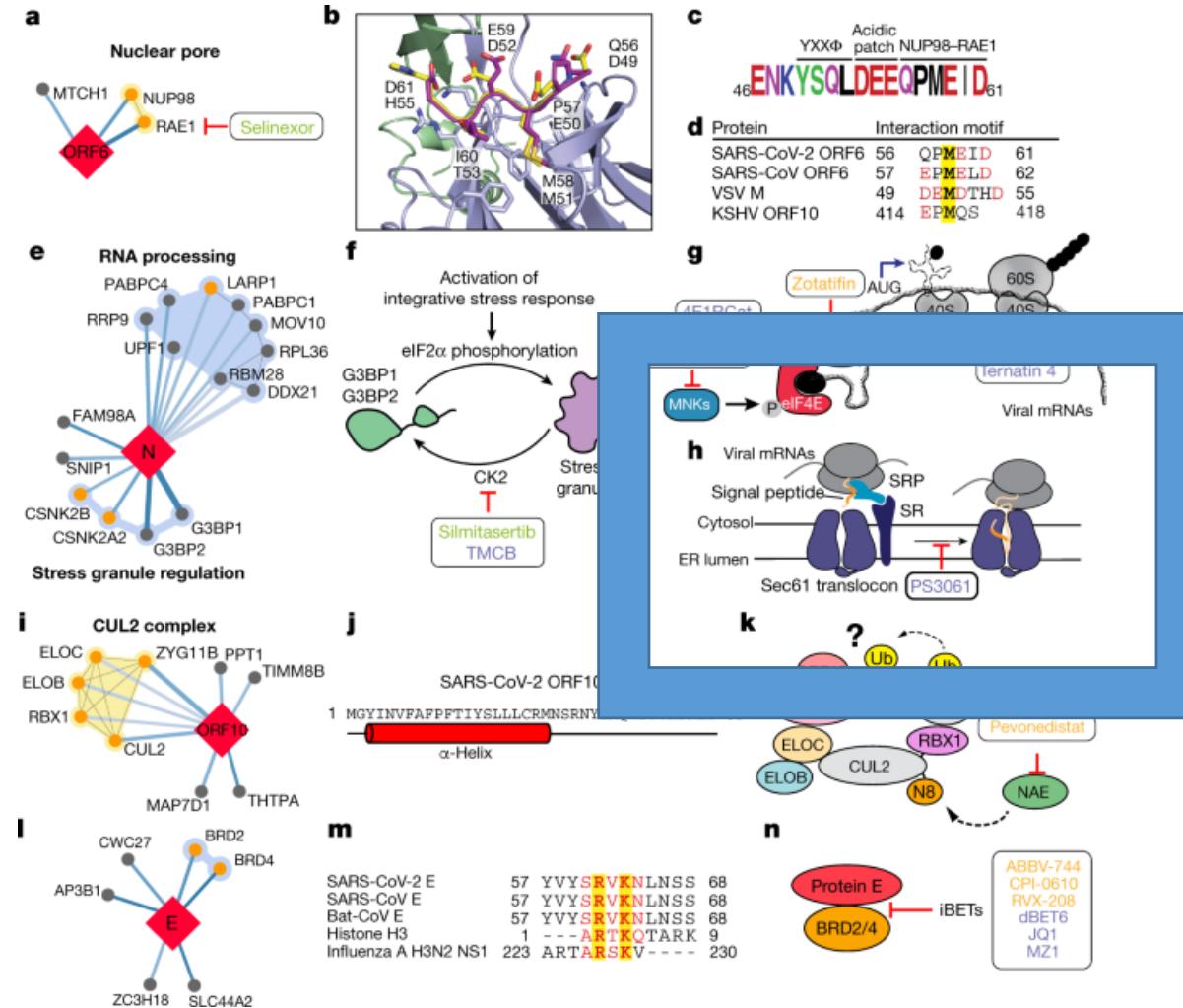
elongation & termination



protein synthesis (eukaryotes)



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



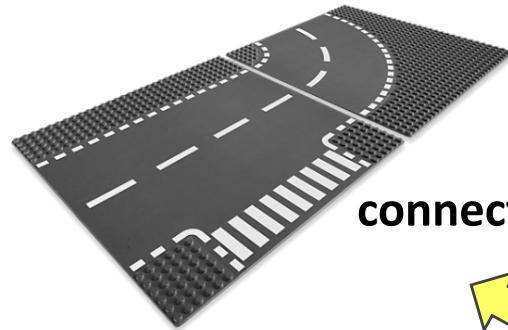
Gordon, D.E., Jang, G.M., Bouhaddou, M. et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 583, 459–468 (2020).
<https://doi.org/10.1038/s41586-020-2286-9>

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



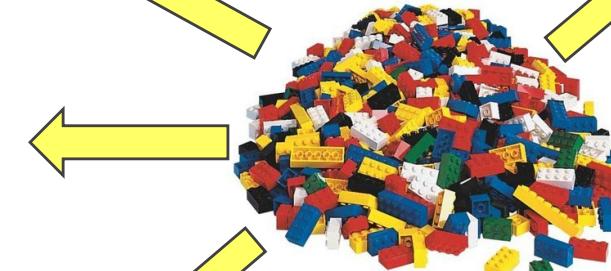
transport



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



**synthesis of new
compounds**



signalling



motion



assembly

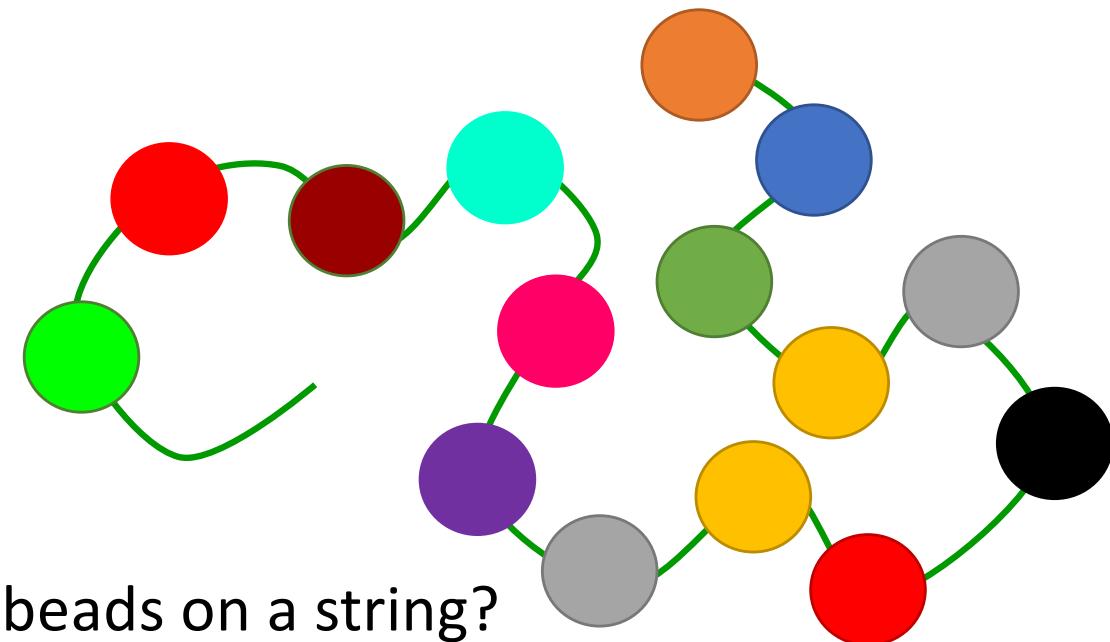


storage

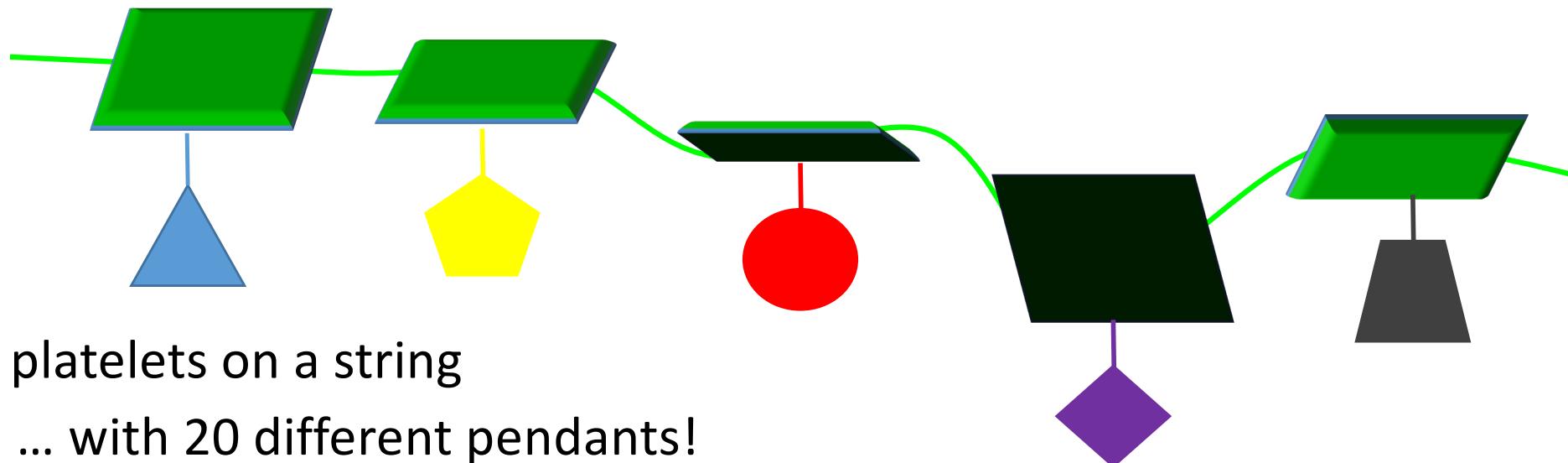
proteins



lego pieces?



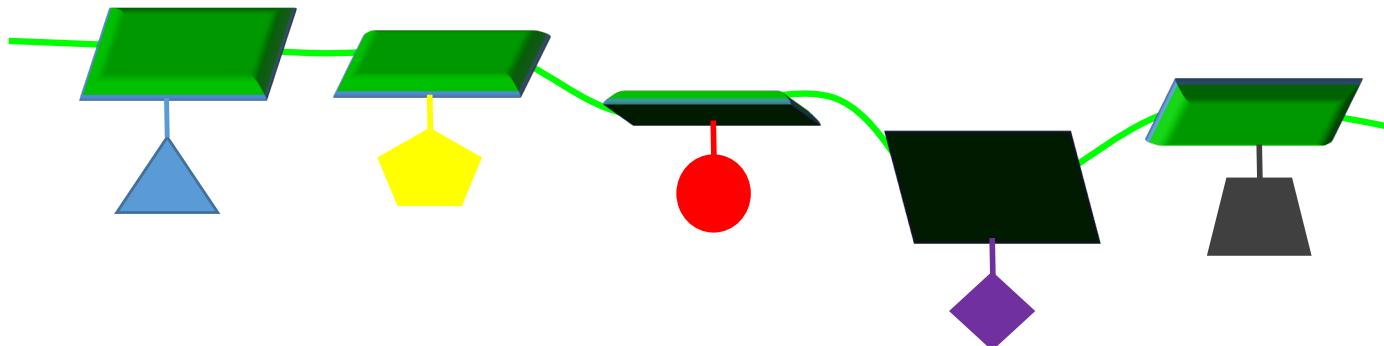
beads on a string?



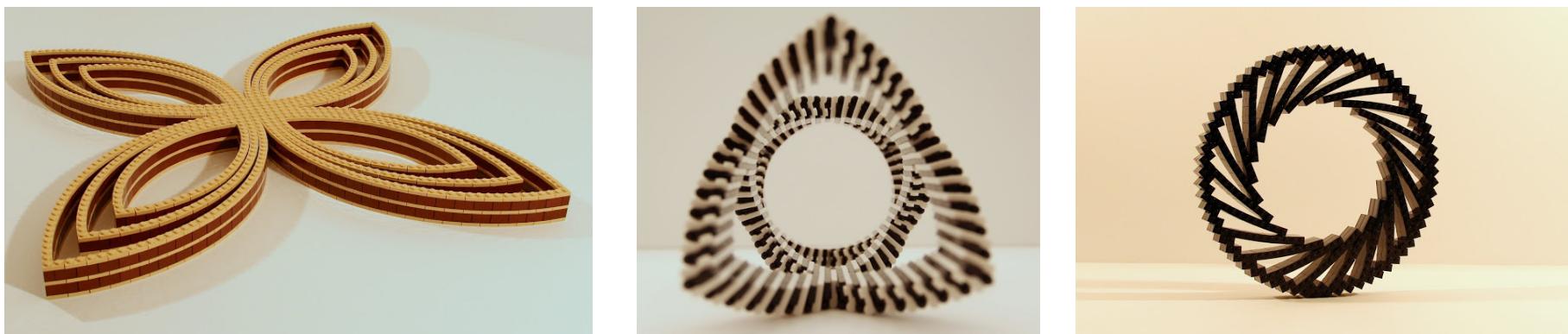
platelets on a string

... 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

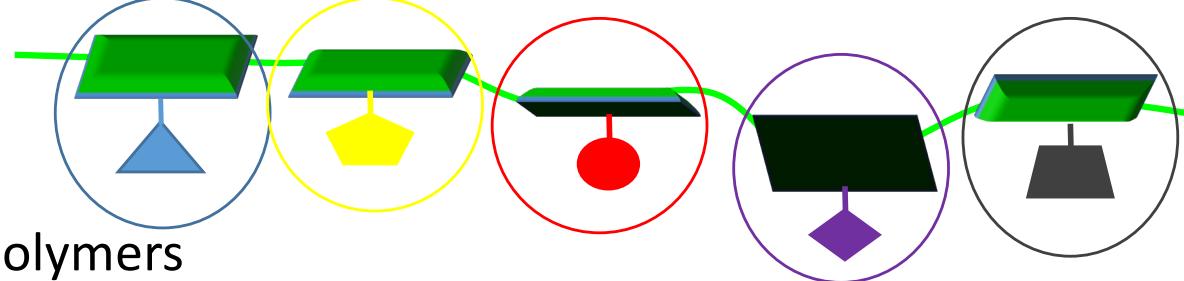


folded
(native)



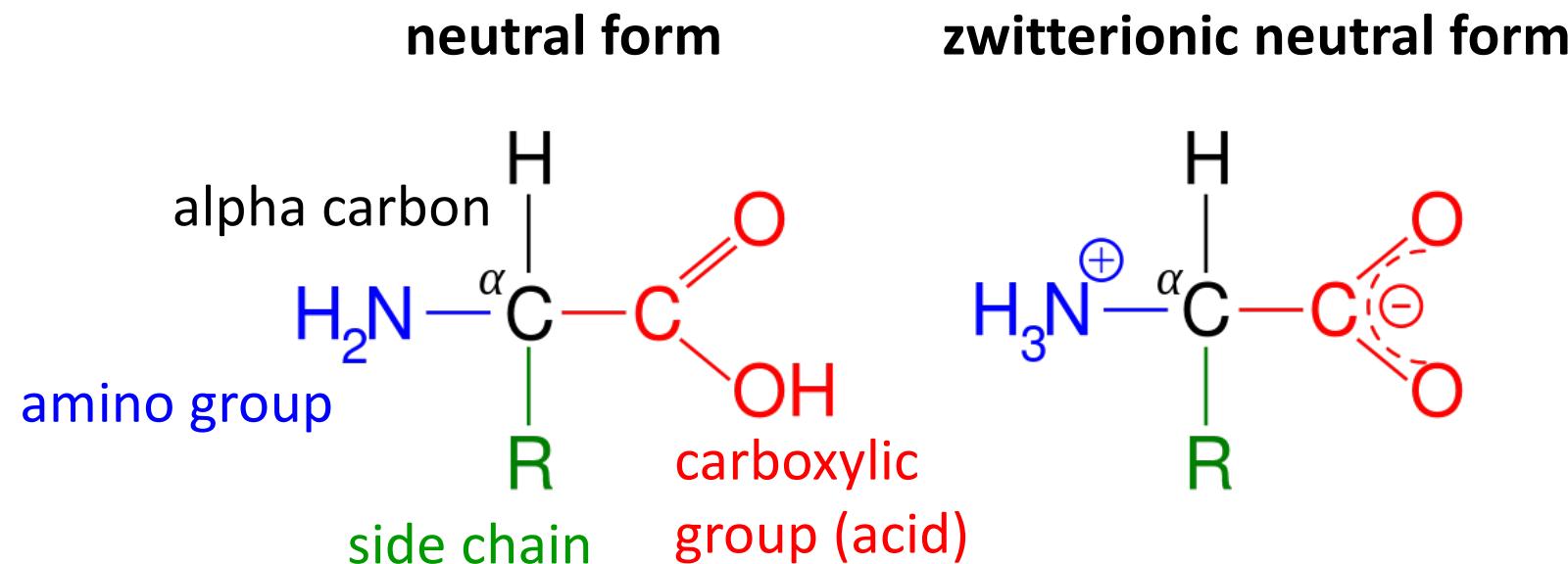
- interactions within protein and with its environment → the pendants on our long chain have some preferences for other pendants and general surrounding, platelets also prefer some angles over other...

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

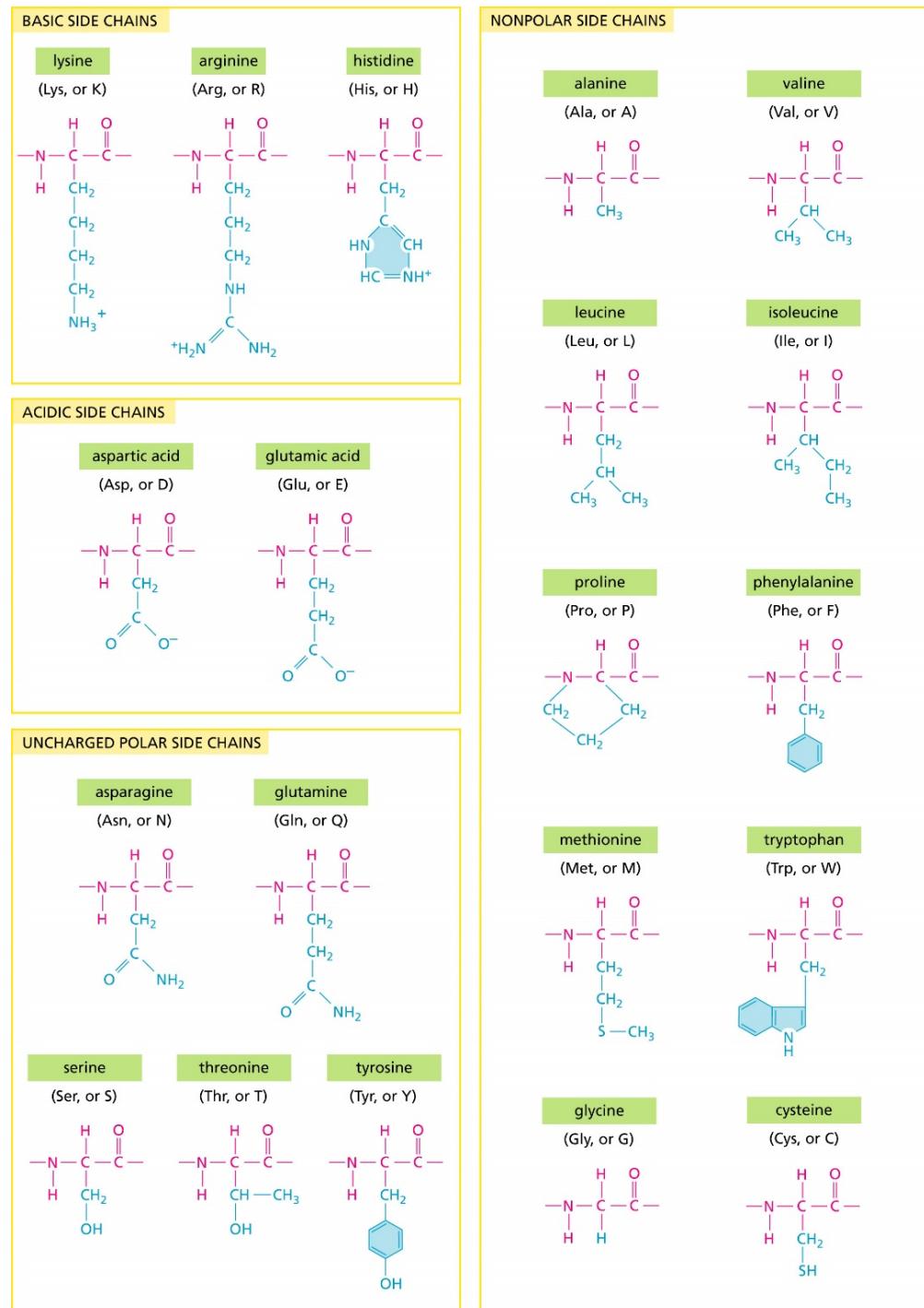
- amino acids can be classified in a few different ways
- E.g. classification into 4 groups based on the chemical properties of their side chain:
 - basic (positively charged);
 - acidic (negatively charged);
 - polar (uncharged);
 - nonpolar (uncharged)

base: proton acceptor

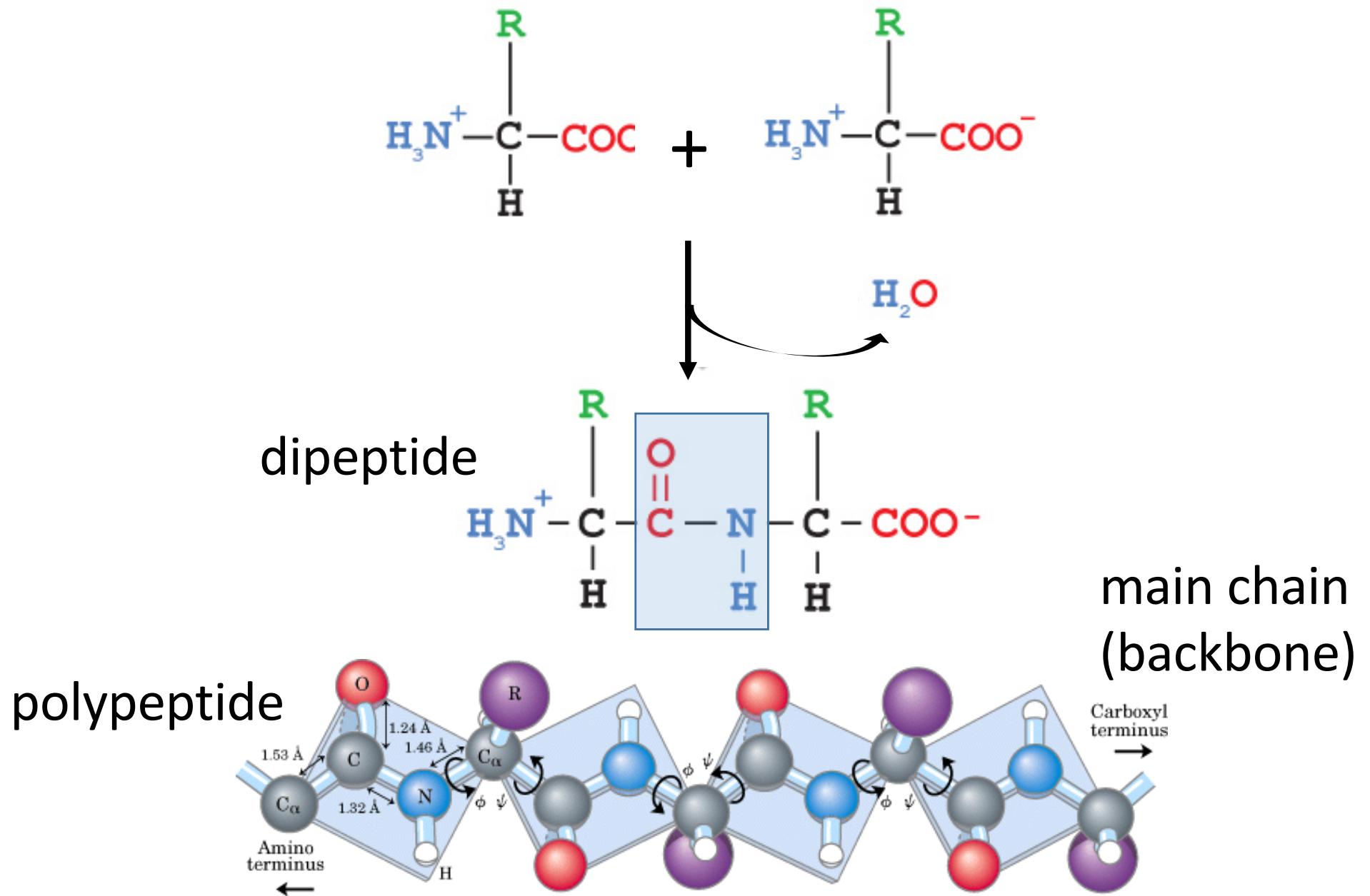
acid: proton donor

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

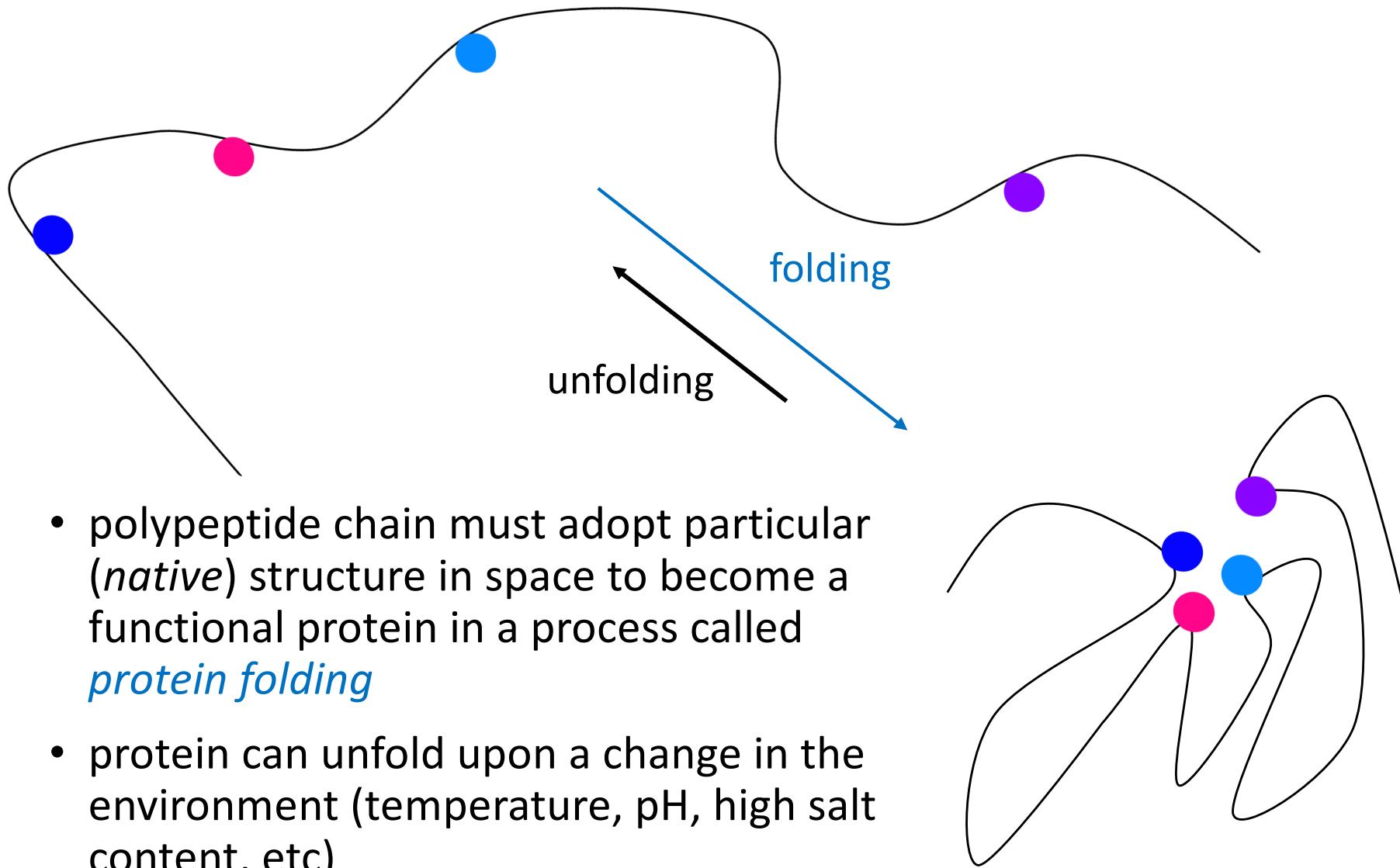
nonpolar residues: neutral molecules without a dipole (weak interactions, hydrophobic)



peptide bond

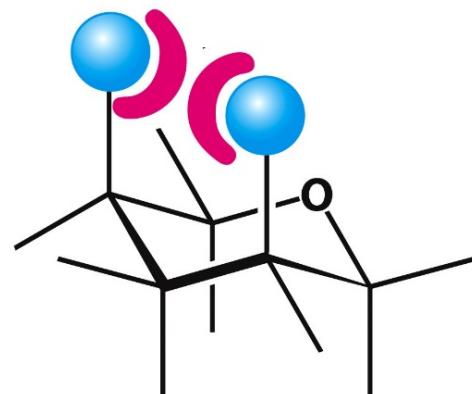


protein vs. polypeptide

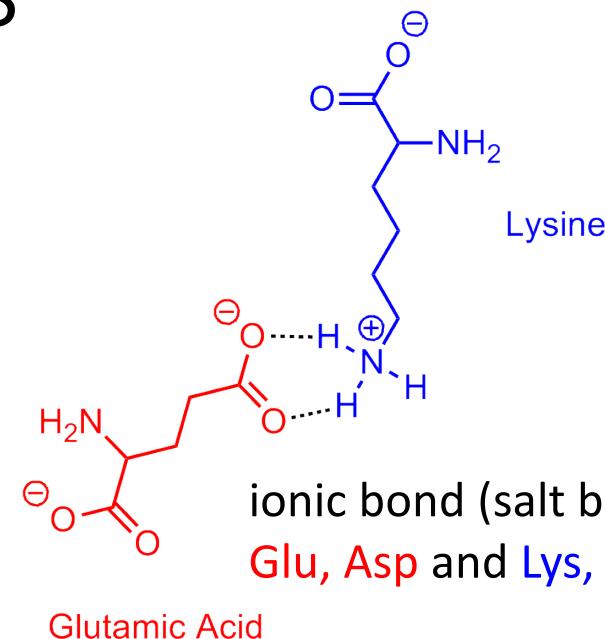
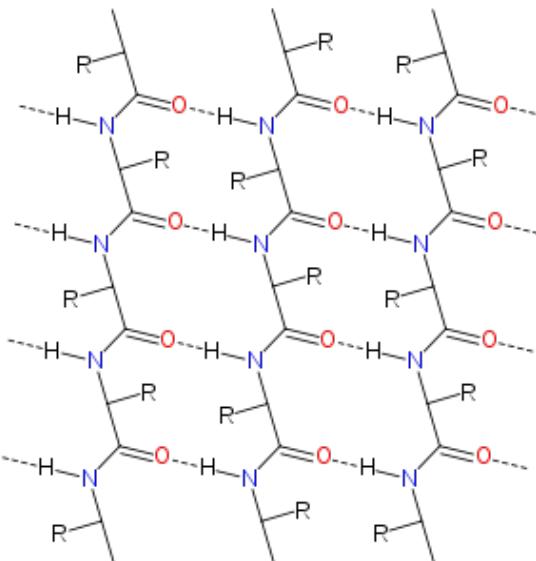


interactions & motions

steric hindrance



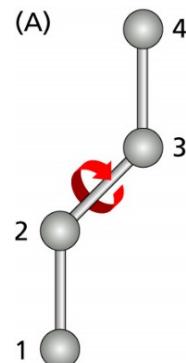
Unnumbered 11 p324
Biochemistry, Seventh Edition
© 2012 W. H. Freeman and Company



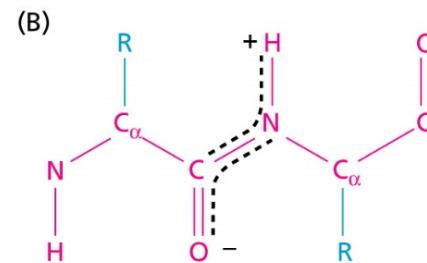
ionic bond (salt bridge):
Glu, Asp and Lys, Arg

Glutamic Acid

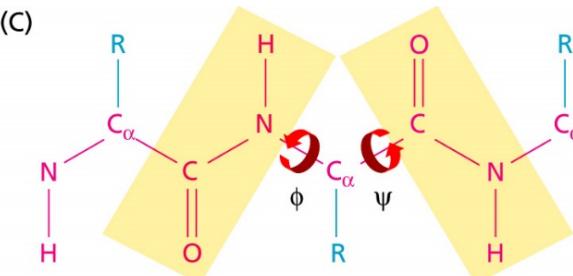
hydrogen bonding



rotation around
single bond

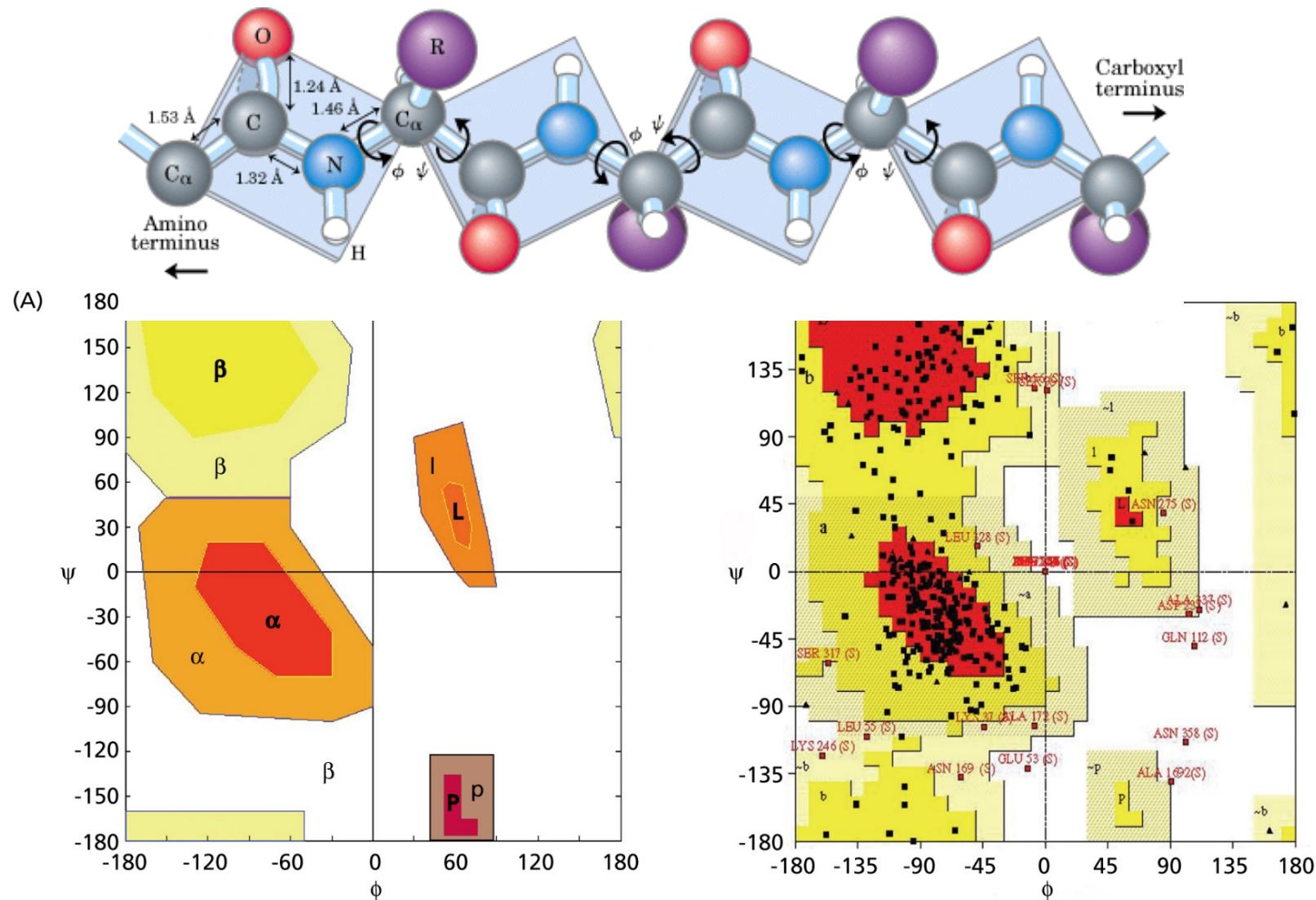


peptide bond is planar due to
the electron resonance (partial
double bond character)



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

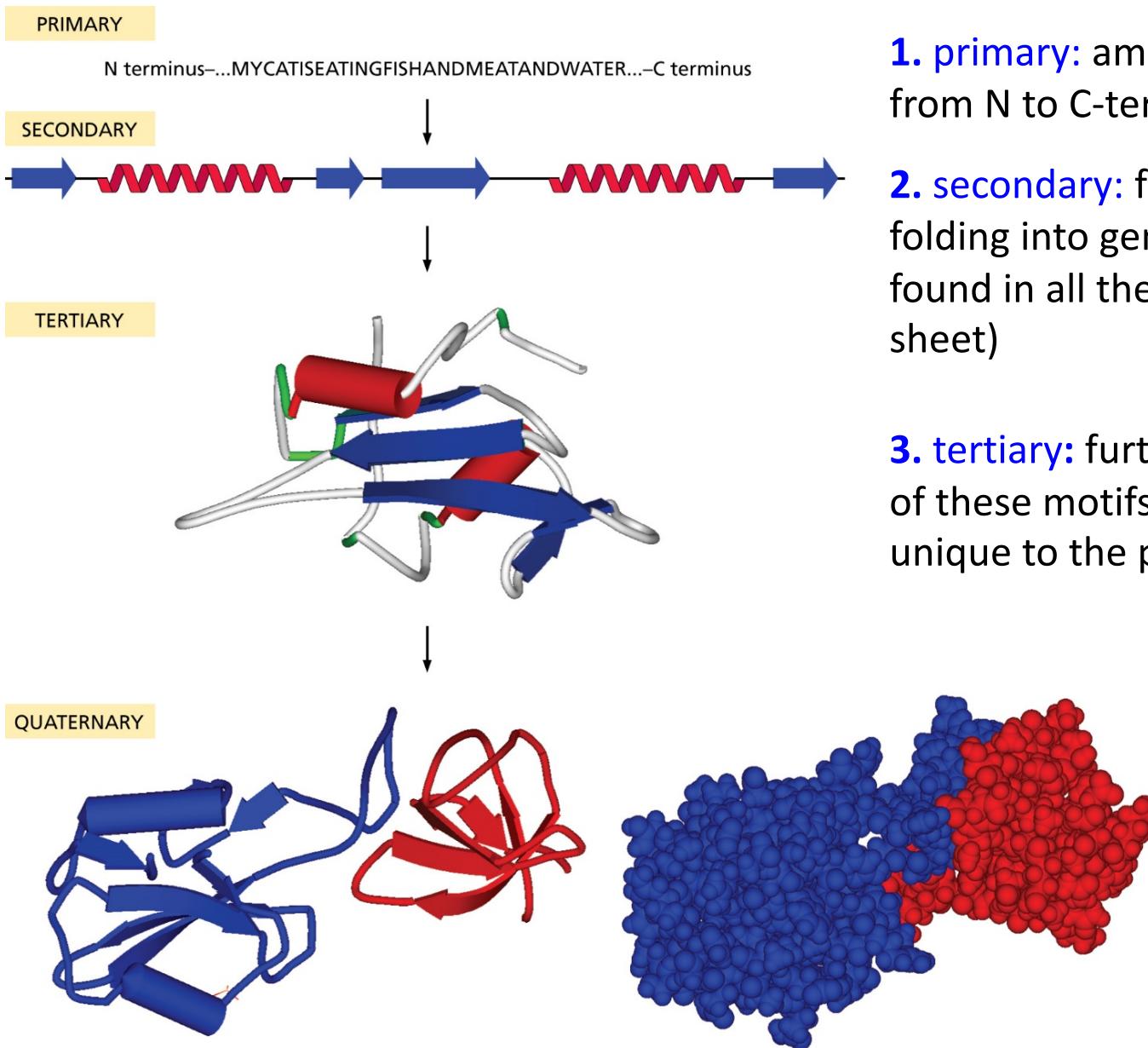
Ramachandran plot



existence of the preferred orientations between the peptide bonds

protein structure

- there are four levels of protein structure:



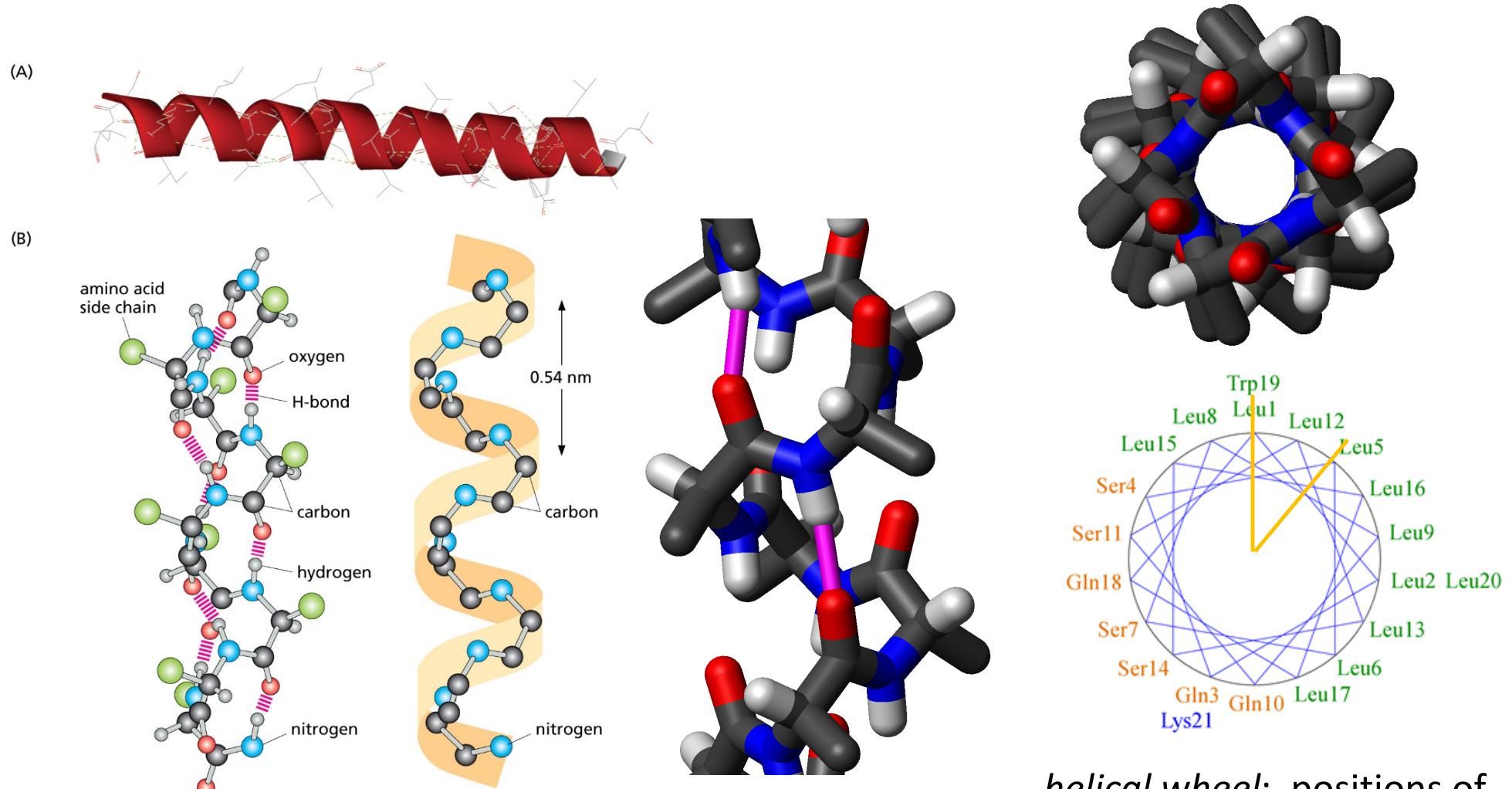
1. primary: amino acid sequence read from N to C-terminus

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

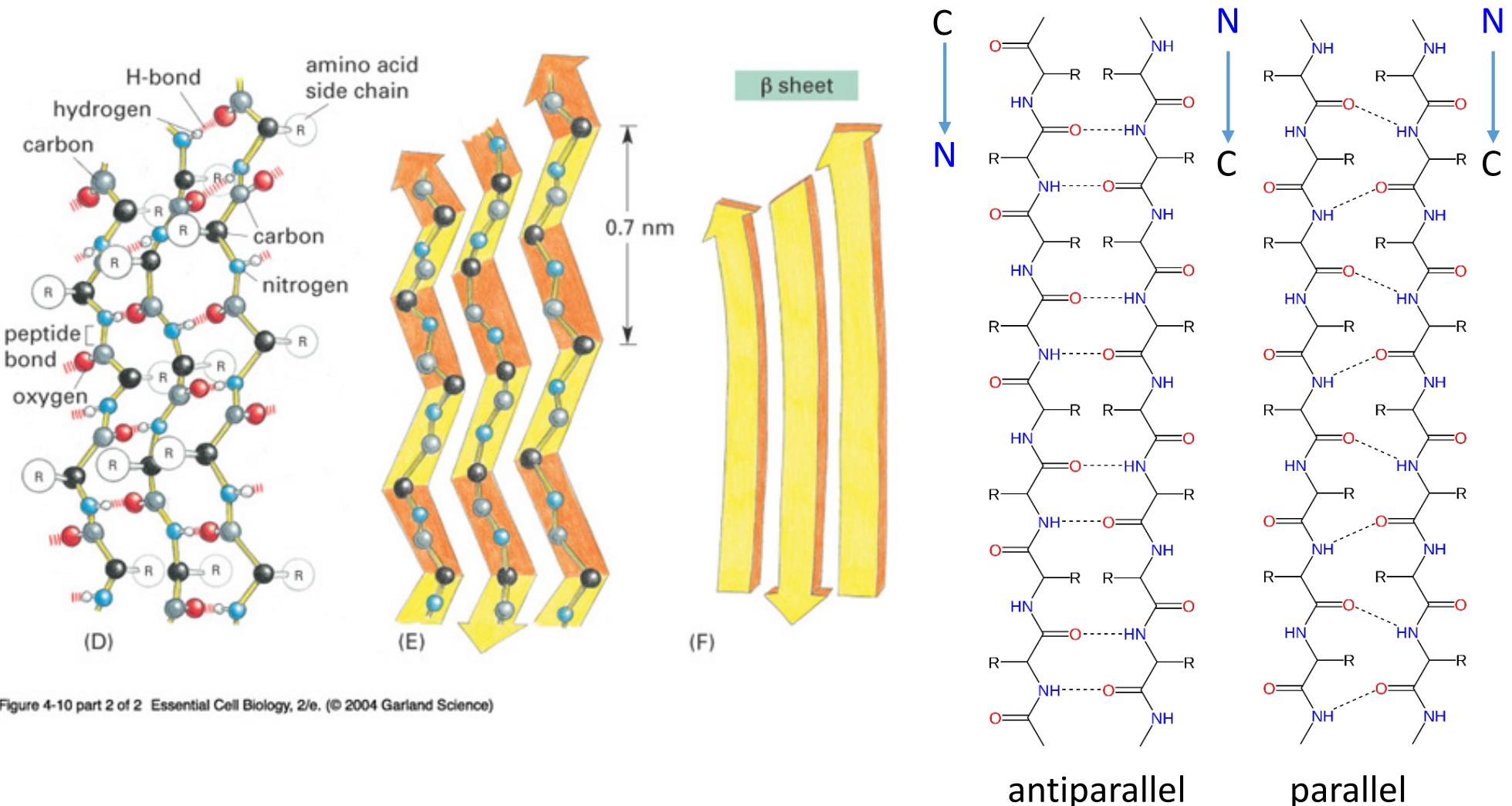
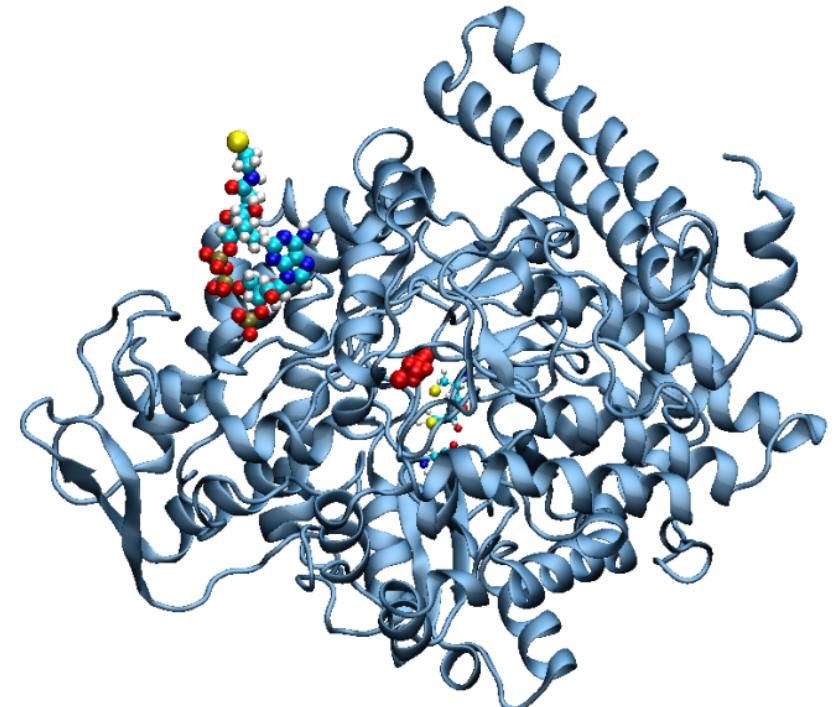
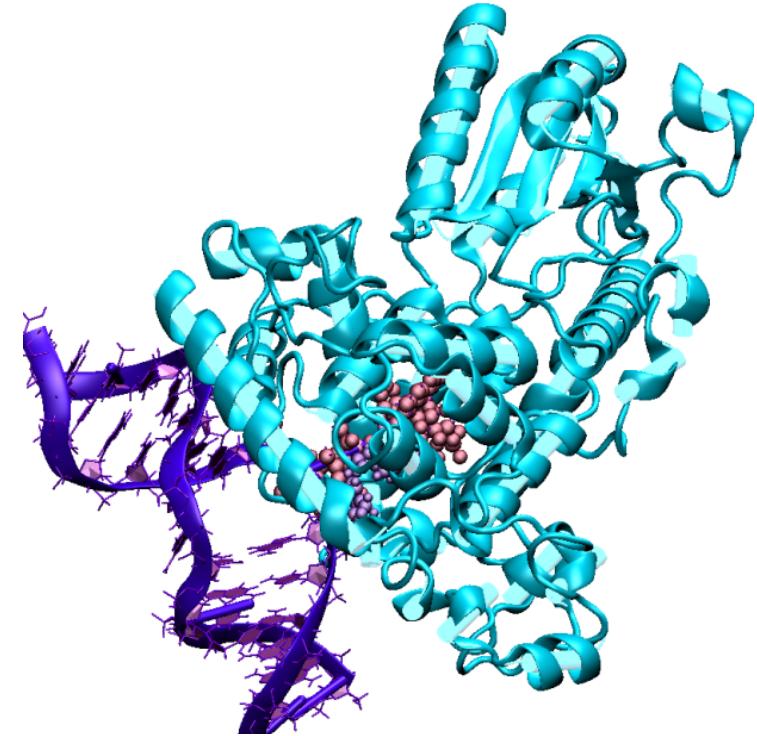


Figure 4-10 part 2 of 2 Essential Cell Biology, 2/e. (© 2004 Garland Science)

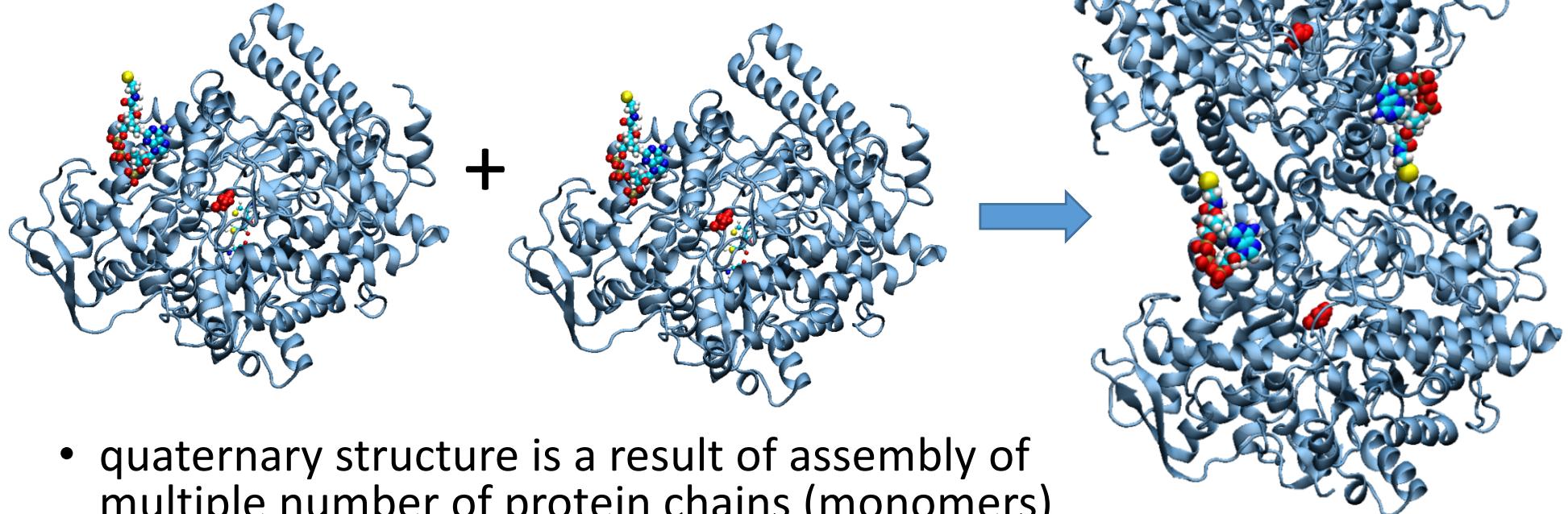
- perfect β -sheets contain all-parallel or all-antiparallel strands, but in nature they come as mixture of both

tertiary structure

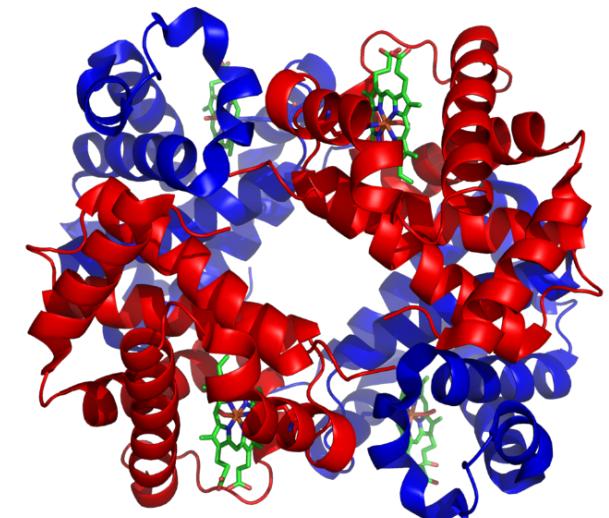
- tertiary structure is a product of packing various combinations of secondary structures (domains)
- we are still unable to predict the tertiary structure of the protein based only on its sequence → *homology modelling*
- *homologous* proteins have a common ancestor → similarity in structure and function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



quaternary structure



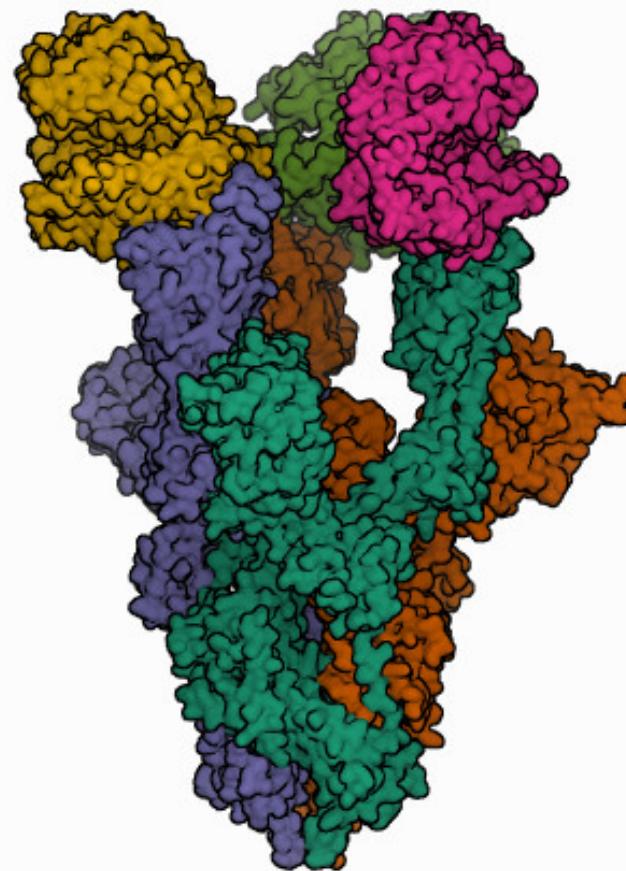
- quaternary structure is a result of assembly of multiple number of protein chains (monomers) into large supramolecular complexes → *oligomers*
- the subunits can be identical (for example, homodimers) or they can be different (heterodimers)
- tetrameric haemoglobin (right) with two α and two β subunits



Coronavirus update: 3D structures

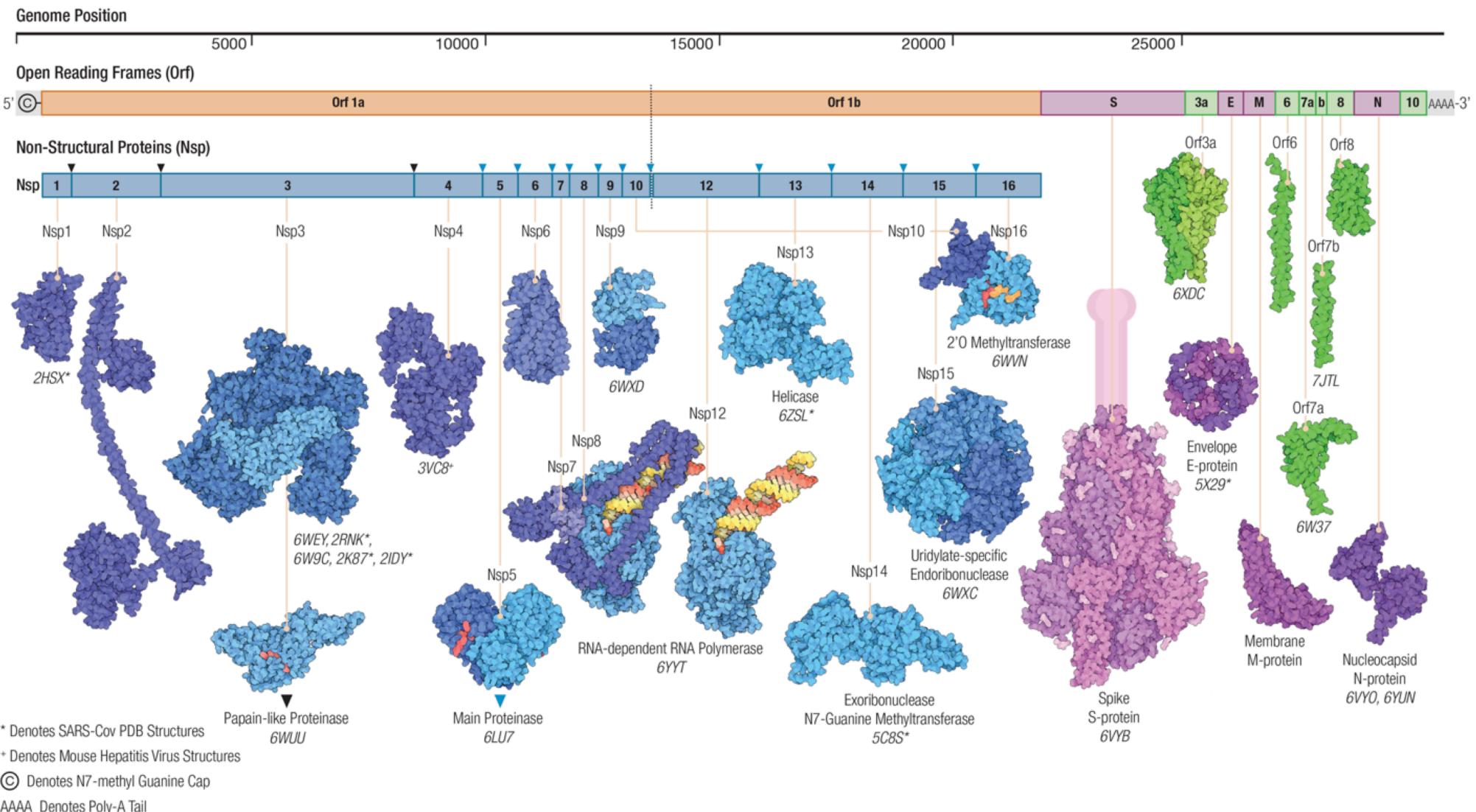


ACE2 bound to spike
protein (binding domain)



3 x ACE2 bound to spike
protein (full protein)

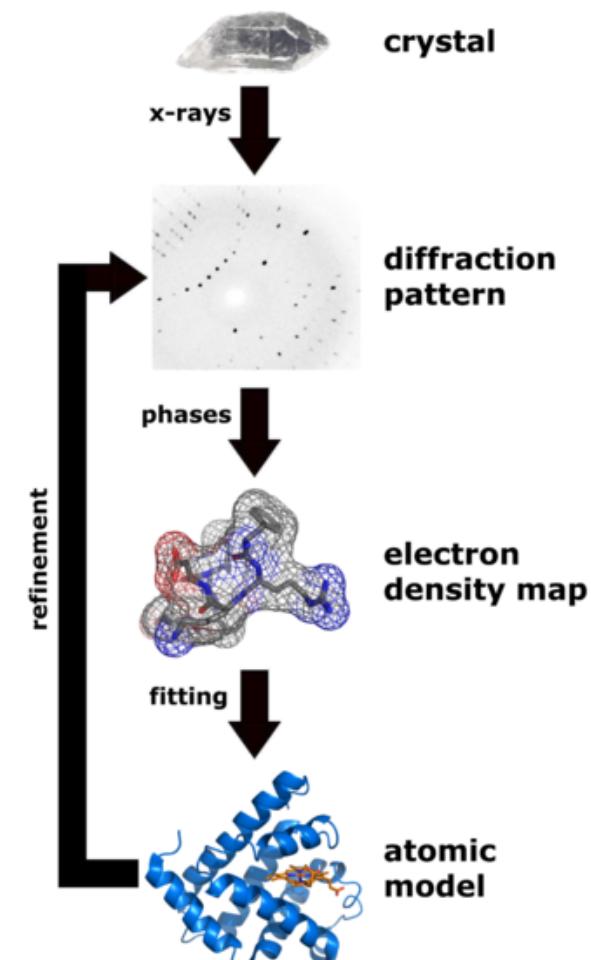
Coronavirus update: 3D structures



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



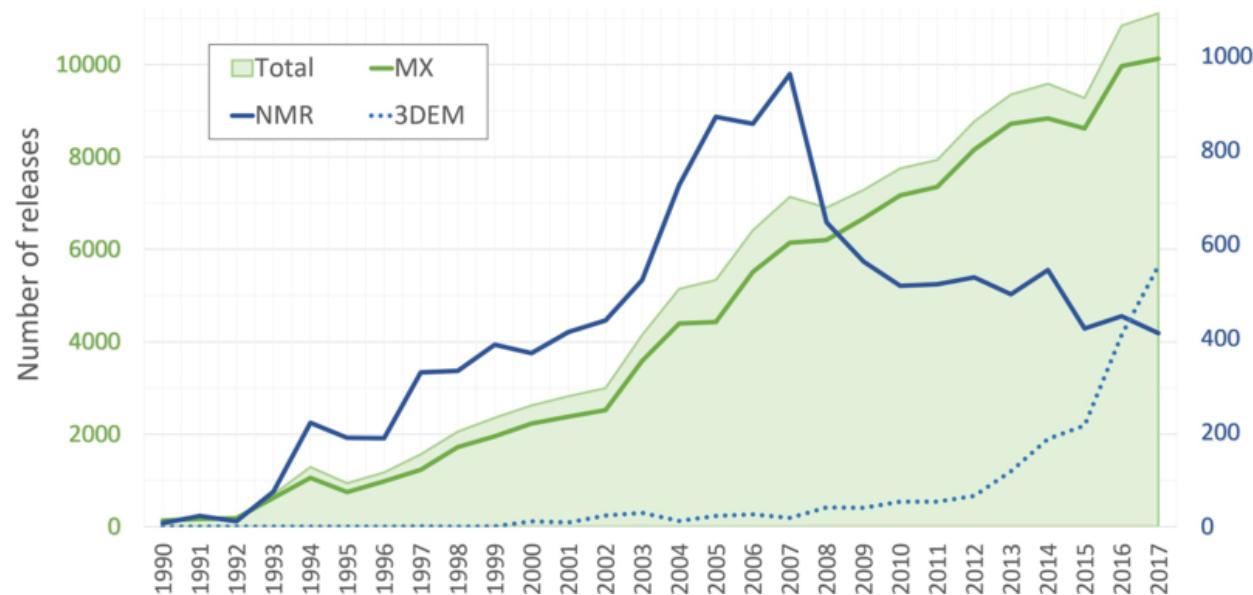
protein databases

- 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>

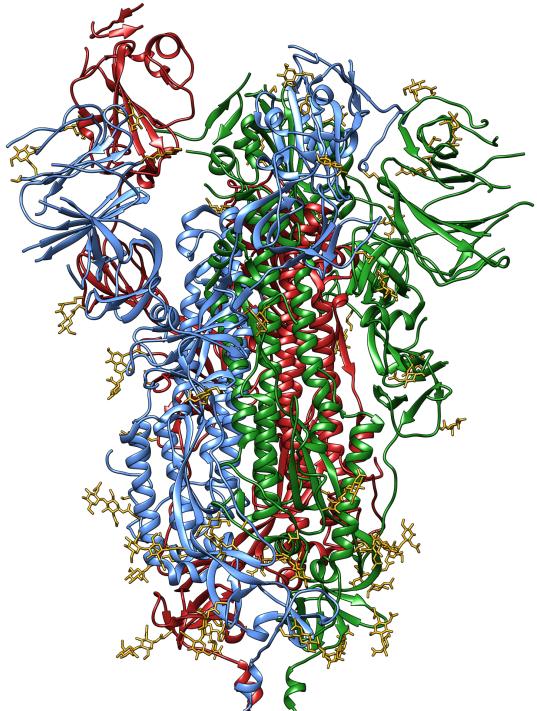


174,000 structures available in 2021 (77,000 in 2011)

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)).