EPFL

Biology for MX

MSE – 212

Prof. Maartje M.C. Bastings

Programmable Biomaterials Laboratory

Course 2: Proteins and Protein Based Materials

Course Content

EPFL

Intro Ex.	*Defining our starting point (online questionnaire & video)*	*(February 19)*

BLOCK 1: Introduction and engineering with cellular components

Lecture 1.	Intro to biology and cells	(February 26)
Lecture 2.	Proteins and protein based materials	(March 5)
Lecture 3.	DNA and DNA-based materials	(March 12)
Exercise 1.	*Proteins, peptides and DNA*	*(March 19)*

BLOCK 2: Inter- and intracellular action

Lecture 4.	ECM, adhesion and artificial matrices	(March 26)
Lecture 5.	Virus, antibodies and immune engineering	(April 2)
Lecture 6.	Bacteria	(April 9)
Exercise 2.	*Nanoparticles and Scaffolds*	*(April 16)*

BLOCK 3: Physics of biological processes

Lecture 7.	Receptors and targeting	(April 30)
Lecture 8.	Endocytosis	(May 7)
Lecture 9.	Signaling and communication	(May 14)
Exercise 3.	*Engineering functionality*	*(May 21)*

Lecture 10.	Revision and conclusion	(May 28)

On Today's Menu:

Fundaments of Proteins

Engineering with Proteins

Part 1

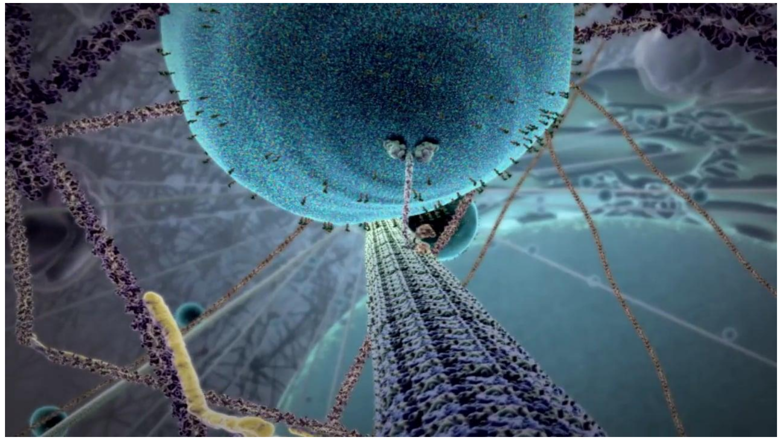
What are Proteins?
- Structure
- Function
- Production

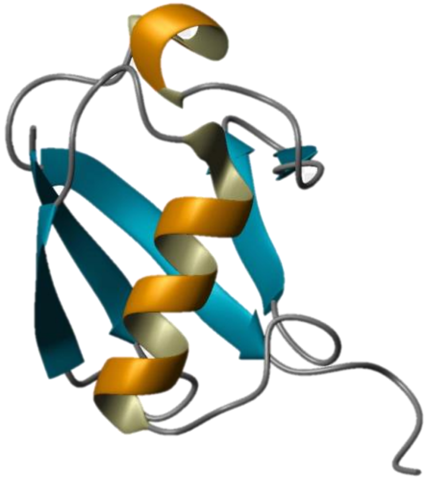
Part 2

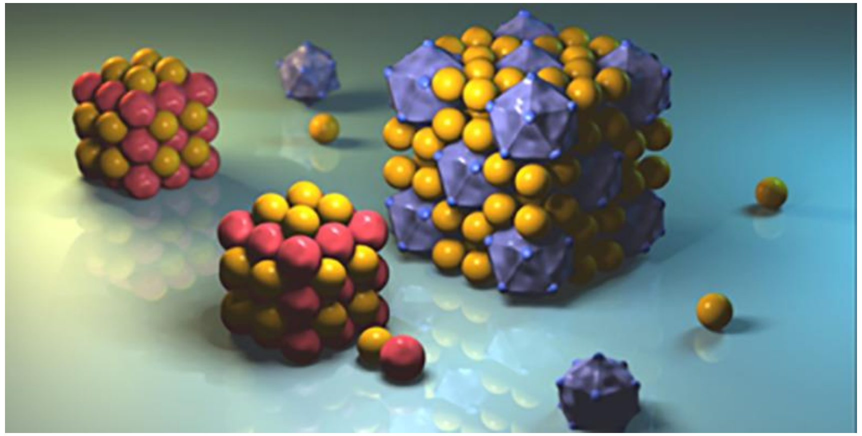
Protein Engineering
- Mutants
- Synthesis
- Folding Prediction (AI)

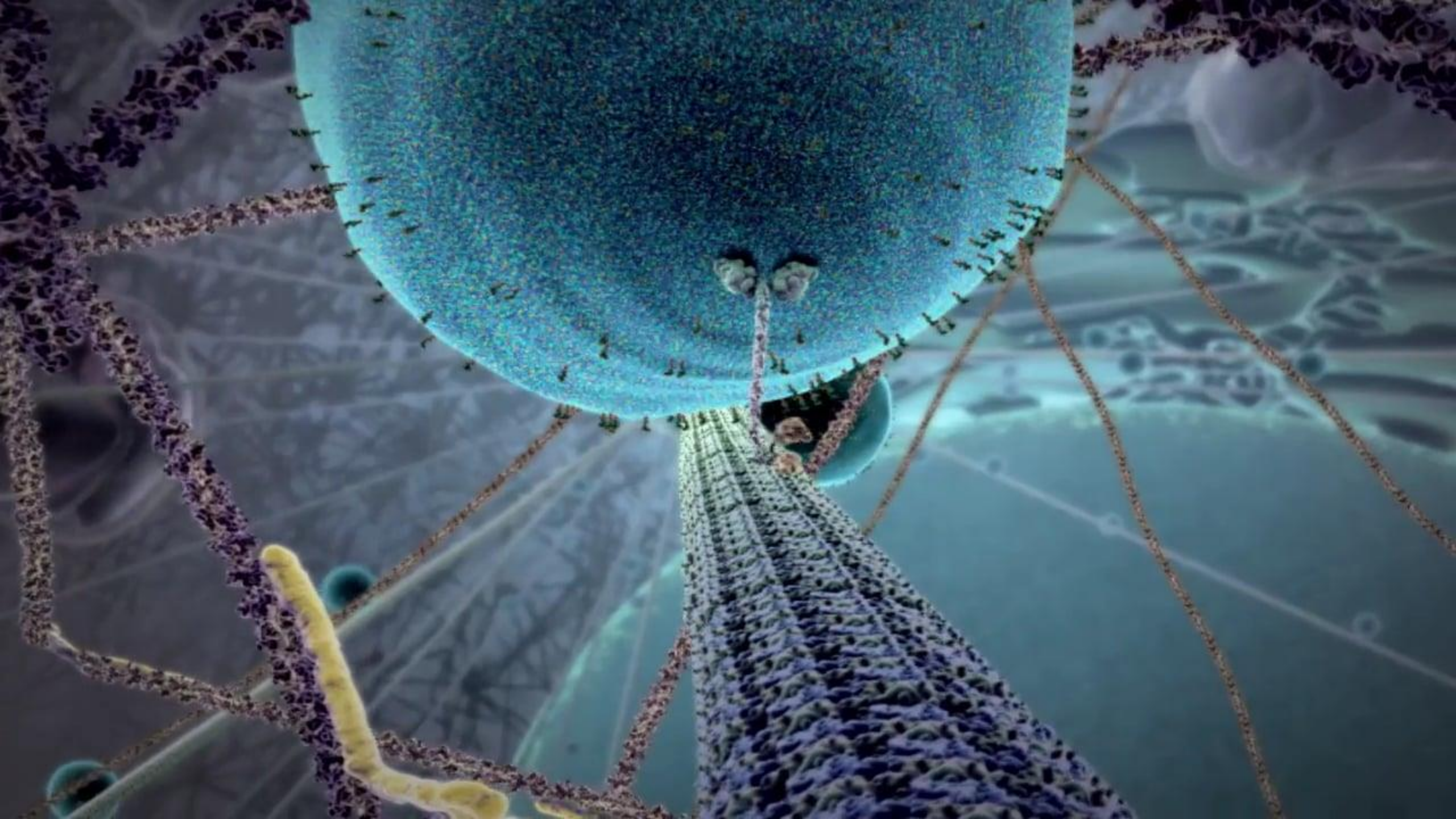
Part 3

Protein-based Materials
- Particles
- Surfaces
- Fun stuff

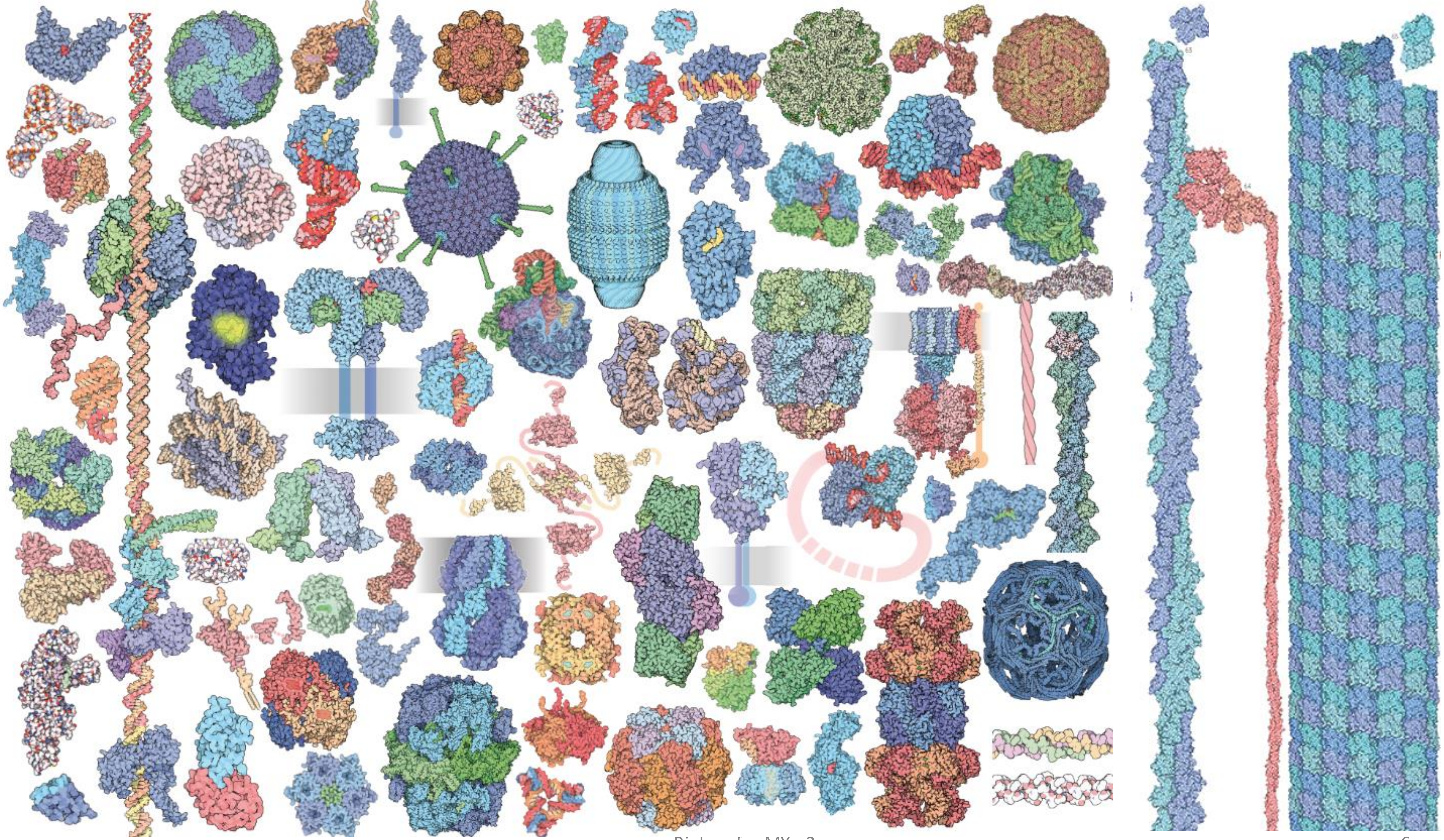








Why do we need to eat protein?

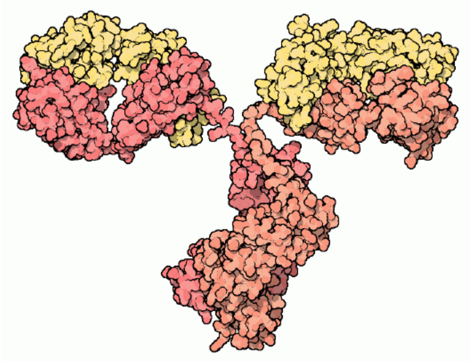


https://pdb101.rcsb.org/motm/motm-about

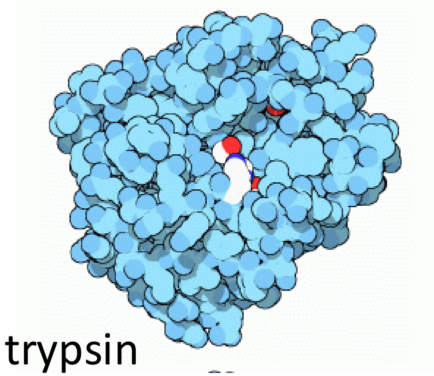
7 Types of Proteins

All figures from David Goodsell

Antibodies:
Protection

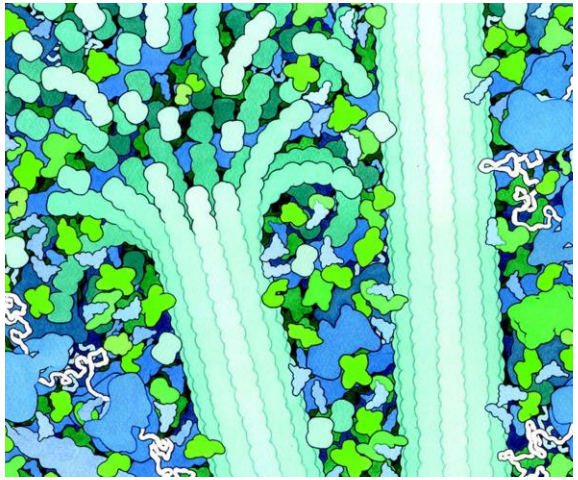


Enzymes:
Chemical reactions



trypsin

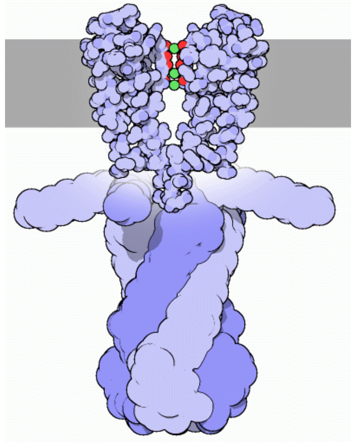
Structural proteins:
Support and Strength

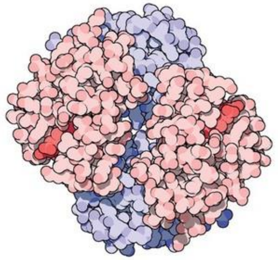


Hormonal proteins:
Regulation, development

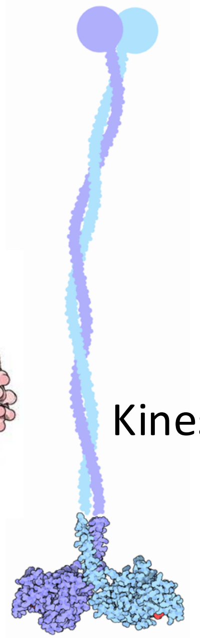
insulin

Transport proteins
Communication



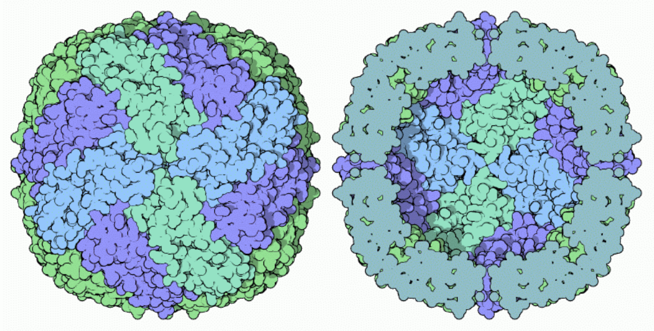


Hb



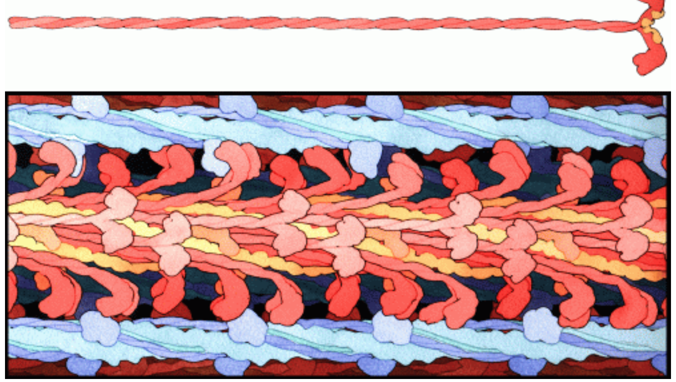
Kinesin

Storage proteins
Reserves



Ferritin

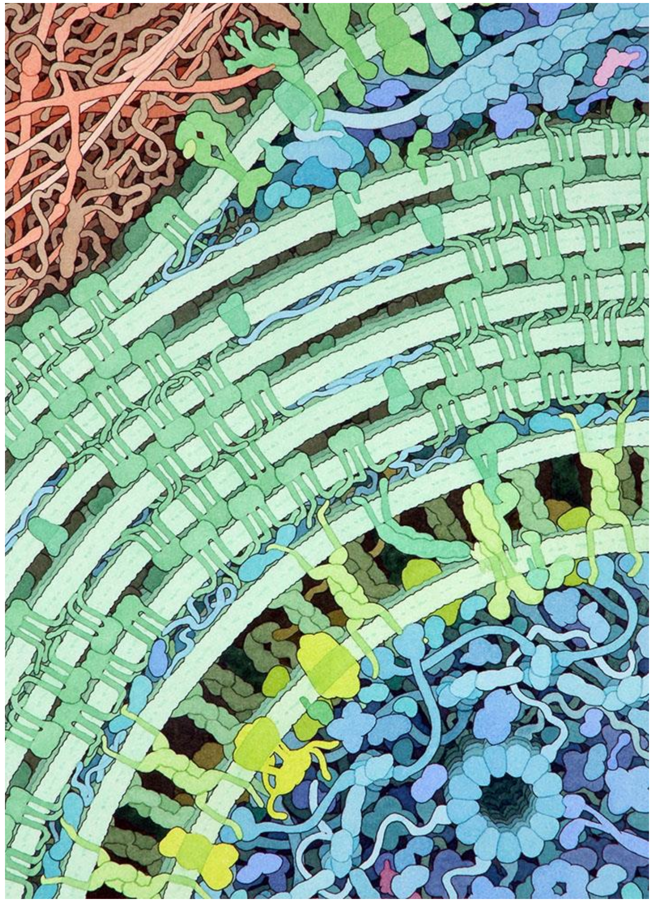
Contractile proteins
movement



Myosin

David Goodsell

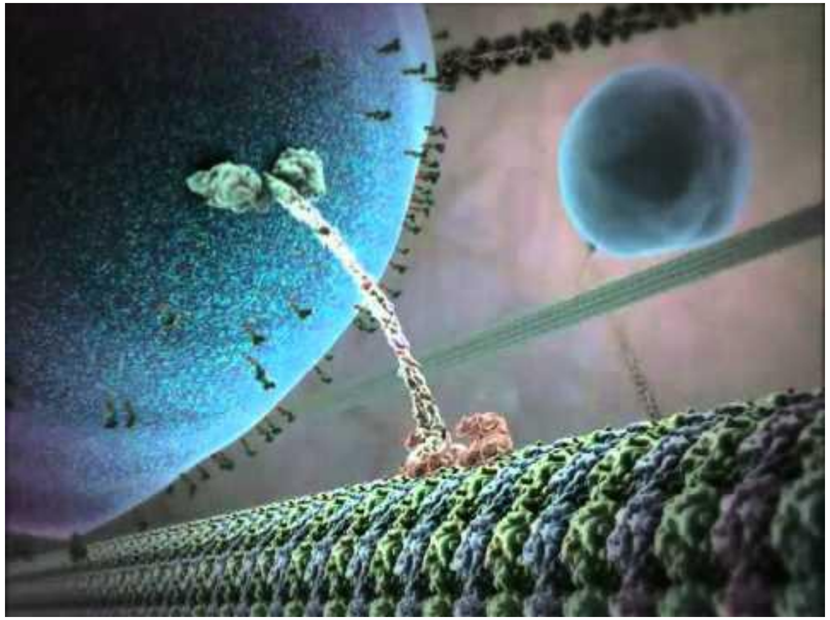
Professor of Computational Biology at the Scripps Research Institute,
Research Professor at Rutgers University, Scientific Outreach Lead at the RCSB Protein Data Bank.

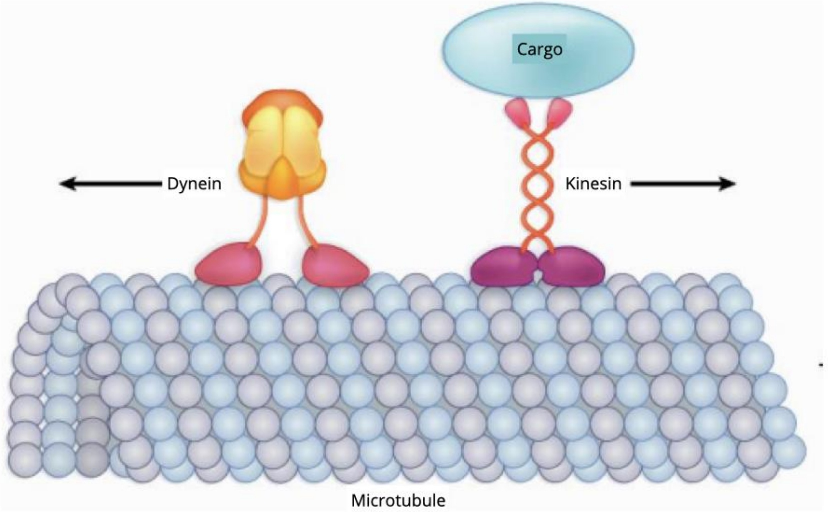


https://ccsb.scripps.edu/goodsell/

Motor Proteins: Kinesin and Dynein

Nature's robots!

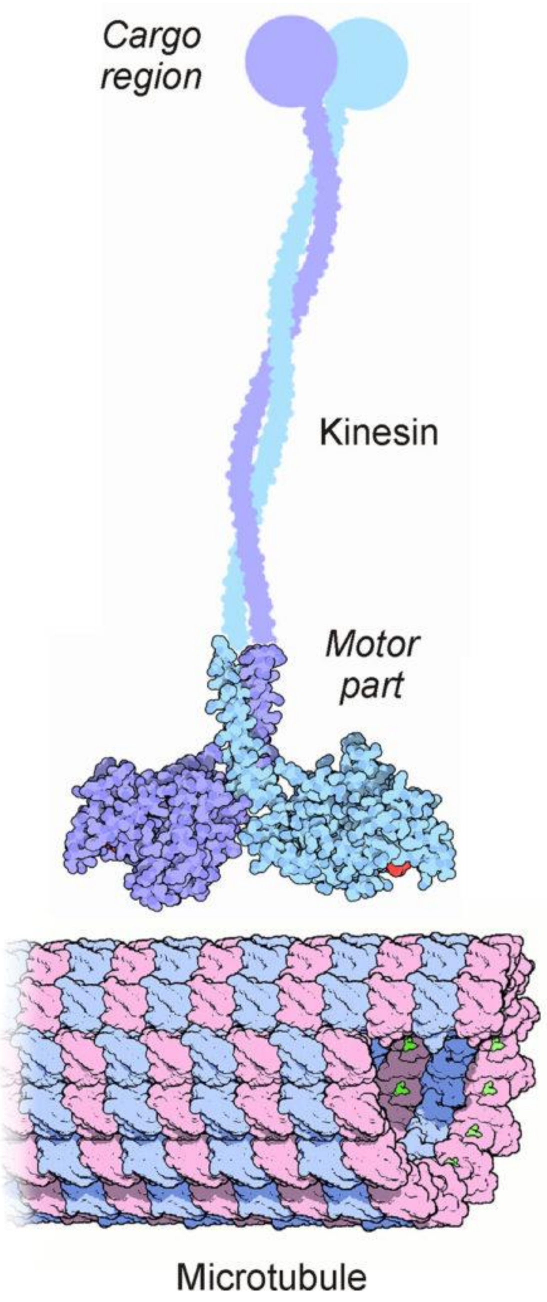




Motor proteins use the **energy of ATP hydrolysis** to move along microtubules or actin filaments. They mediate the sliding of filaments relative to one another and the transport of membrane-enclosed matter along filament tracks.

1.motor proteins on **actin filaments** are members of the **myosin superfamily**.

2.The motor proteins on **microtubules** are members of the **kinesin superfamily or the dynein family**.



iBiology course on Motor Proteins
https://youtu.be/9RUHJhskW00?si=cSsXC4tNnkldJqyp

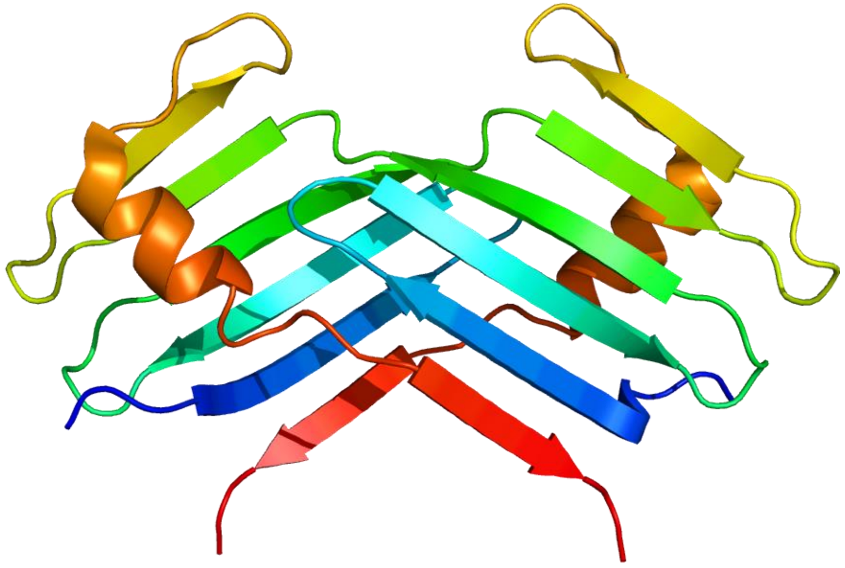
What are Proteins?

Proteins are large biomolecules consisting of one or more **long chains of amino acid residues**.

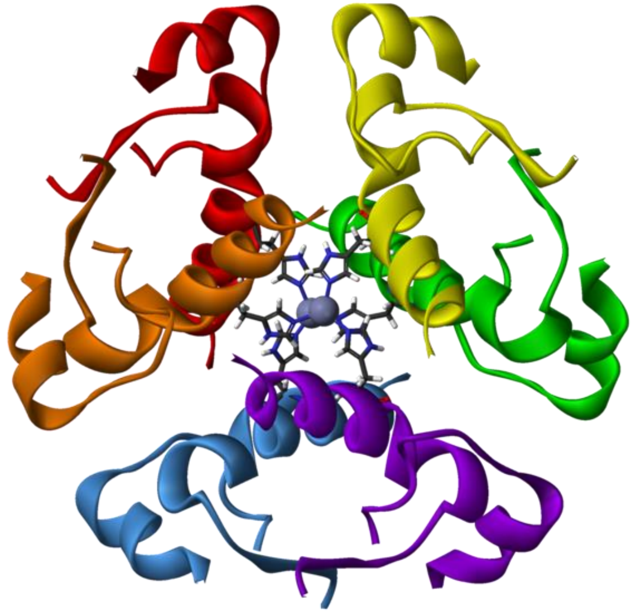
Proteins **differ** from one another primarily in their **sequence of amino acids**, which is dictated by the nucleotide sequence of their genes, and which usually results in protein **folding** into a specific **3D structure that determines its activity.**

7 types of proteins:
antibodies, contractile proteins, enzymes, hormonal proteins, structural proteins, storage proteins, and transport proteins

Serine/Threonine Protein Kinase Plk4

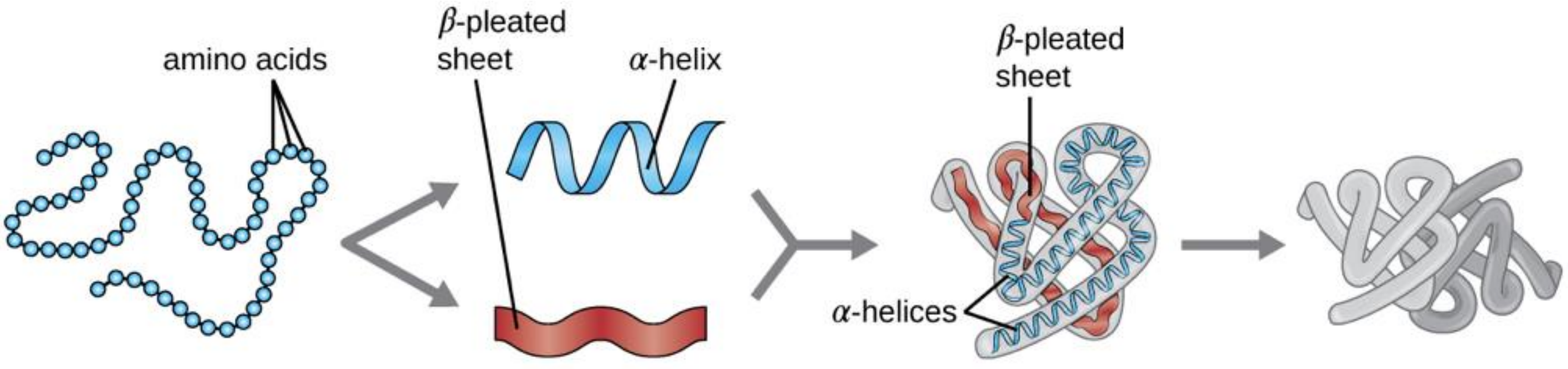


Insulin hexamer



Fundamentals of Proteins

Sequence → Structure



Primary Protein Structure
Sequence of a chain of amino acids

Secondary Protein Structure
Local folding of the polypeptide chain into helices or sheets

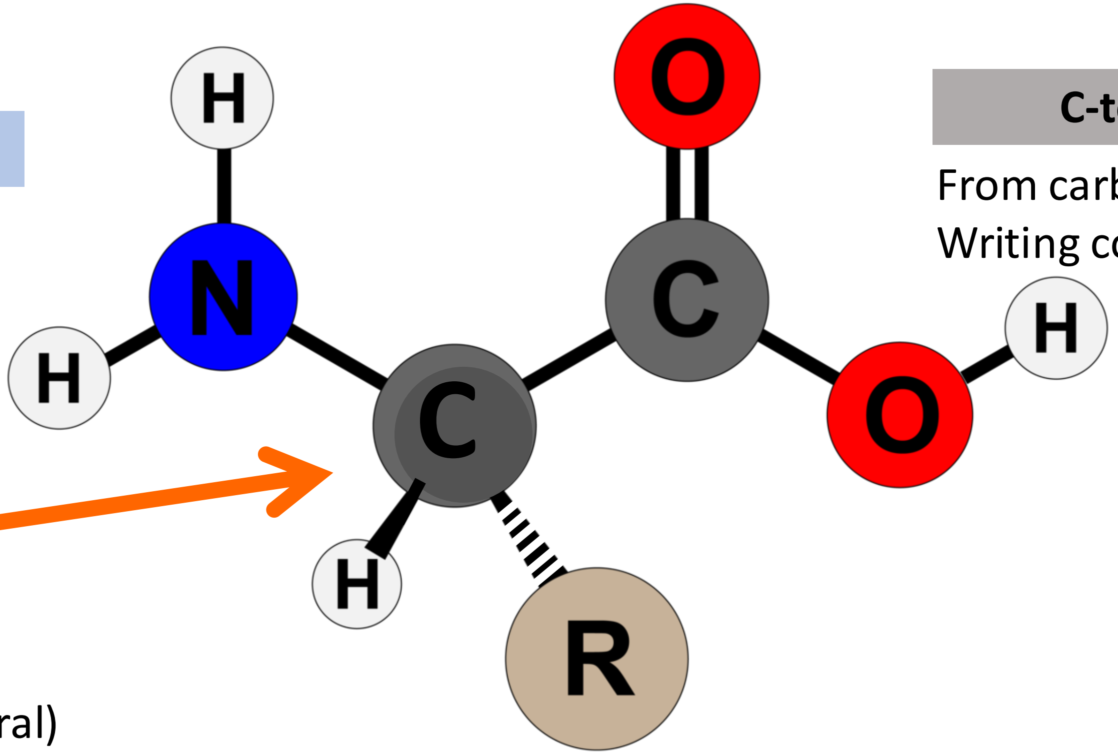
Tertiary Protein Structure
three-dimensional folding pattern of a protein due to side chain interactions

Quaternary Protein Structure
protein consisting of more than one amino acid chain

Amino Acids

N-terminus

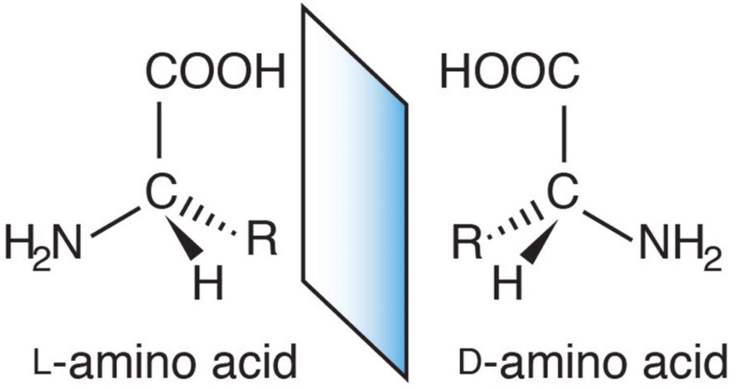
From amine.



C-terminus

From carboxyl.
Writing convention is always N → C

Chiral central carbon!
Natural AAs are all in **L-form**
(Except Glycine, which is not chiral)



Side-chain

Can be polar (uncharged), apolar (hydrophobic), Basic or Acidic

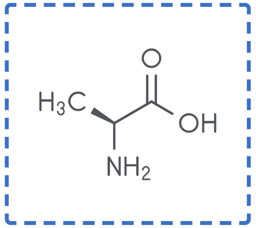
20 Amino Acids

EPFL

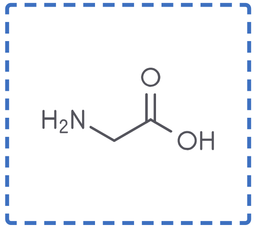
Chart Key

Alkyl | Aromatic | Neutral | Acidic | Basic | Essential | Non-Essential

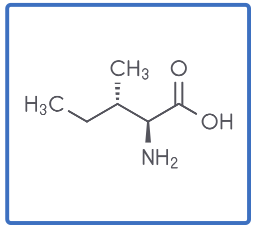
Note: The NH_2 and COOH values listed below are pK_a values.



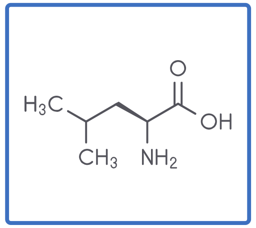
Alanine
Ala A
NH_2: 9.87 COOH: 2.35



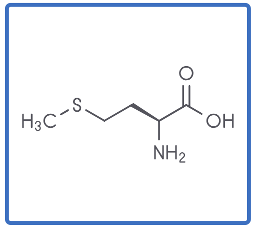
Glycine
Gly G
NH_2: 9.60 COOH: 2.34



Isoleucine
Ile I
NH_2: 9.76 COOH: 2.32

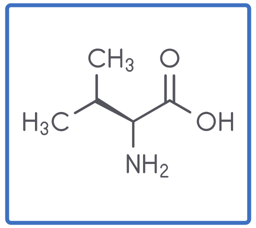


Leucine
Leu L
NH_2: 9.60 COOH: 2.36

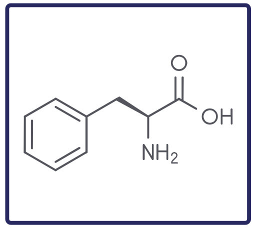


Methionine
Met M
NH_2: 9.21 COOH: 2.28

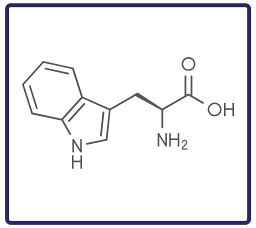
Proline
Pro P
NH_2: 10.60 COOH: 1.99



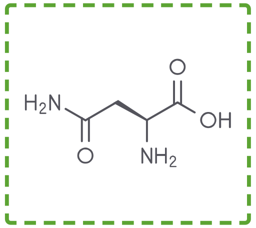
Valine
Val V
NH_2: 9.72 COOH: 2.29



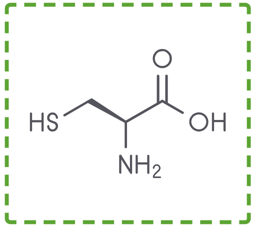
Phenylalanine
Phe F
NH_2: 9.24 COOH: 2.58



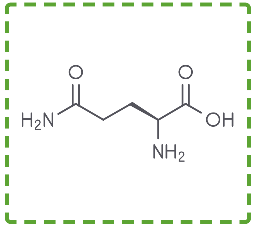
Tryptophan
Trp W
NH_2: 9.39 COOH: 2.38



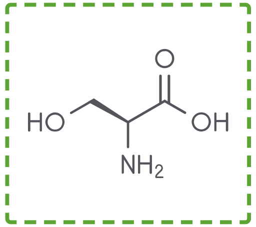
Asparagine
Asn N
NH_2: 8.80 COOH: 2.02



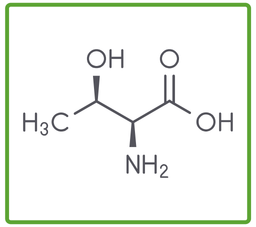
Cysteine
Cys C
NH_2: 10.78 COOH: 1.71



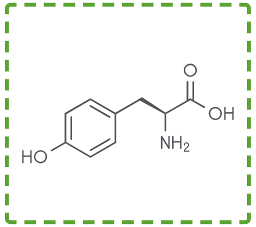
Glutamine
Gln Q
NH_2: 9.13 COOH: 2.17



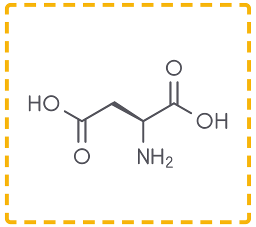
Serine
Ser S
NH_2: 9.15 COOH: 2.21



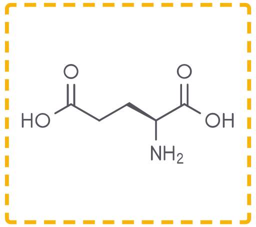
Threonine
Thr T
NH_2: 9.12 COOH: 2.15



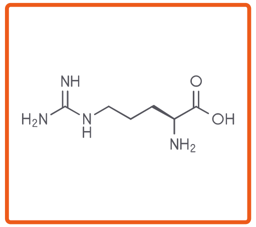
Tyrosine
Tyr Y
NH_2: 9.11 COOH: 2.20



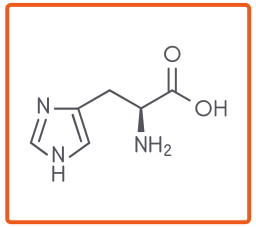
Aspartic Acid
Asp D
NH_2: 9.60 COOH: 1.88



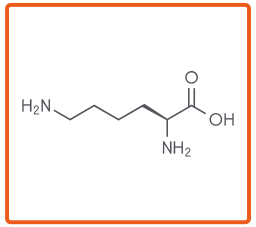
Glutamic Acid
Glu E
NH_2: 9.67 COOH: 2.19



Arginine
Arg R
NH_2: 9.09 COOH: 2.18



Histidine
His H
NH_2: 8.97 COOH: 1.78



Lysine
Lys K
NH_2: 10.28 COOH: 8.90

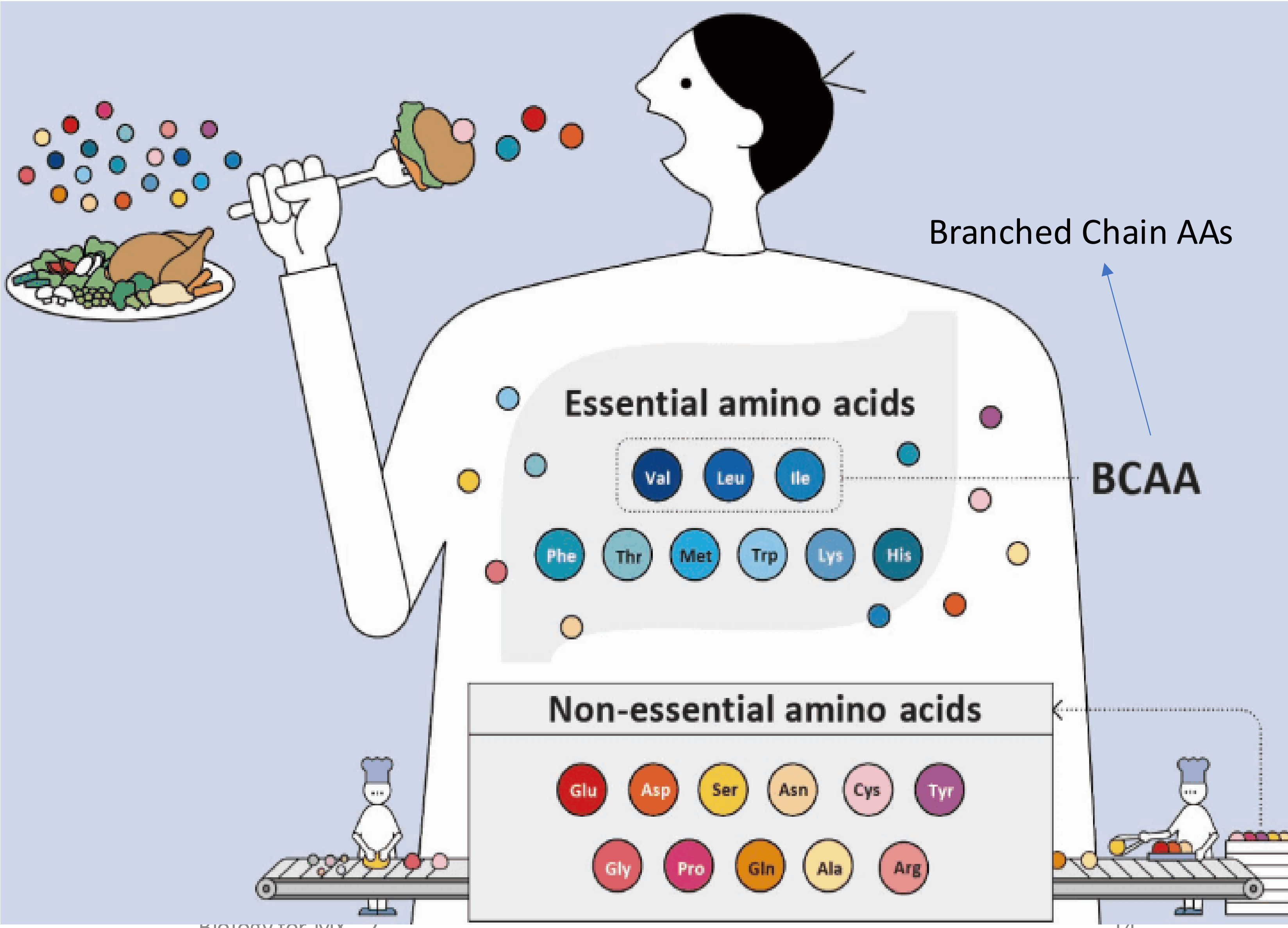
Source: Fisher Scientific

20 Amino Acids

Non Essential: our body can make them
Alanine; Asparagine; Aspartic acid; Glutamic acid; Serine

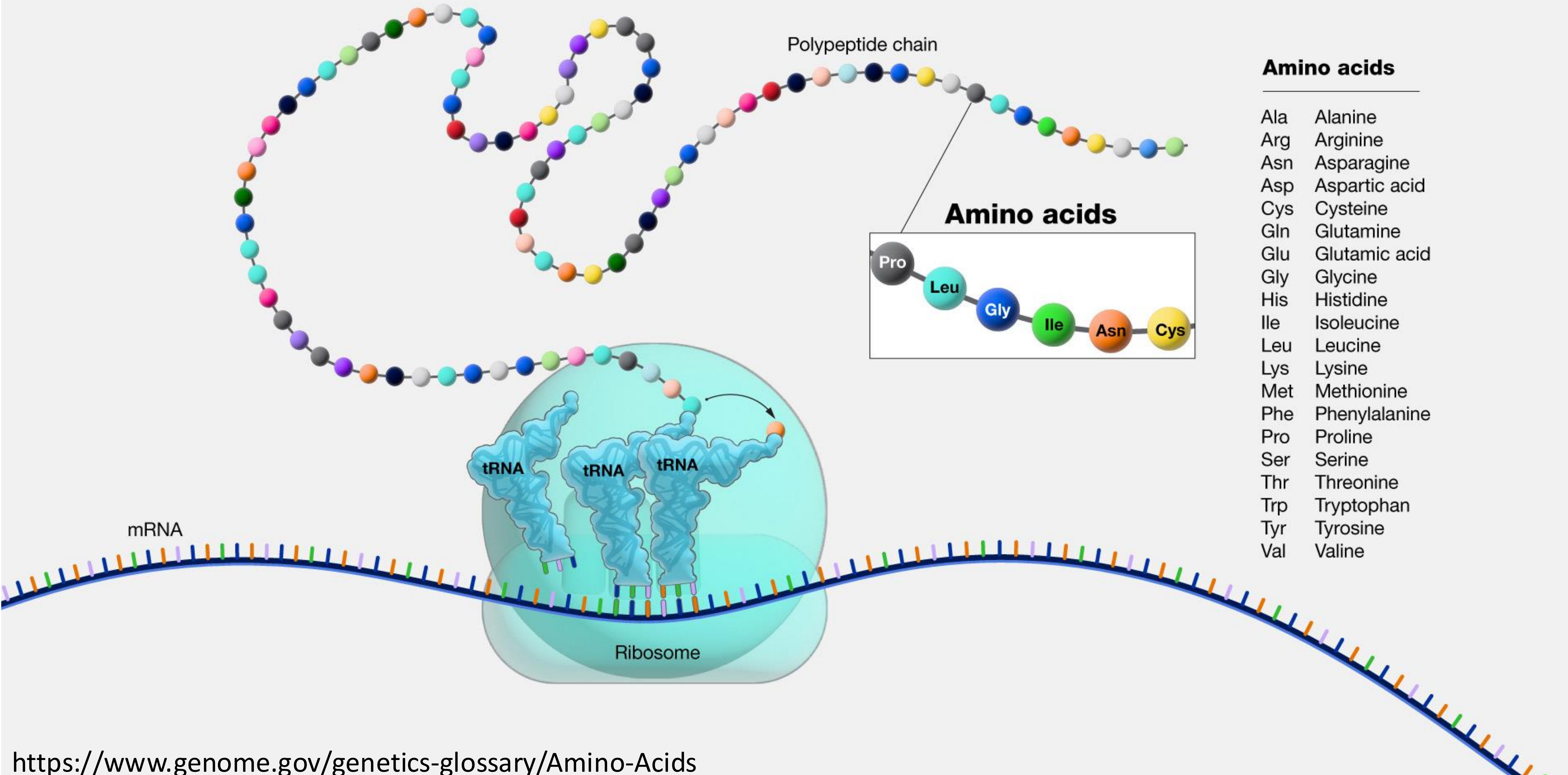
Conditionally-Essential : healthy bodies can make them under normal physiologic conditions. They become essential under certain conditions like starvation or inborn errors of metabolism.
Arginine; Cysteine ; Glutamine; Glycine; Proline; Tyrosine

Essential : cannot be made by us. Dietary protein provides these amino acids, which are needed to make certain hormones and other important molecules.
Histidine; Isoleucine; Leucine; Lysine; Methionine; Phenylalanine; Threonine; Tryptophan; Valine



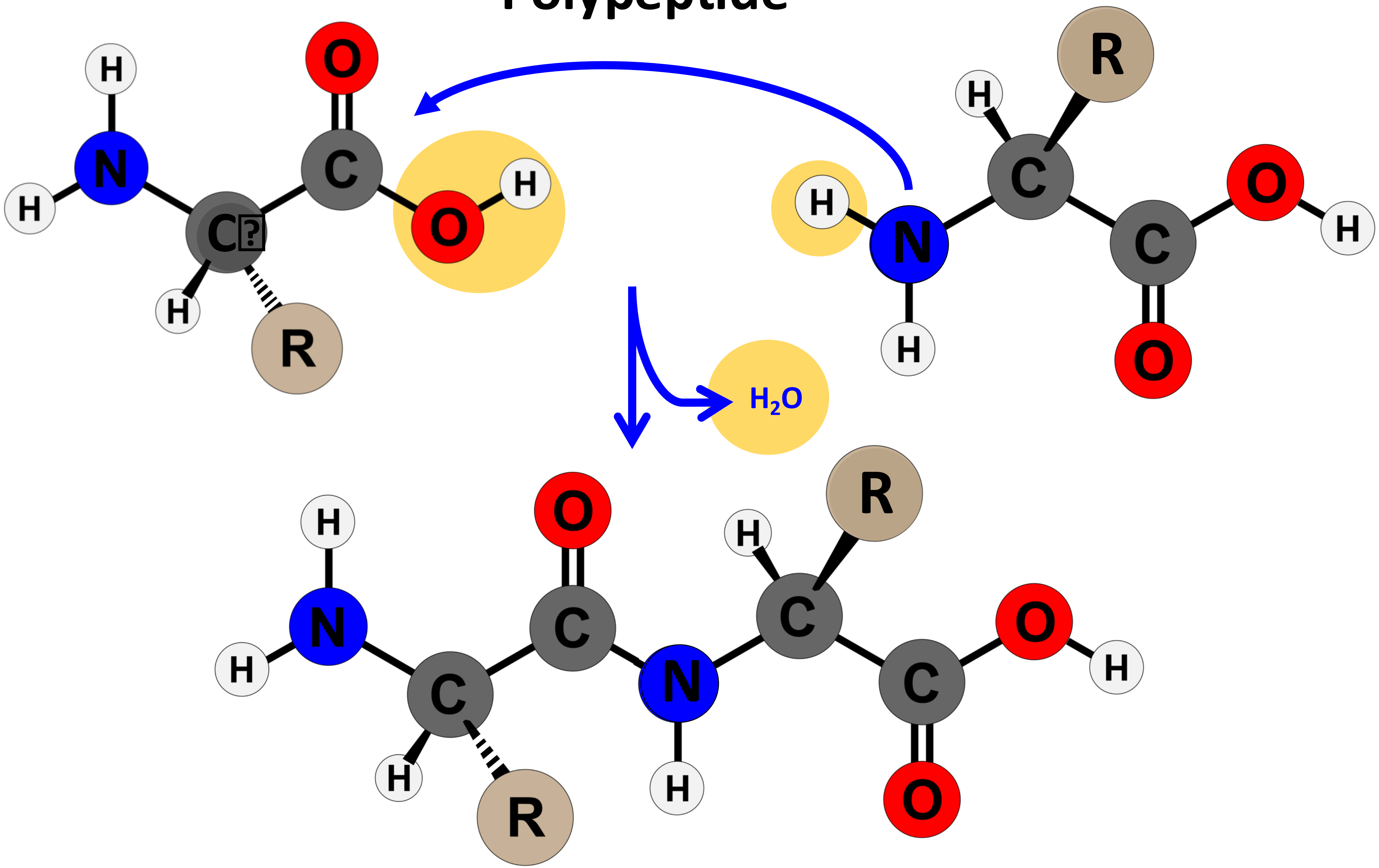
https://www.ajinomoto.com/amino-acids/what-are-amino-acids

The Making of a Protein



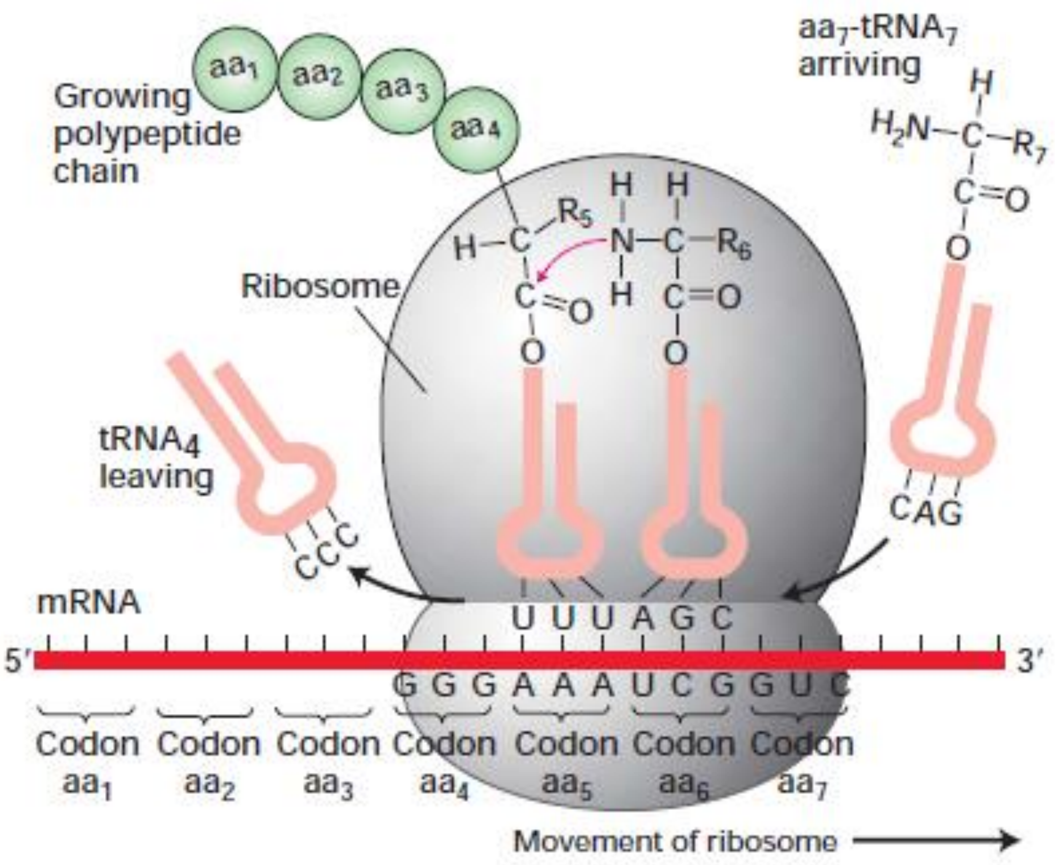
https://www.genome.gov/genetics-glossary/Amino-Acids

Polypeptide

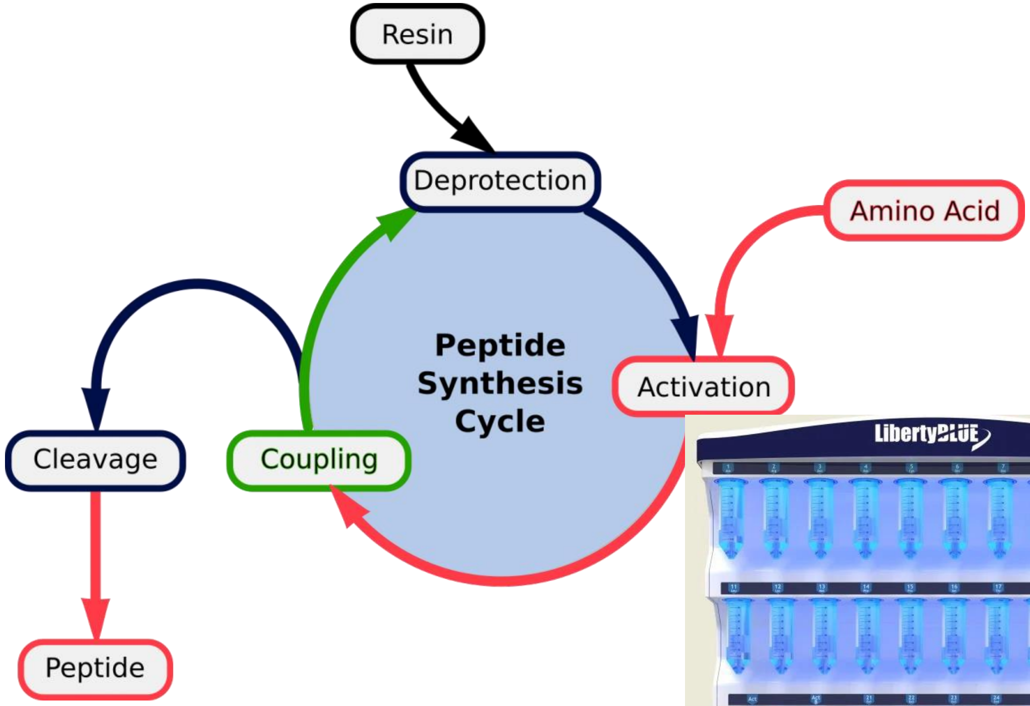


Peptide Production

In the cell by the ribosome: N → C



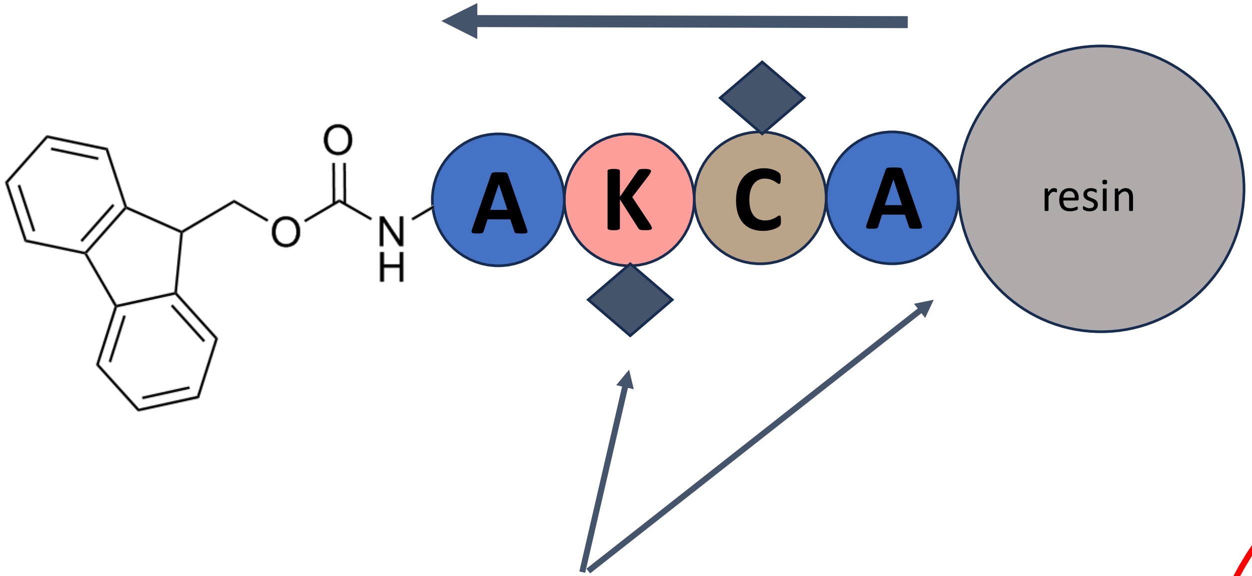
In the chemistry lab : C → N



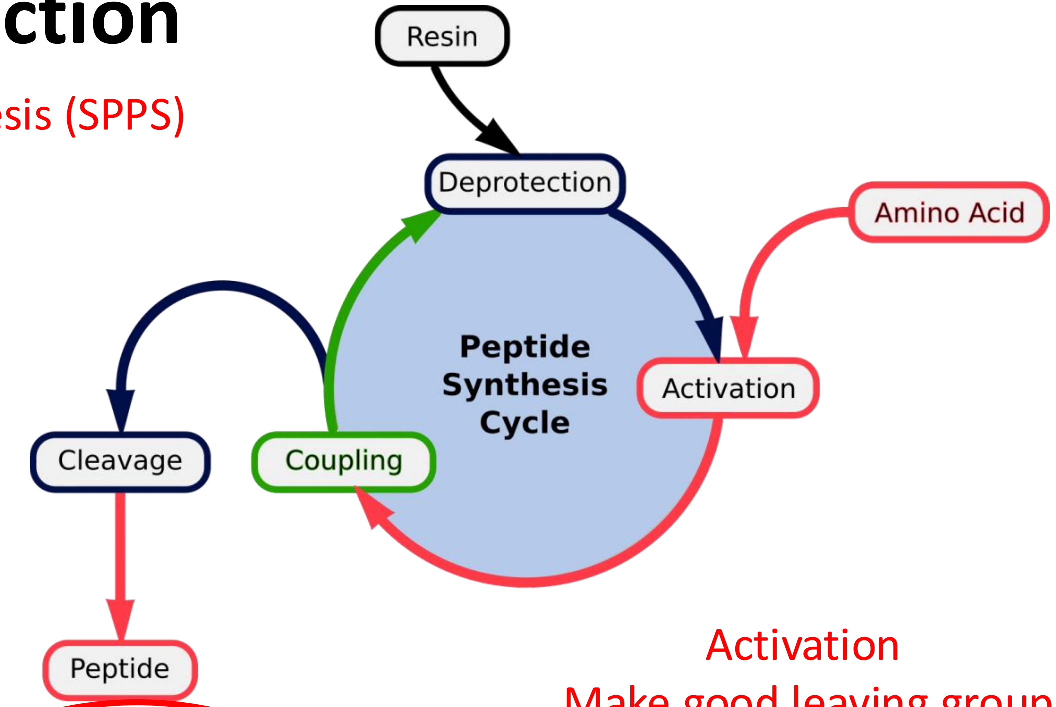
Peptide Production

Solid Phase Peptide Synthesis (SPPS)

Direction of synthesis

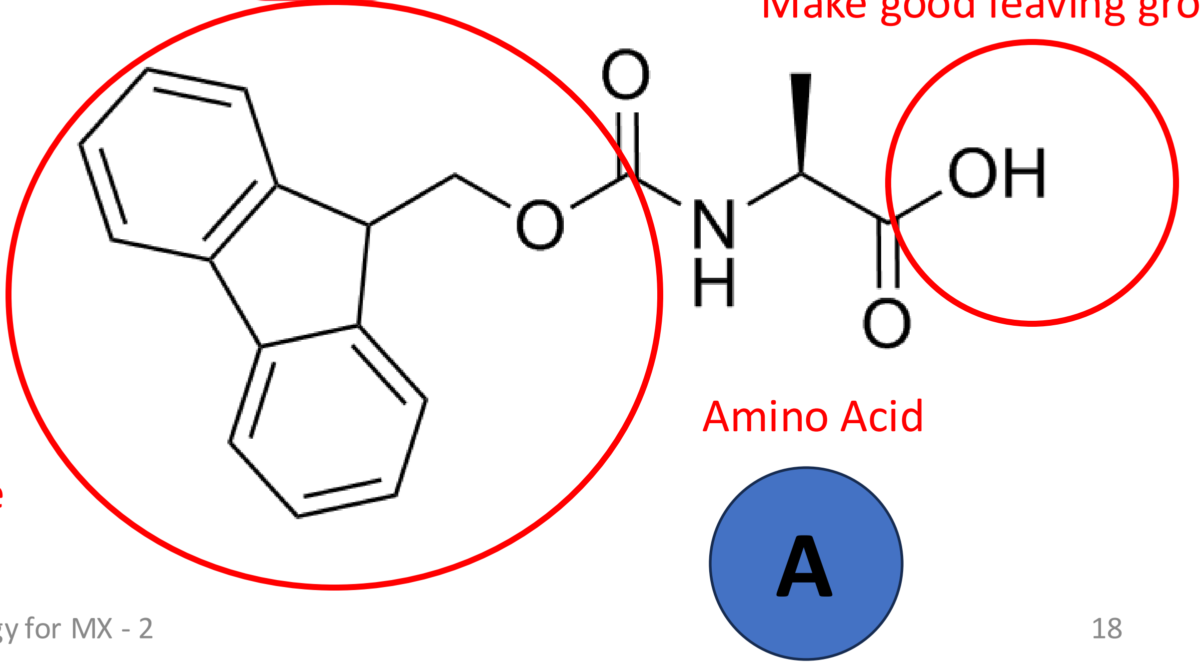


Last step: cleave from resin and remove all side chain protective groups



Activation
Make good leaving group

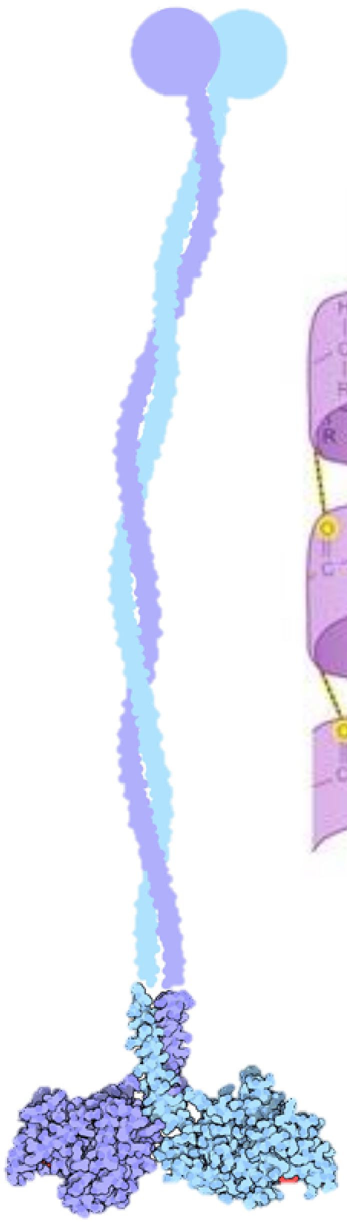
Fmoc
protecting
group
Base labile

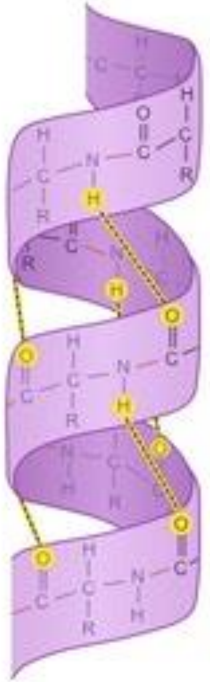


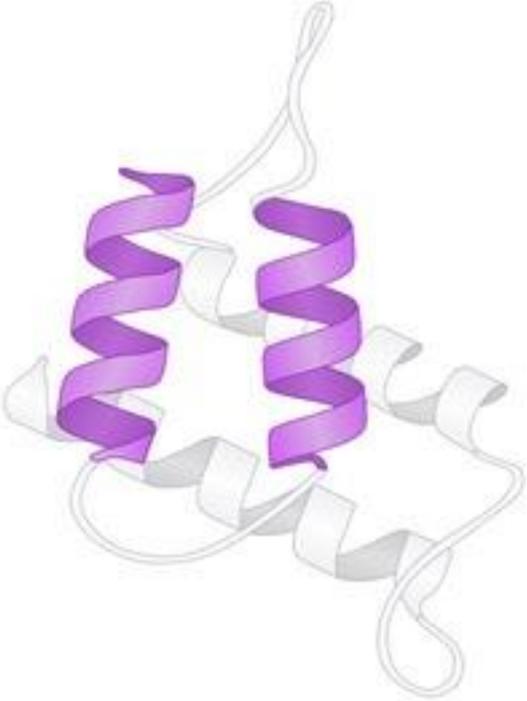
Amino Acid

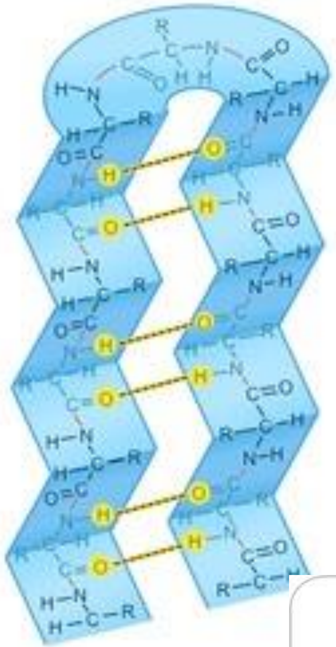
Helix and Sheet

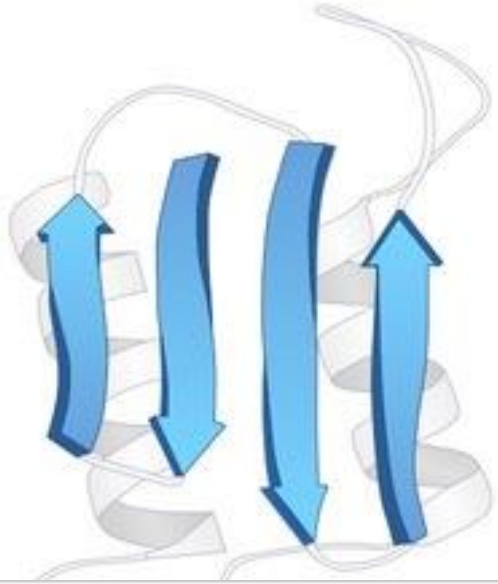
hydrogen bonding between the amine and carboxyl groups of *non-adjacent* amino acids

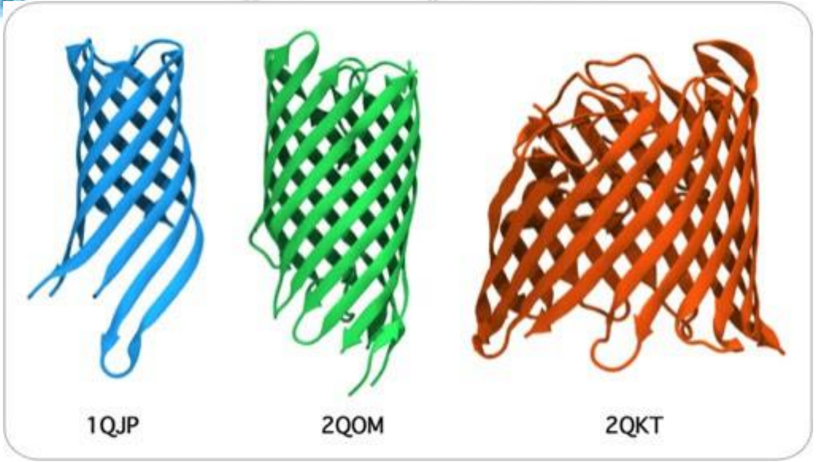






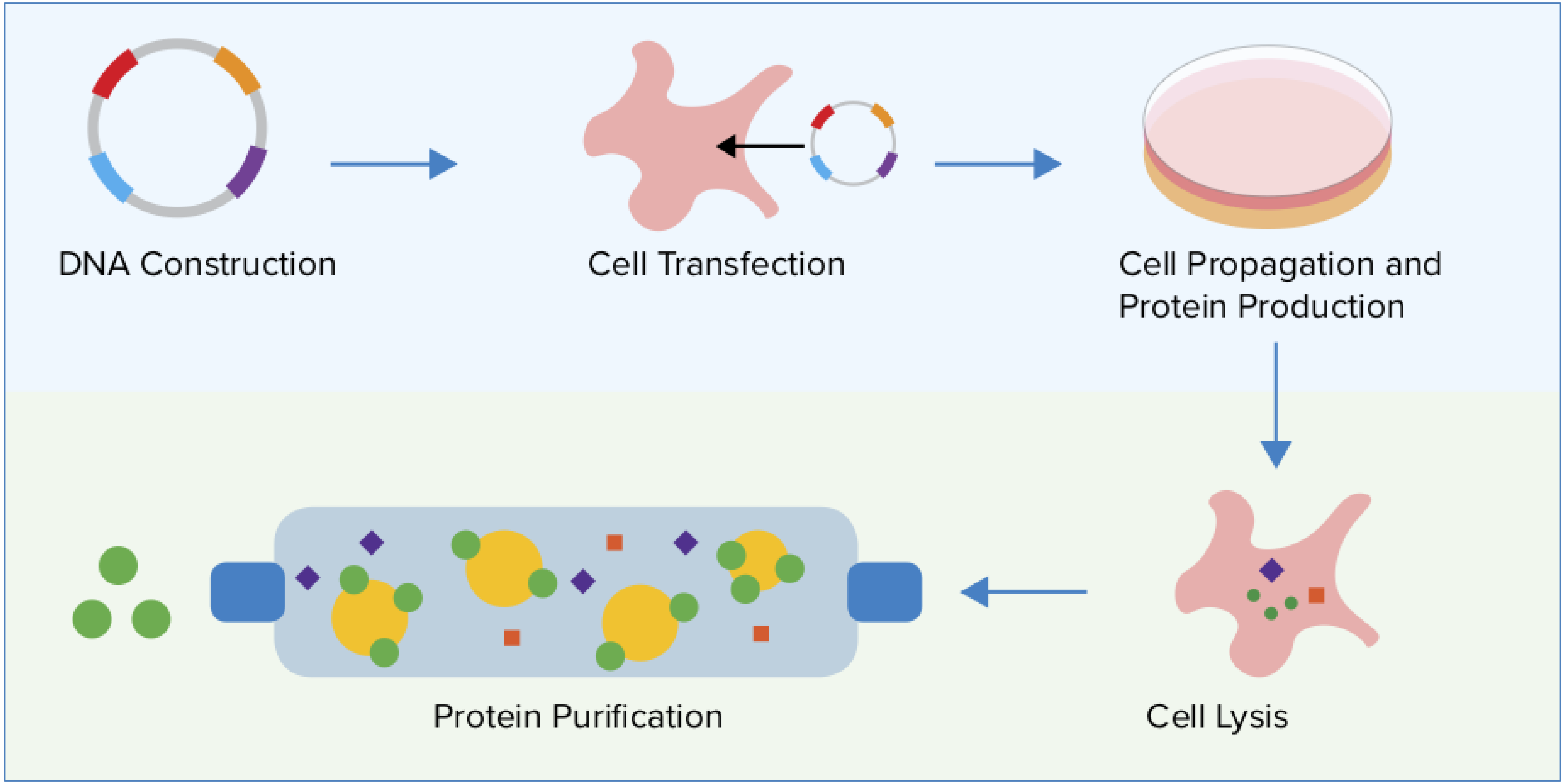






How to make a protein in the lab?

Recombinant proteins are **proteins** encoded by **recombinant** DNA that has been cloned in an expression vector that supports expression of the gene.

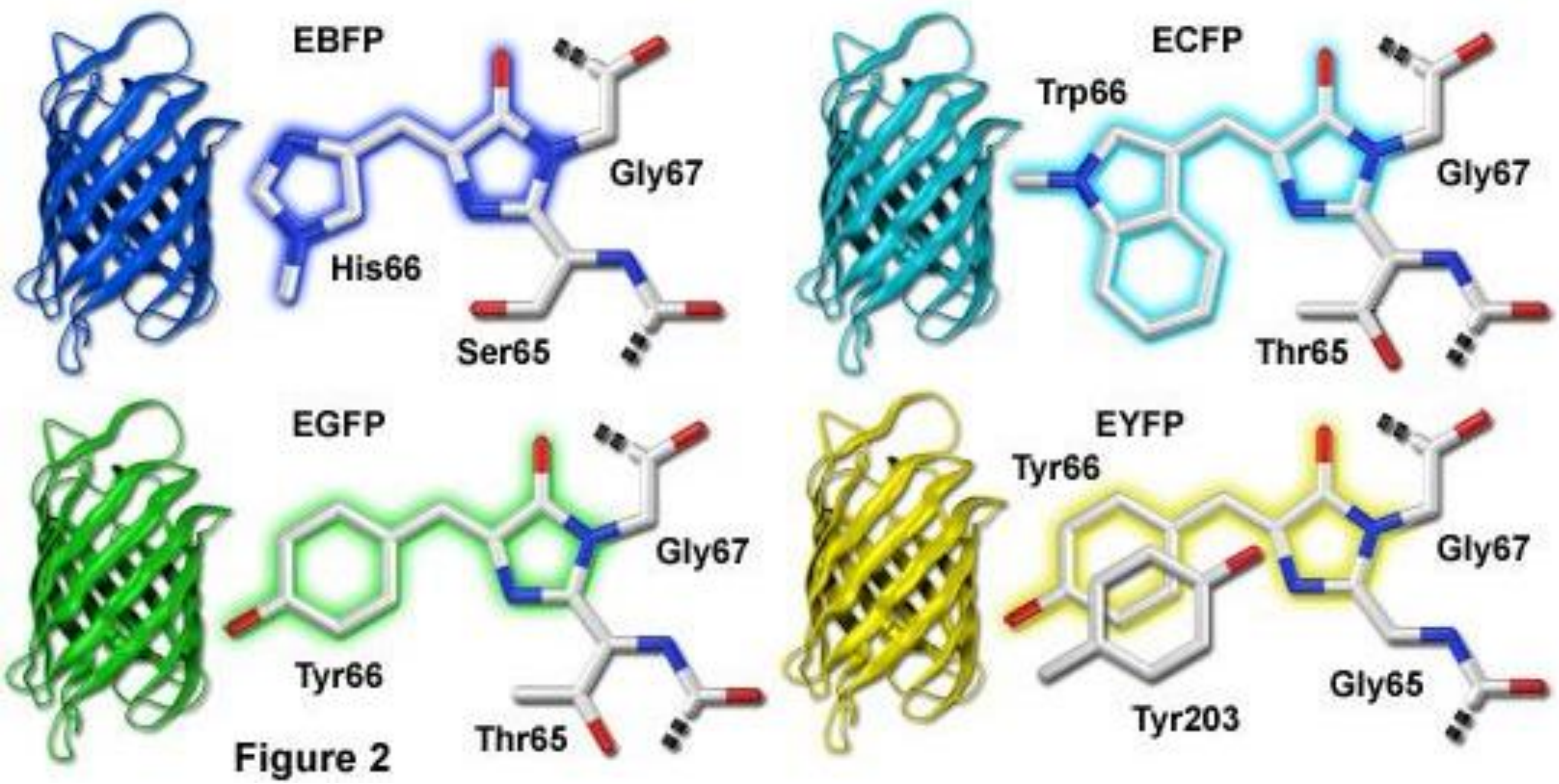


https://blog.addgene.org/plasmids-101-protein-expression

Protein Engineering – Mutations

Protein engineering is the conception and production of unnatural polypeptides, often through modification of amino acid sequences that are found in nature.

Change the color by mutations of specific amino acids
" point mutations "



Protein Engineering – Mutations

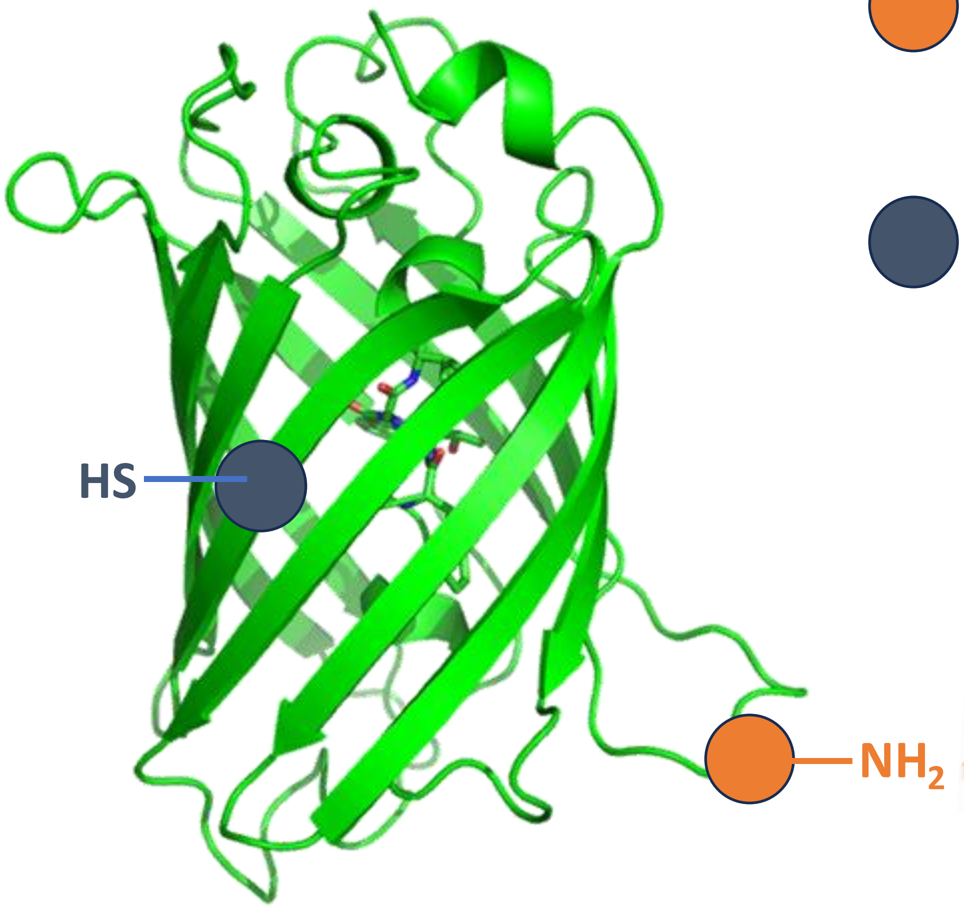
Chemically interesting mutations:

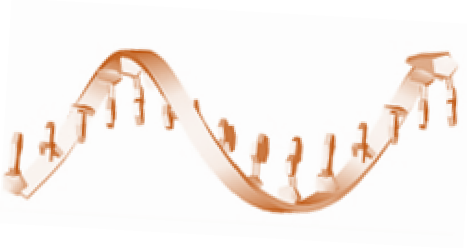
- Identify an amino acid on a random section

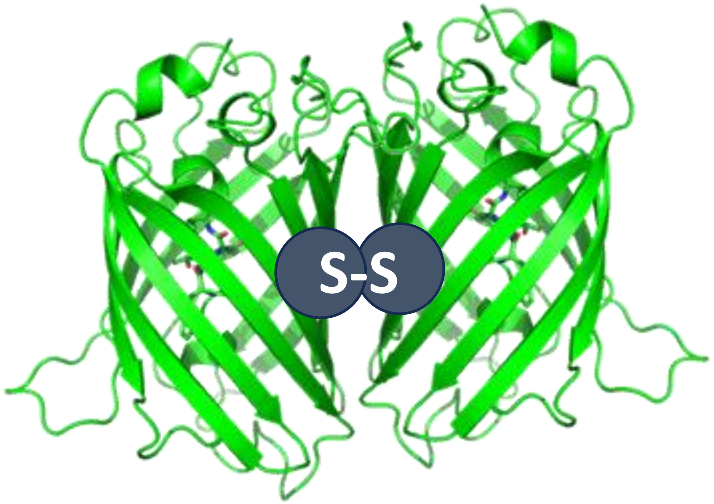
  Change to Lysine → chemical ligations to other molecules

- Identify an amino acid on a outward facing surface

  Change to Cysteine → Induce sulphor bridge = dimerization



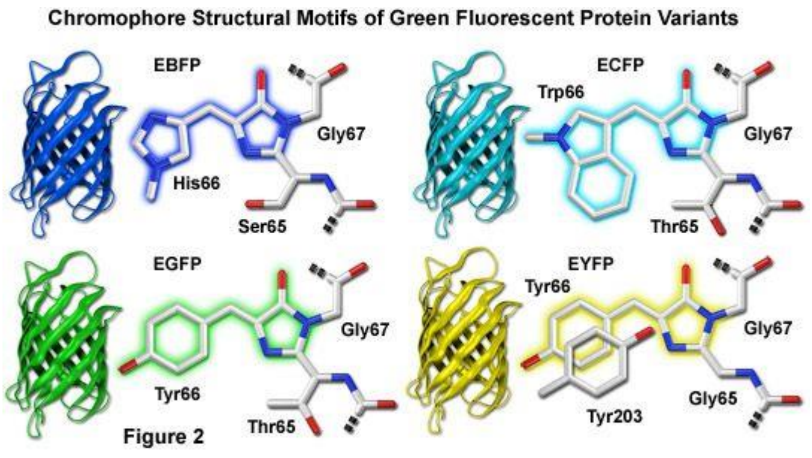




Protein Engineering – What?

Protein engineering is the conception and production of unnatural polypeptides, often through modification of amino acid sequences that are found in nature.

Synthetic **protein** structures and functions can now be **designed** entirely on a **computer** or produced through **directed evolution** in the laboratory.



Directed Evolution

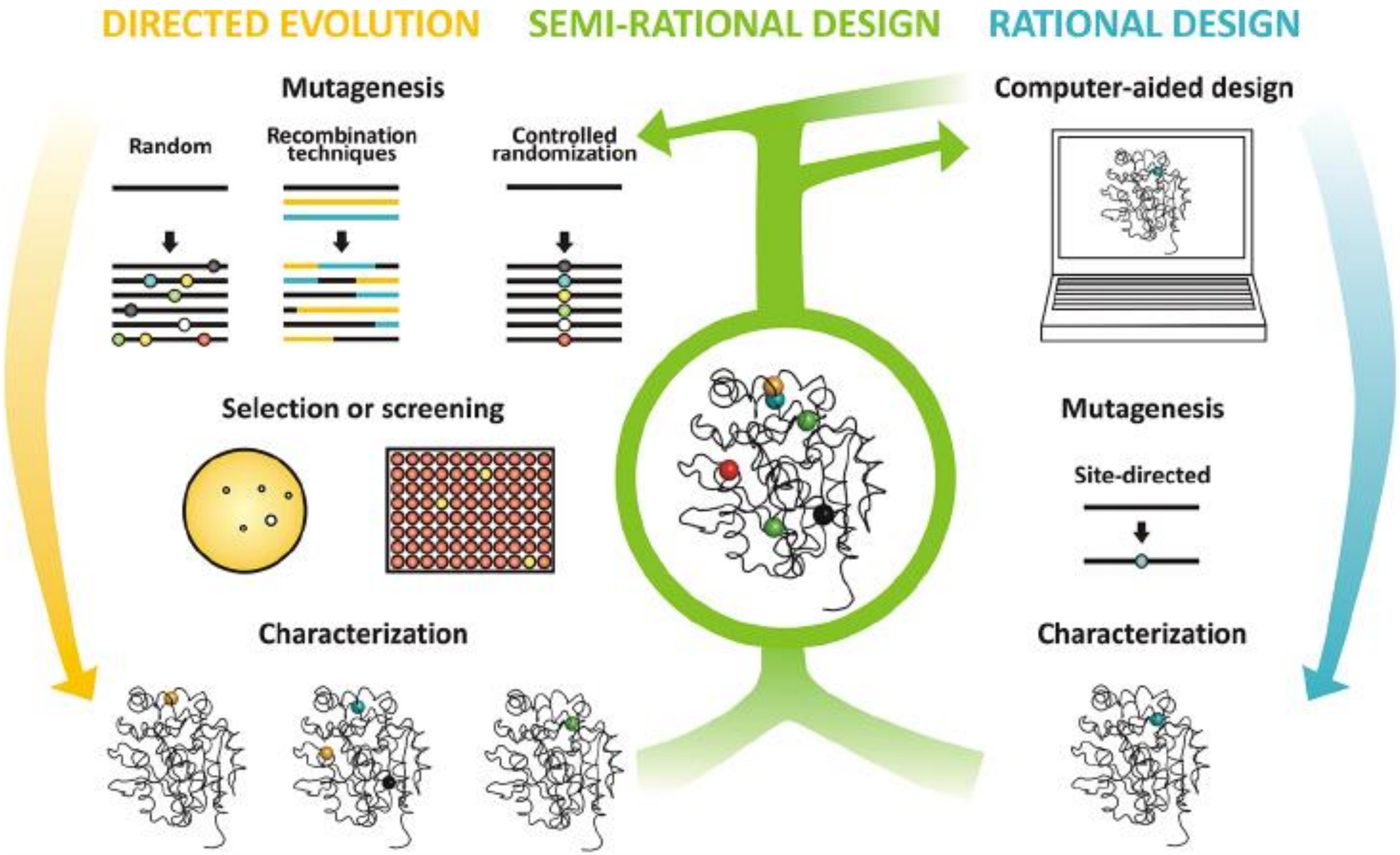
In **directed evolution**, random mutagenesis (by error-prone PCR or sequence saturation mutagenesis) is applied to a protein, and a **selection regime** is used to select variants having **desired traits**.

Further rounds of mutation and selection are then applied which mimics natural evolution and produces superior results (in function) to rational design.

Rational Design

In **rational protein design**, a scientist uses detailed knowledge of the structure and function of a protein to make desired changes.

The drawback is that **detailed structural knowledge** of a protein is often **unavailable**, and it can be difficult to predict the effects of mutations since structural information provides a **static picture** of the structure.



DO - 10.1021/cs400684x

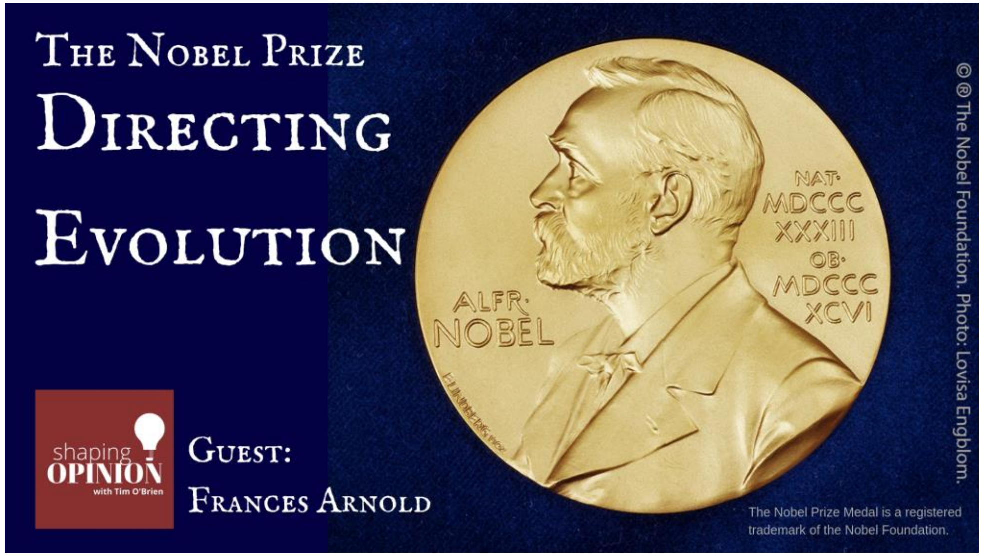
2018 Nobel Prize: Frances Arnold

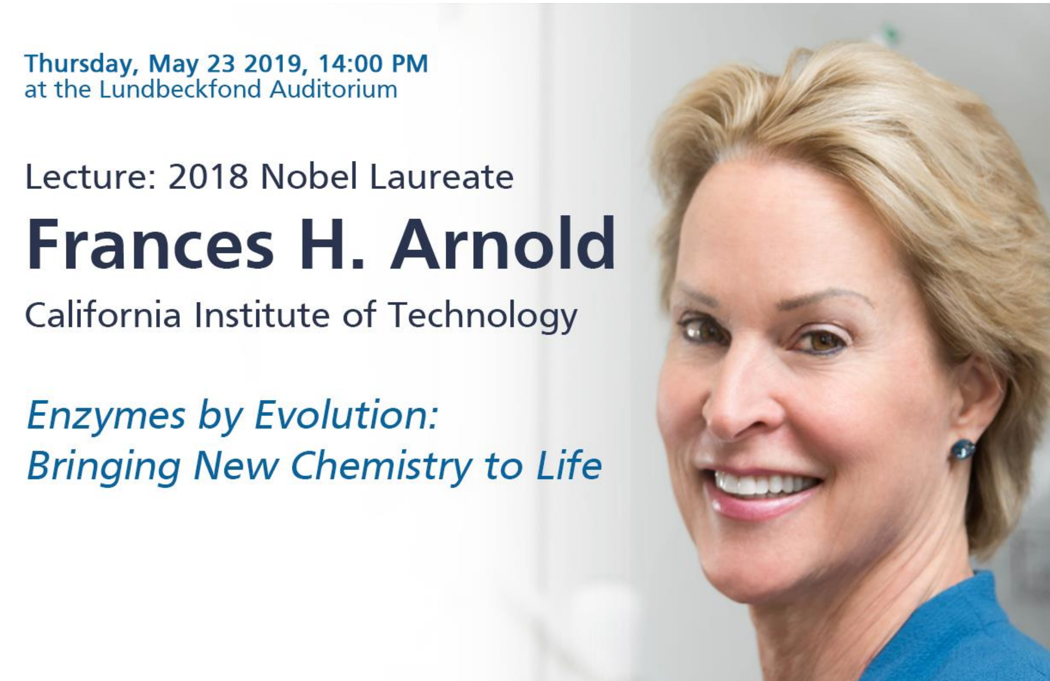
Evolution – the adaption of species to different Environments – has created an enormous diversity of life.

Frances Arnold has used the same principles – genetic change and selection – to **develop proteins that solve humankind's chemical problems.**

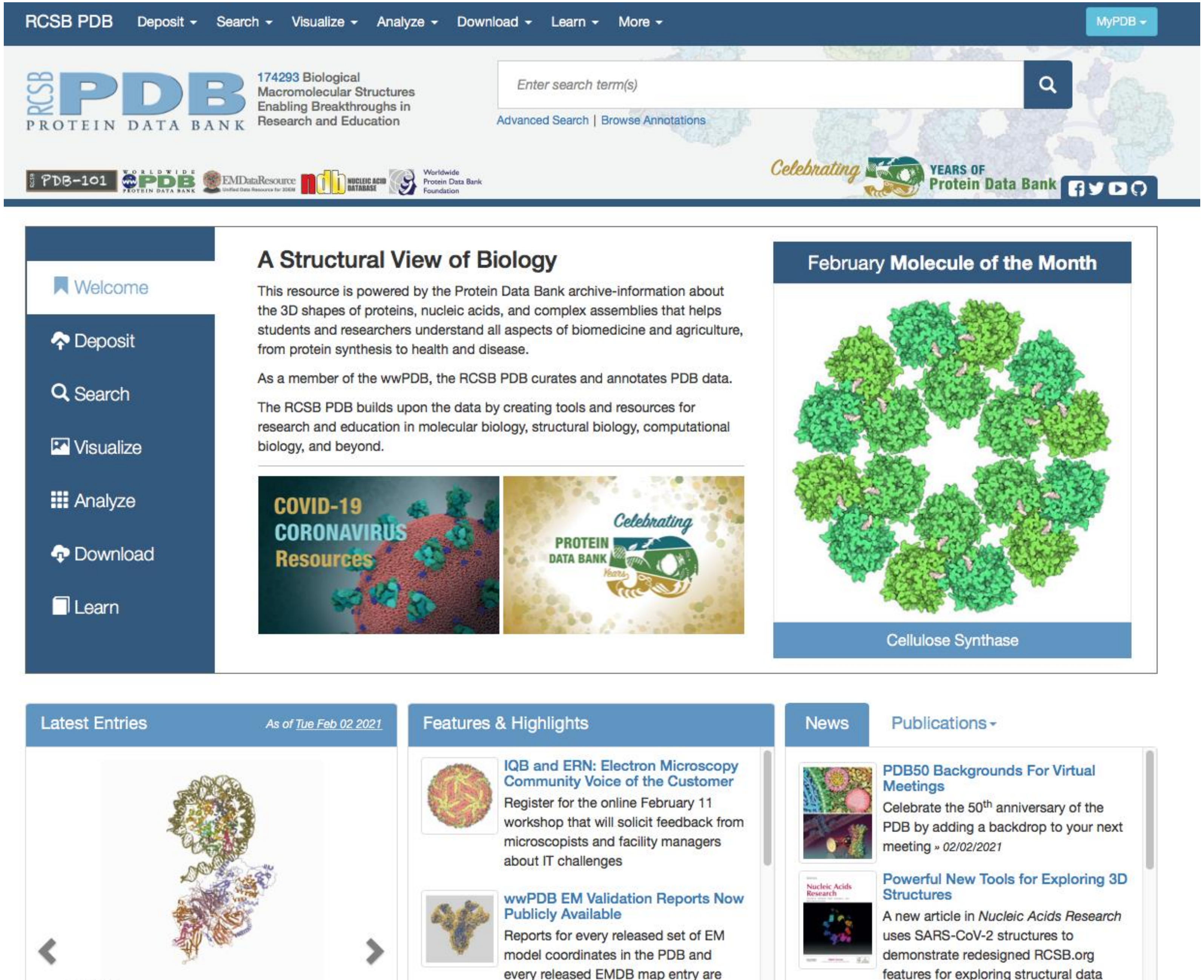
In 1993, Arnold conducted the **first directed evolution of enzymes**, which are proteins that catalyze chemical reactions.

The uses of her results include more environmentally friendly manufacturing of chemical substances, such as pharmaceuticals, and the production of renewable fuels.





Helpful Tools: The Protein Databank



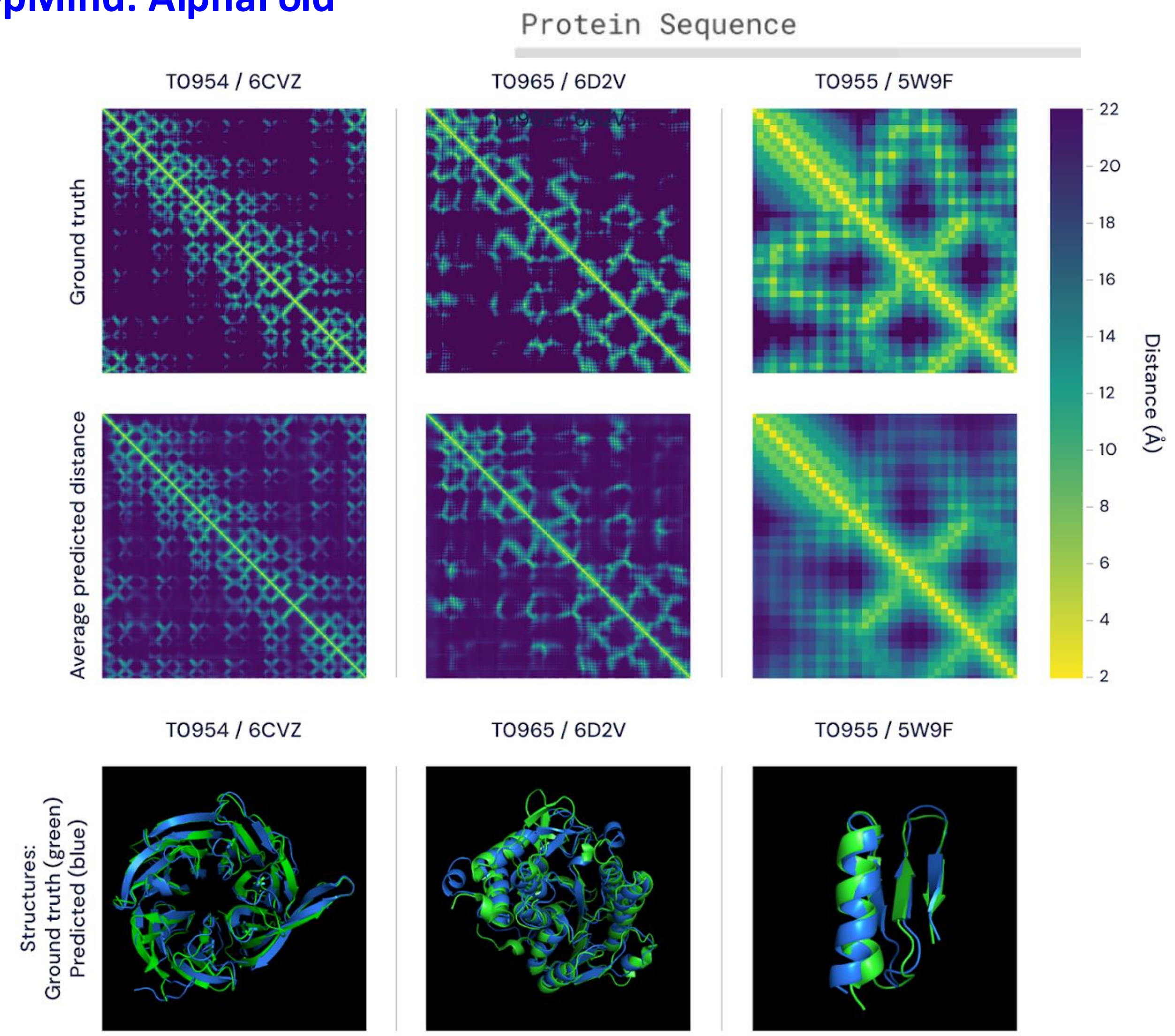
Prediction of Folding: AI in Biology

Google DeepMind: AlphaFold

AlphaFold works in two steps

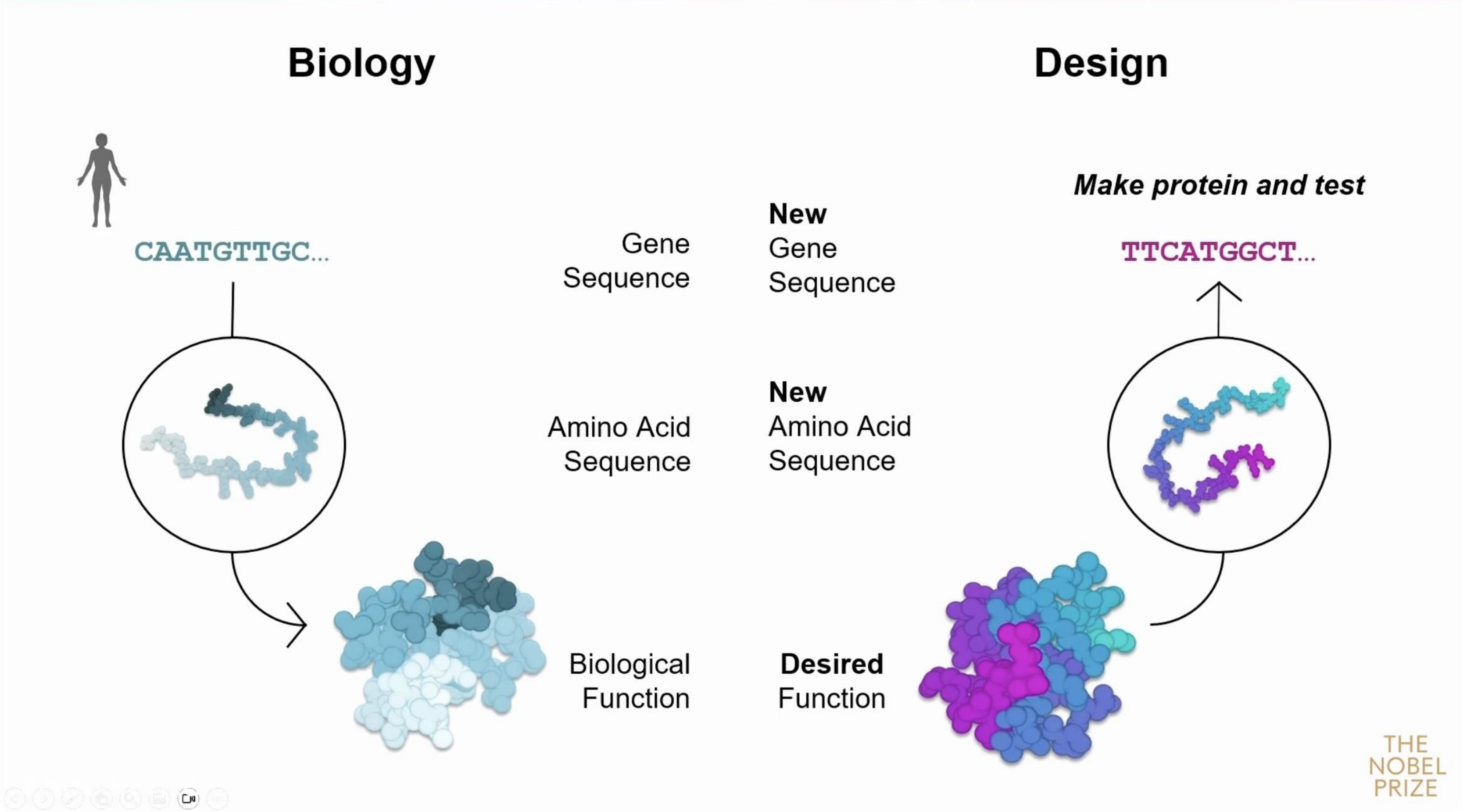
1) multiple sequence alignments: comparison of protein's sequence with similar ones in a **database** to reveal **pairs** of amino acids that don't lie next to each other in a chain, but that tend to appear in tandem. DeepMind trained a **neural network** to take such pairings and predict the distance between two paired amino acids in the folded protein.

2) Create a physically possible — but nearly random — **folding arrangement** for a sequence. Instead of another neural network, it used an optimization method called **gradient descent** to iteratively refine the structure so it came close to the (not-quite-possible) predictions from the first step.

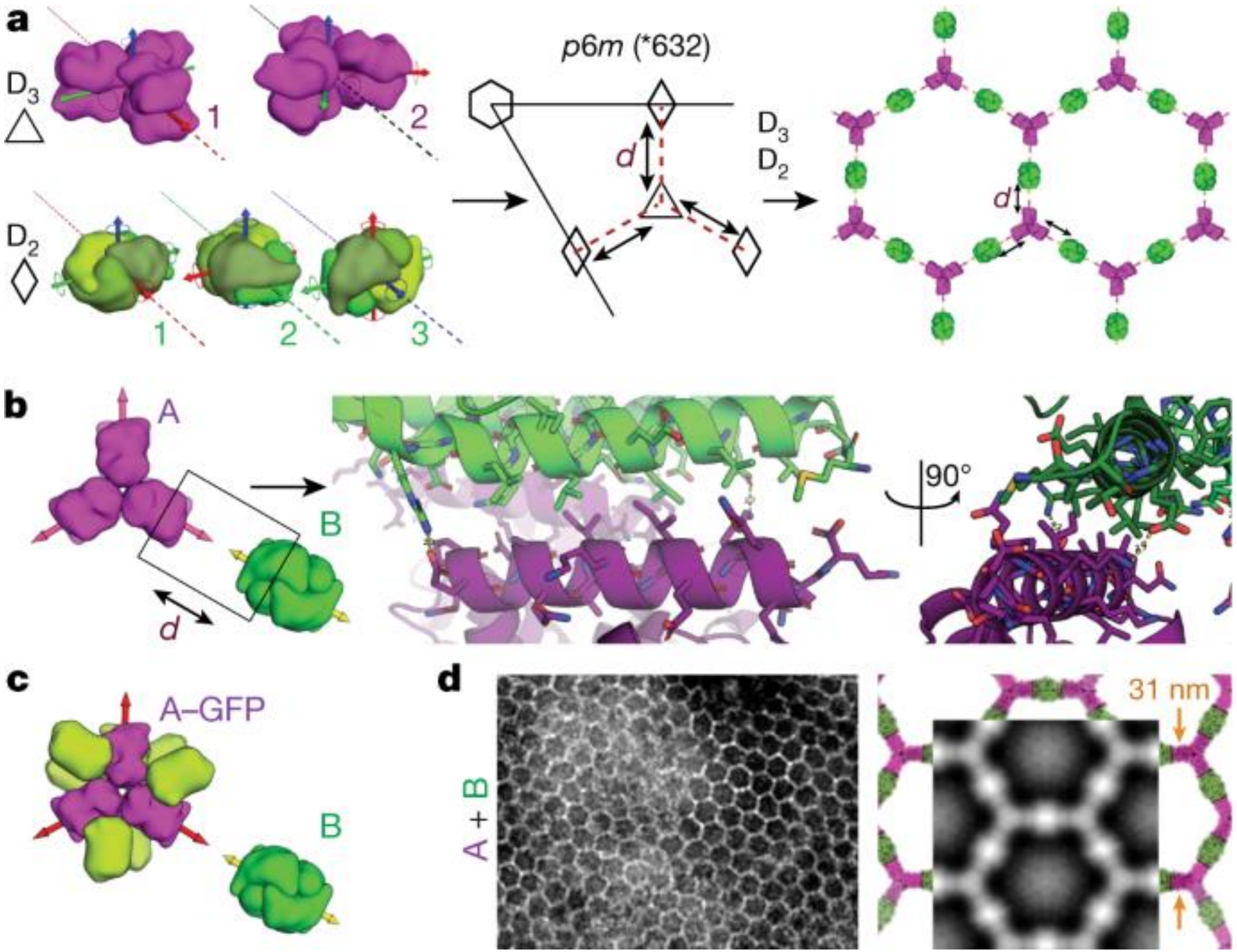


Nature **588**, 203-204 (2020)
doi: *https://doi.org/10.1038/d41586-020-03348-4*

2024 Nobel Prize: David Baker



2D Crystalline Protein Surfaces



Crystalline binary layers by designing rigid interfaces between pairs of dihedral protein building blocks, in a *p6m* lattice.

Because the material is designed from the ground up, the components can be readily **functionalized** and their **symmetry** reconfigured, enabling formation of ligand arrays with distinguishable surfaces.

With the rapid developments in de novo design of protein building-blocks and quantitative microscopy techniques, this provides the basis for a future of programmable biomaterials for synthetic and living systems.

Nature 589, 468–473(2021)

Conclusion

Proteins are extremely important macromolecules that enable life and show remarkable diversity in properties and function.

They can be easily engineered and produced with relative low cost in the lab.

This enables a central use of proteins as material tools to create modular architectures with on-demand function.

With DNA as a template for synthesis, all copies of a protein are identical, providing uniformity in materials building blocks.

Through recent advances in artificial intelligence, de novo protein design will likely experience a huge jump in performance the next year.