AMIDD 2024 Lecture 4: Proteins as drug targets

Ill. Niklas Elmehed © Nobel Prize Outreach

John J. Hopfield

Ill. Niklas Elmehed © Nobel Prize Outreach

Geoffrey E. Hinton

The Nobel Prize in Physics 2024: "for foundational discoveries and inventions that enable machine learning with artificial neural networks"

Ill. Niklas Elmehed © Nobel Prize Outreach

David Baker

Ill. Niklas Elmehed © Nobel Prize Outreach

Demis Hassabis

Ill. Niklas Elmehed © Nobel Prize Outreach

John M. Jumper

The Nobel Prize in Chemistry 2024 was divided, one half awarded to David Baker "for computational protein design", the other half jointly to Demis Hassabis and John M. Jumper "for protein structure prediction"

Dr. Jitao David Zhang, Computational Biologist

[1] Pharmaceutical Sciences, Pharma Research and Early Development, Roche Innovation Center Basel, F. Hoffmann-La Roche

[2] Department of Mathematics and Informatics, University of Basel

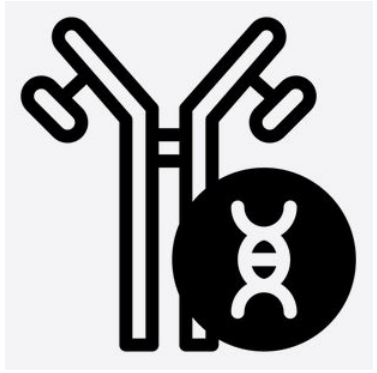
Topics of lecture 4

- The central dogma for drug discovery
- ODE-based mechanistic models
- Key considerations for a drug to work

Five key questions in drug discovery

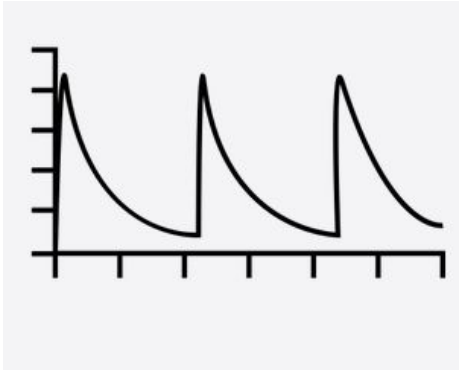
Medical Need

What is the unmet medical need to be addressed?



Target & modality

What is the target? What is the modality?

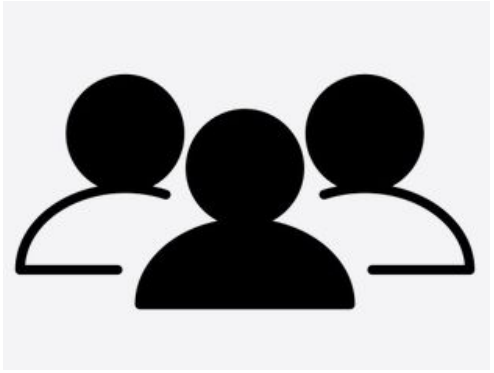


PK/PD

How much drug reach which body part? What does body do to the drug (PK)? What does the drug do to the body (PD)?

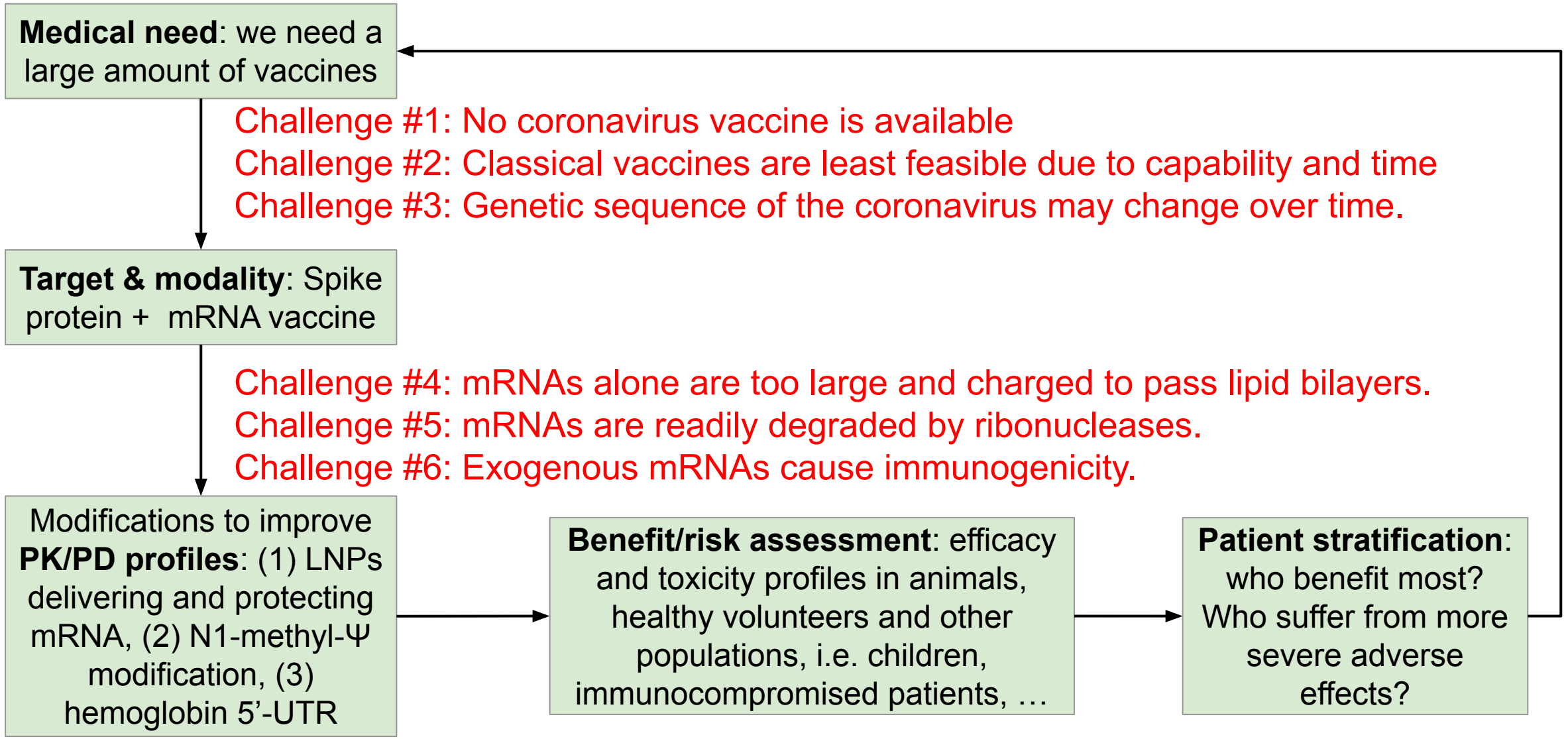
Benefit/risk

What is the toxicity of the drug? Is it justifiable given the benefits?



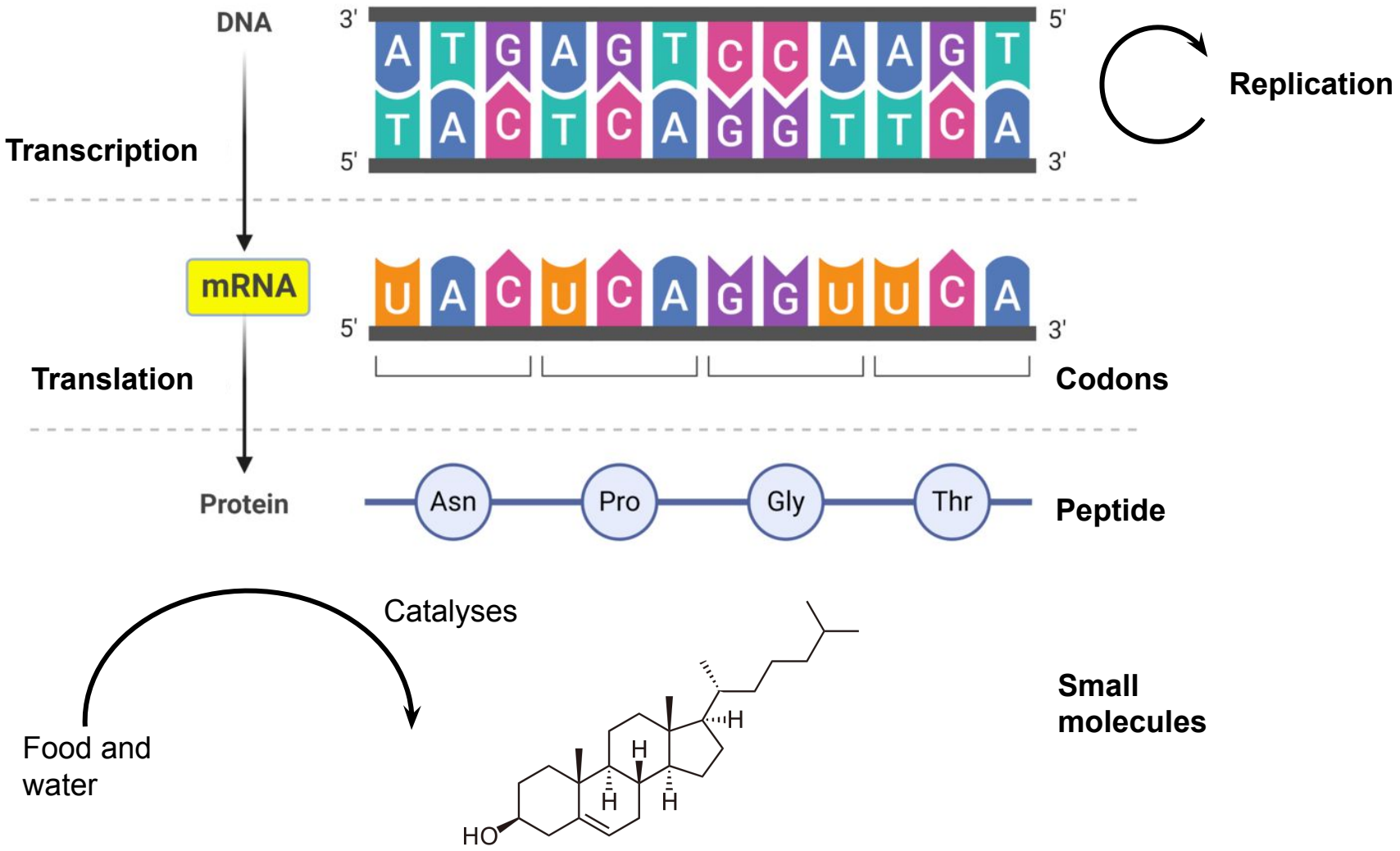
Patient stratification

Who are responsive to the drug? Who are susceptible to adverse events?

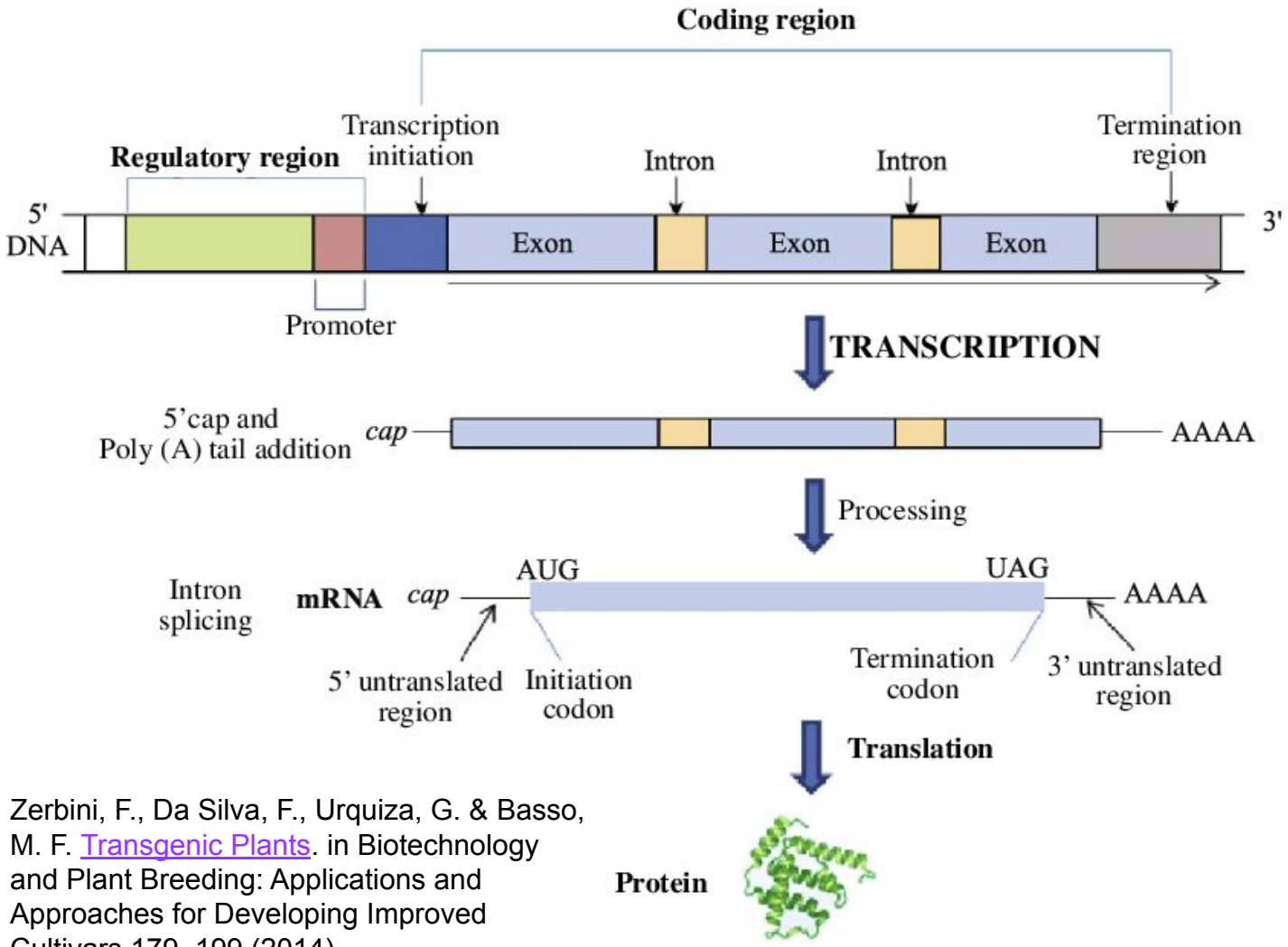


LNP: lipid nanoparticles

The central dogma



Coding sequence of the spike protein alone is not enough: mRNA transcription depends on 5’-UTR and 3’-UTR, too

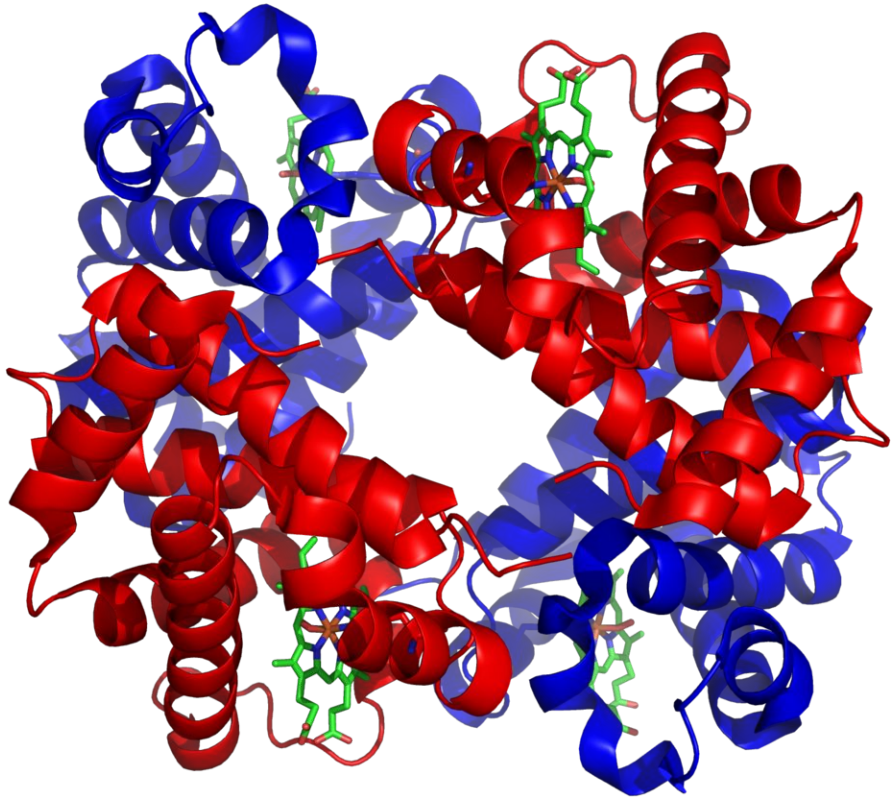


Zerbini, F., Da Silva, F., Urquiza, G. & Basso, M. F. Transgenic Plants. in Biotechnology and Plant Breeding: Applications and Approaches for Developing Improved Cultivars 179–199 (2014).

The process of gene expression in eukaryotes:

1. RNA polymerase, an enzyme, binds to the promoter region of the gene. It reads the DNA from the 5′ untranslated region (UTR) to the 3′ UTR to synthesize pre-mRNA.
2. Pre-mRNA receives a modified nucleotide (7-methylguanosine triphosphate) at the 5′ end as a cap, and a repeated adenine sequence (poly-A tail) at the 3′ end.
3. Pre-mRNA is spliced to remove introns. Mature mRNA contains the 5’ cap, 5’-untranslated region (5’-UTR), coding sequence, 3’-untranslated region (3’-UTR), and a poly-A tail.
4. Mature mRNA is transported from the nucleus to the cytoplasm for translation.

5’-UTR of human hemoglobin is a good choice to make sure that the vaccine sequence is stable and highly translated



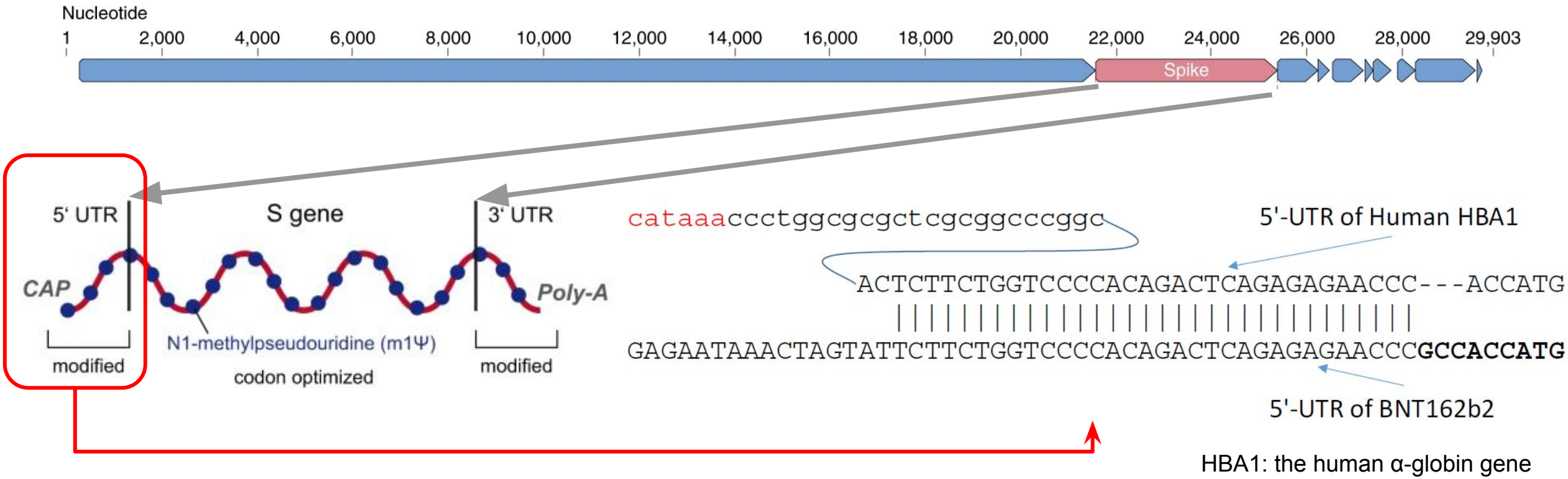
```
1    -MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFD------LSHGS   53  P69905   HBA_HUMAN
1    MVHLTPEEKSAVTALWGKVNV--DEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGN   58  P68871   HBB_HUMAN
1    MVHLTPEEKTAVNALWGKVNV--DAVGGEALGRLLVVYPWTQRFFESFGDLSSPDAVMGN   58  P02042   HBD_HUMAN
      : *:* :*: *.* **** .     *.*** *::: :* *: :*  *          *.

54   AQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAH  113  P69905   HBA_HUMAN
59   PKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHH  118  P68871   HBB_HUMAN
59   PKVKAHGKKVLGAFSDGLAHLDNLKGTFSQLSELHCDKLHVDPENFRLLGNVLVCVLARN  118  P02042   HBD_HUMAN
      :**.*****  *::::.:**:*::  ::: **:**..**:*** **:**.. *: .** .

114  LPAEFTPAVHASLDKFLASVSTVLTSKYR  142  P69905   HBA_HUMAN
119  FGKEFTPPVQAAYQKVVAGVANALAHKYH  147  P68871   HBB_HUMAN
119  FGKEFTPQMQAAYQKVVAGVANALAHKYH  147  P02042   HBD_HUMAN
     :  **** ::*: :*.:*.*:..*: **:
```

- Hemoglobin (left) is a protein that transports oxygen.
- Hemoglobin consists of three subunits: alpha, beta, and delta. They are encoded by three highly similar genes known as HBA, HBB, and HBD (above).
- Hemoglobin is present in erythrocytes (red blood cells) of almost all vertebrates.
- The protein is essential, therefore the mRNA is relatively stable and highly translated.

LNP, modified RNA, and 5'-UTR of HBA are all essential to make effective *and safe* vaccines against coronavirus



HBA1: the human α-globin gene

References: Heinz, Franz X., and Karin Stiasny. "Distinguishing Features of Current COVID-19 Vaccines: Knowns and Unknowns of Antigen Presentation and Modes of Action." Npj Vaccines 6, no. 1 (August 16, 2021): 1–13. https://doi.org/10.1038/s41541-021-00369-6; Assemblies of putative SARS-CoV2-spike-encoding mRNA sequences for vaccines BNT-162b2 and mRNA-1273 (github.com/NAalytics); Xia, Xuhua. "Detailed Dissection and Critical Evaluation of the Pfizer/BioNTech and Moderna MRNA Vaccines." Vaccines 9, no. 7 (July 3, 2021): 734. https://doi.org/10.3390/vaccines9070734.

A summary of what we have learned so far in the context of coronavirus

1. What is the unmet medical need to be addressed? We need a vaccine to prevent a large population of individuals from being infected by coronavirus, which have severe consequences.
2. What are the target(s) of our drug? Spike protein is conserved: immune reaction is desired.
3. Where should the drug go in patient's body, what does body do to the drug, and what does the drug do to the body? Thanks to LNP, N1-mythel-Ψ, and 5'-UTR of HBA1, the mRNA vaccination can enter cells with minimal side effects. In cells, spike protein RNA is synthesized into proteins, which are digested, presented, and elicit immune response.
4. What is the safety profile of the drug in light of its benefits? Initial study: Polack, F. P. et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. New England Journal of Medicine 383, 2603–2615 (2020), and watch this video.
5. Who are responsive to the drug, or susceptible to adverse events? Updated regularly by regulatory agencies, for instance European Medicines Agency

Interested in learning more? Read this report by WHO on potential benefits and limitations of mRNA vaccines.

Most drugs work by binding to and modulating protein targets

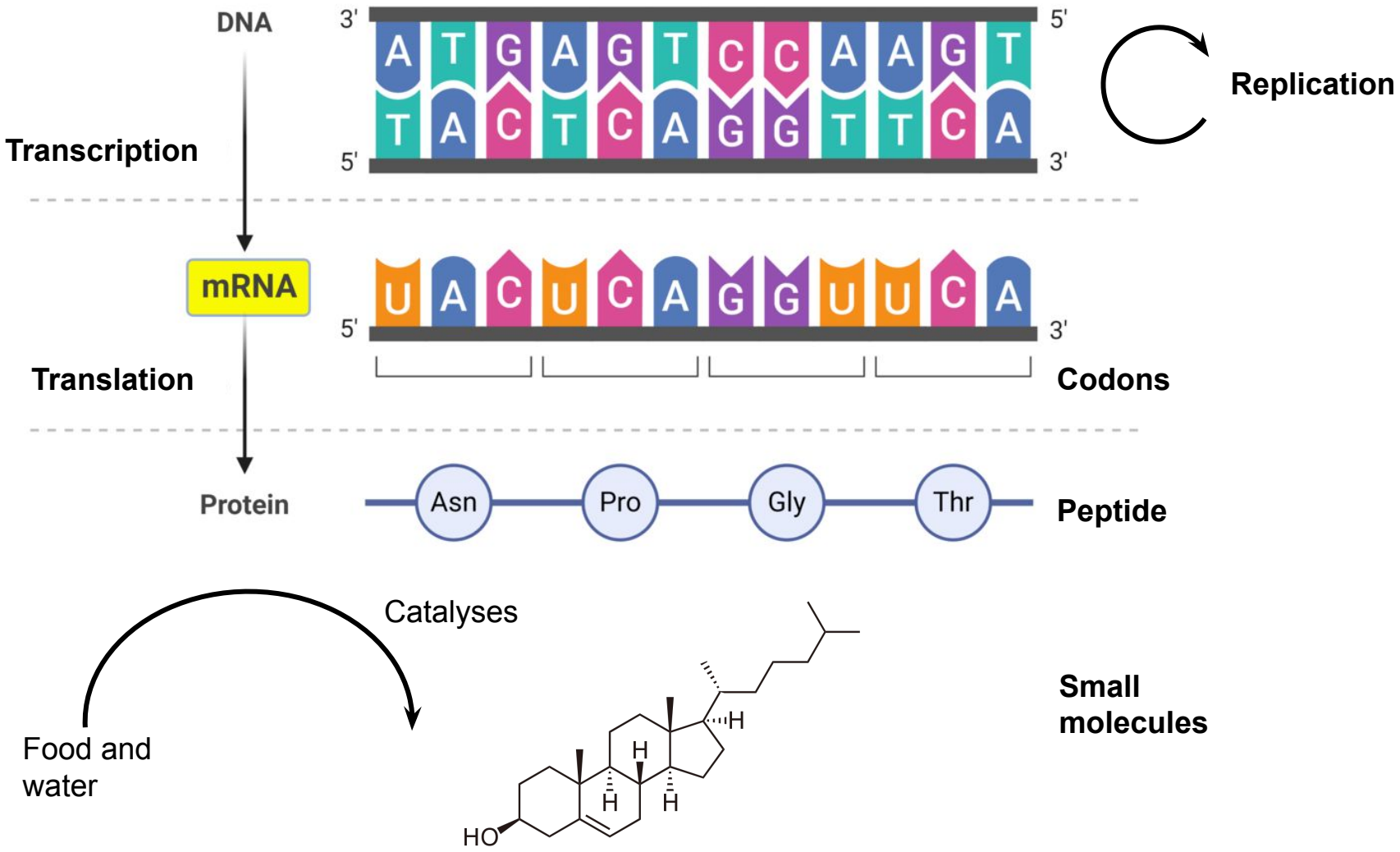
Table 1 | **Molecular targets of FDA-approved drugs**

Drug target class	Targets			Drugs		
	Total targets	Small-molecule drug targets	Biologic drug targets	Total drugs	Small molecules	Biologics
Human protein	667	549	146	1,194	999	195
Pathogen protein	189	184	7	220	215	5
Other human biomolecules	28	9	22	98	63	35
Other pathogen biomolecules	9	7	4	79	71	8

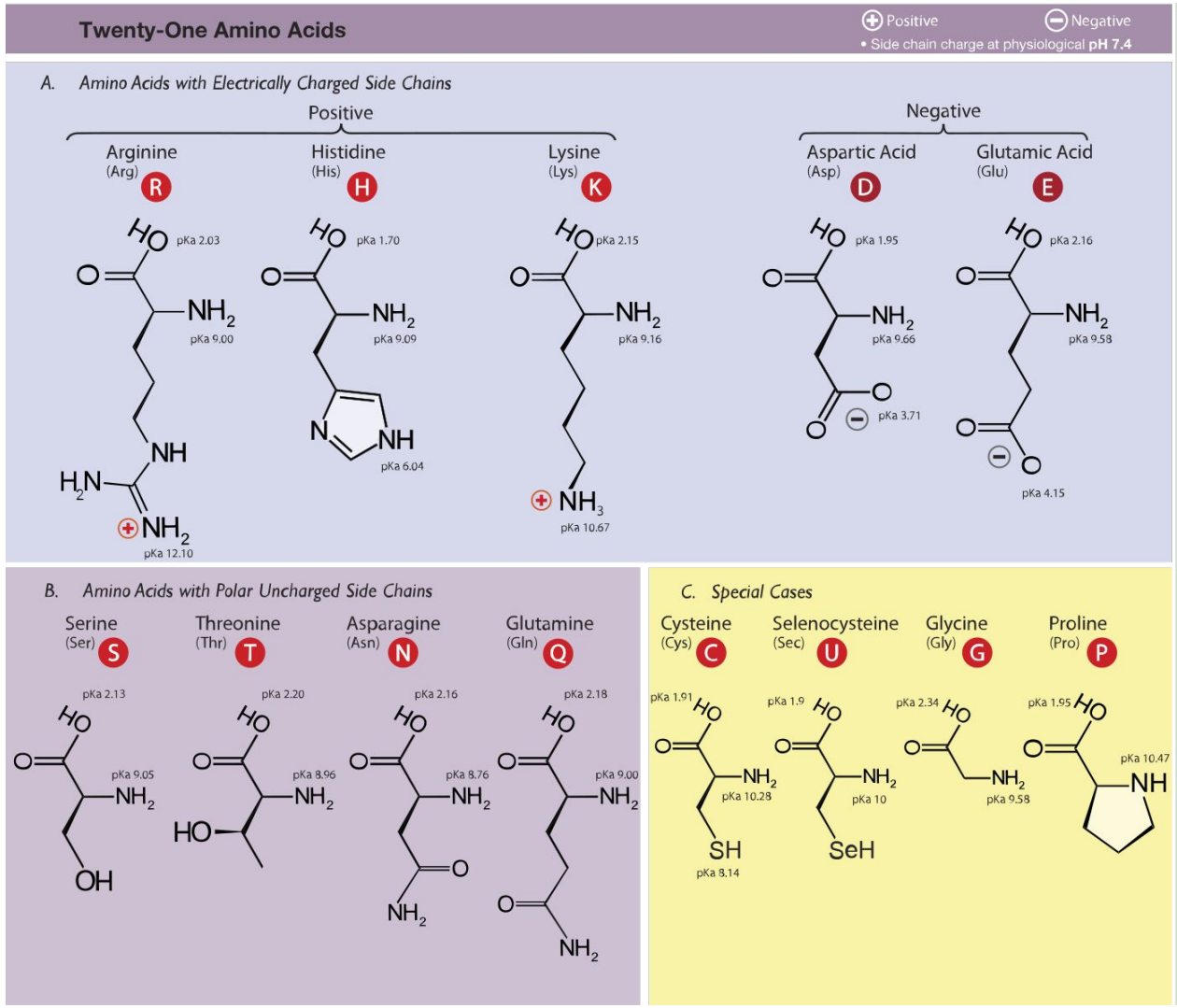
The list also includes antimalarial drugs approved elsewhere in the world.

End of lecture on 11.10.2024

The central dogma



Amino acids, the building blocks of proteins, form peptide bonds



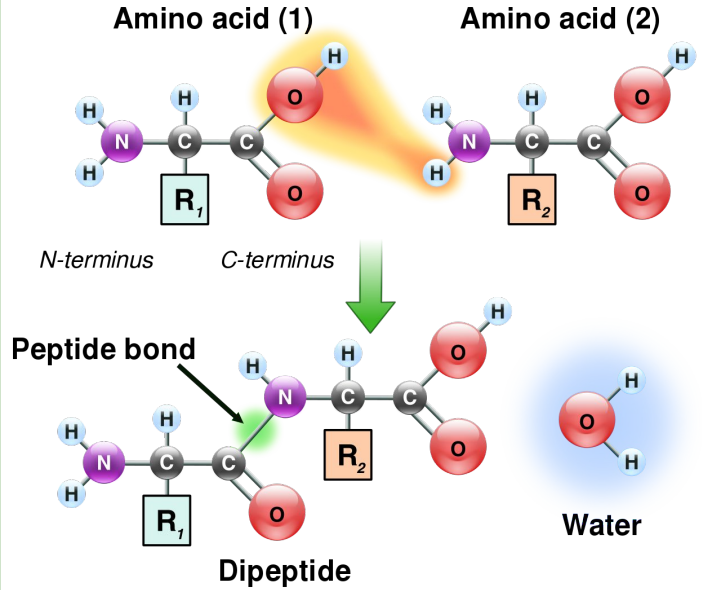
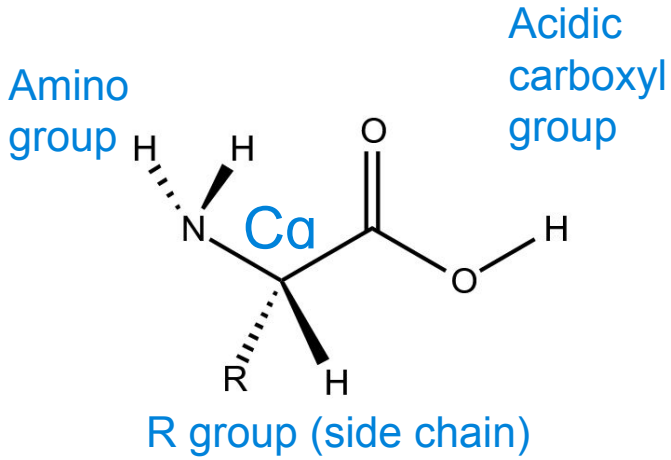
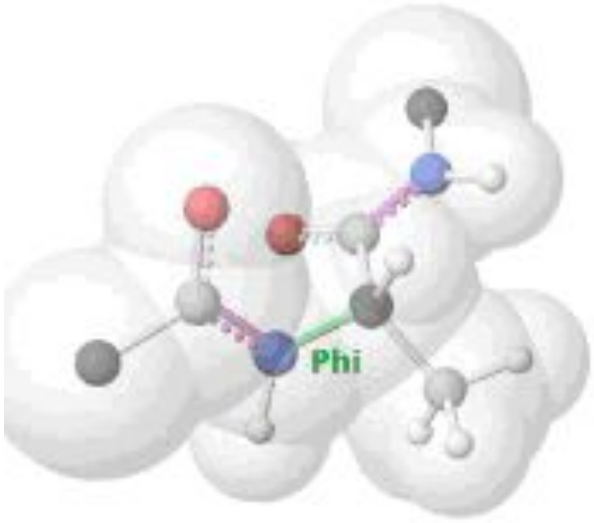


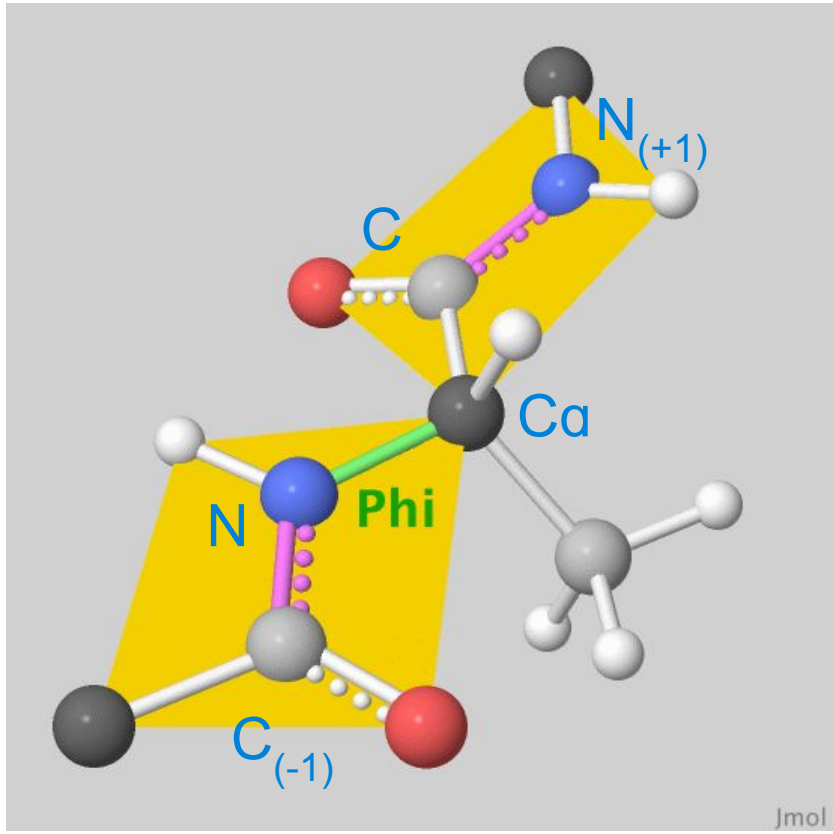
Figure by Dan Cojocari. Reused with CC license from wikimedia

Primary structure of proteins

- (Top left) Human proteins are chains of amino acids (AAs). The backbone remains the same while the side chain varies among AAs.
- (Right) The amino group and the carboxyl group of adjacent amino acids form peptide bonds. Proteins are therefore called *polypeptides*.
- C-Cα bonds Cα-N bonds can rotate at two *dihedral angles,* Ψ (psi) and ϕ (phi), respectively.
- (Bottom left) Due to steric collisions, not all combinations of Phi/Psi are possible.



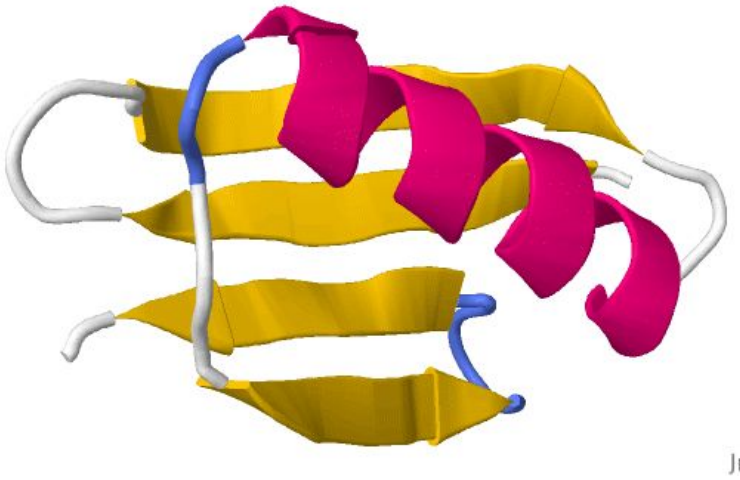


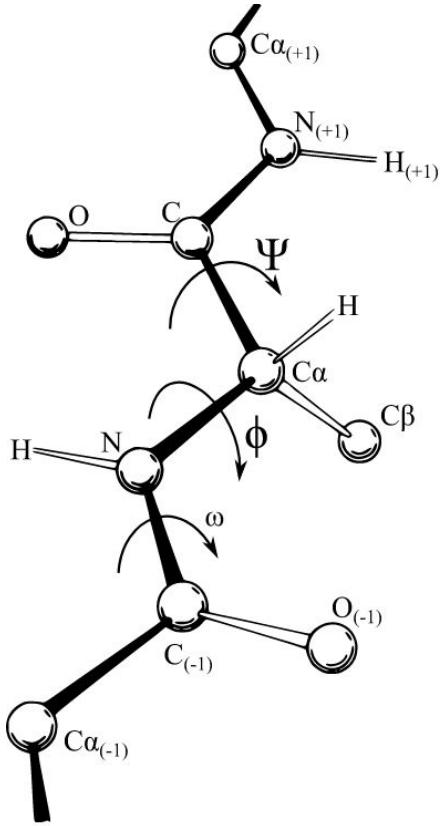


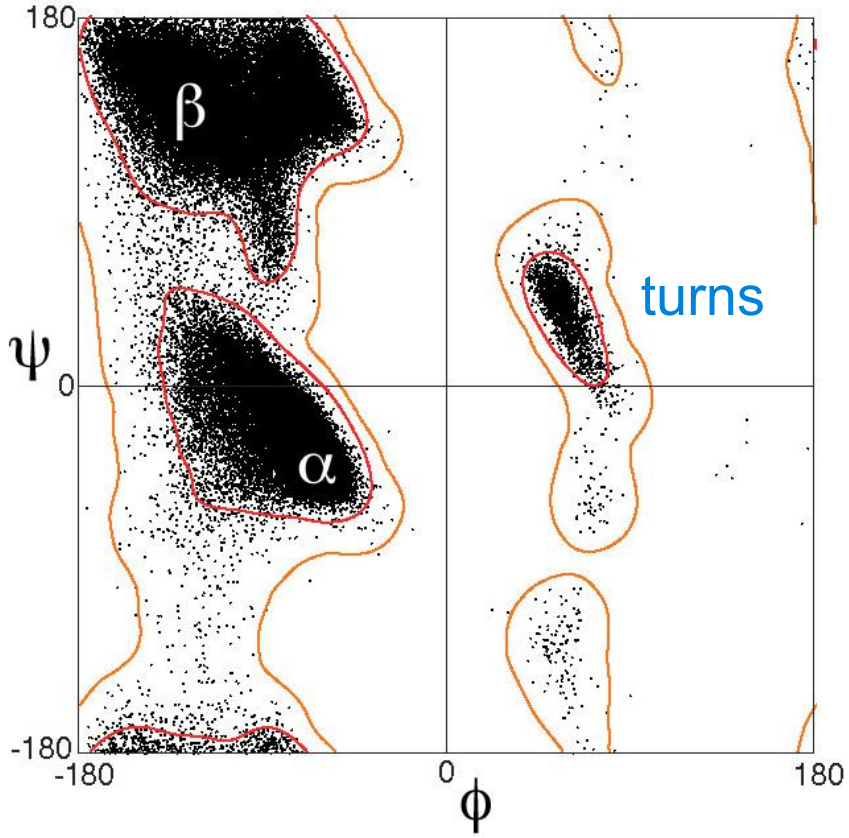
The Ramachandran Principle: alpha helices, beta strands, and turns are the most likely confirmations for a polypeptide

Most other conformations are impossible to due to clashes, known as *steric collisions*, between atoms.

To learn more about the topic, check out the YouTube video tutorial or the Slides by Eric Martz, and finish the Quiz.

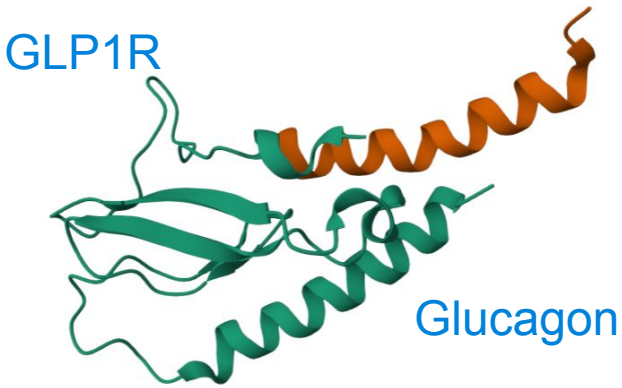




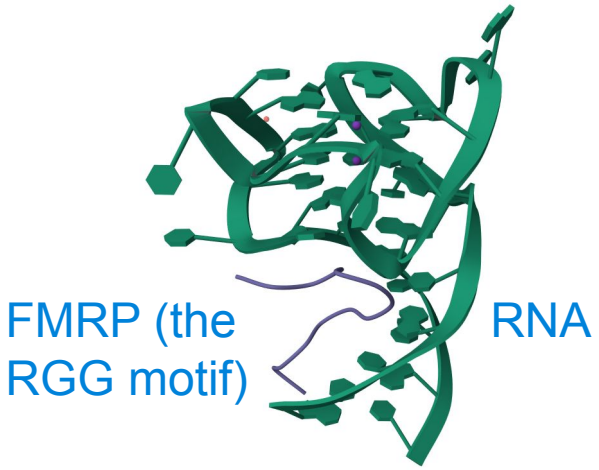


100,000 dots taken from high-resolution crystallographic structures. Wikimedia Commons courtesy Jane and David Richardson (Proteins 50:437, 2003). This plot excludes Gly, Pro, and pre-Pro.

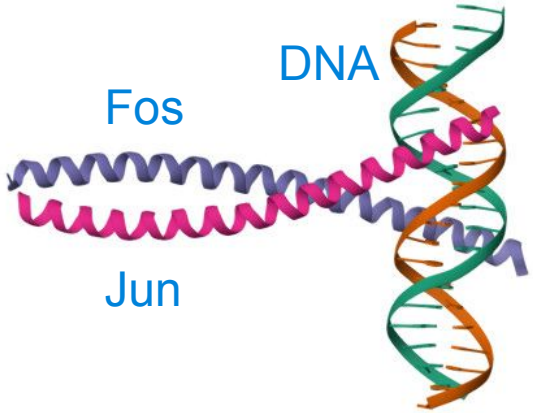
Proteins specifically and tightly bind to other molecules



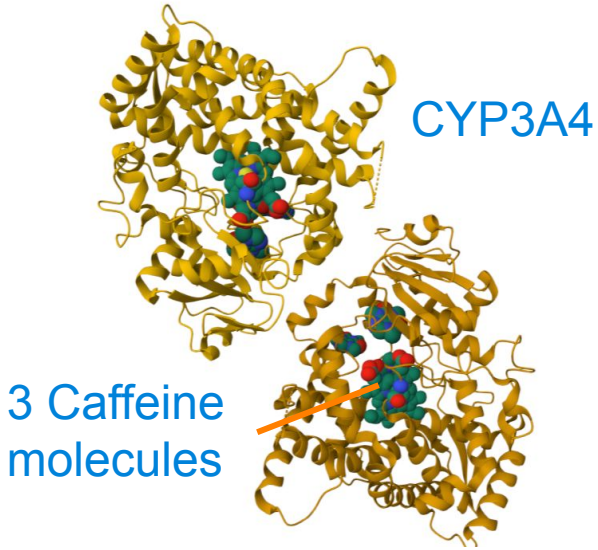
One protein binds to another protein PDB 3iol



Protein binds to RNA. Protein FMRP is encoded by gene *FMR1*. Mutations associated with *FMR1* induce the fragile X syndrome.PDB 5DE5



Protein complex binds to DNA. The complex Fos:Jun is known as AP-1, a transcription factor. PDB 1FOS.



Protein binds to small molecule. Cytochrome P450 3A4 (CYP3A4) is a major drug metabolizing enzyme, which also metabolizes caffeine. PDB 8so1

Major protein classes by functions

Top: an antigen presenting cell.
Bottom: a T cell. The red dot: a virus

Enzymes: catalysis of chemical reactions.

- To learn the basics of enzymes, watch the video *How Enzymes Work*.

Transporters: moving ions, small molecules, and proteins across membranes.

- To learn the basics of transporters and other ways cell transport material across membranes, Watch the video *Biology: Cell Transport*.

Receptors and kinases: signalling allows cells adapt to the environment.

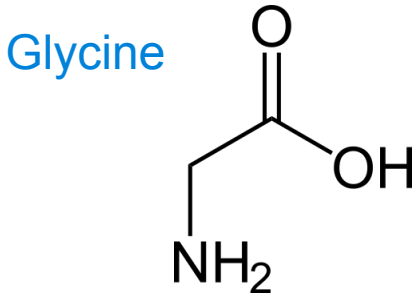
- To learn the basics of cellular signaling, watch the video *Common cell signaling pathway*.

Structural proteins: stiffness, rigidity, and mechanistical forces.

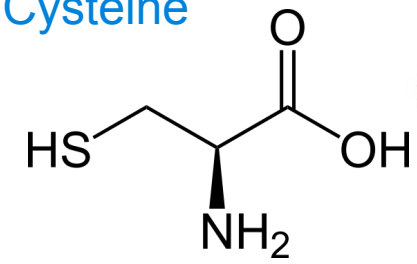
Figure: Immunological Synapse, David S. Goodsell, 2020

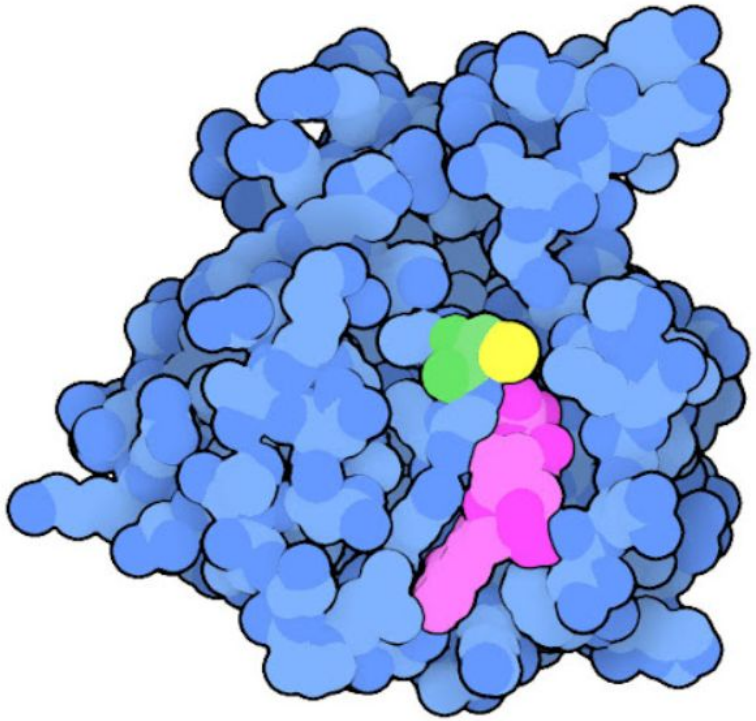
Some diseases are caused by perturbed functions of single protein

Glycine

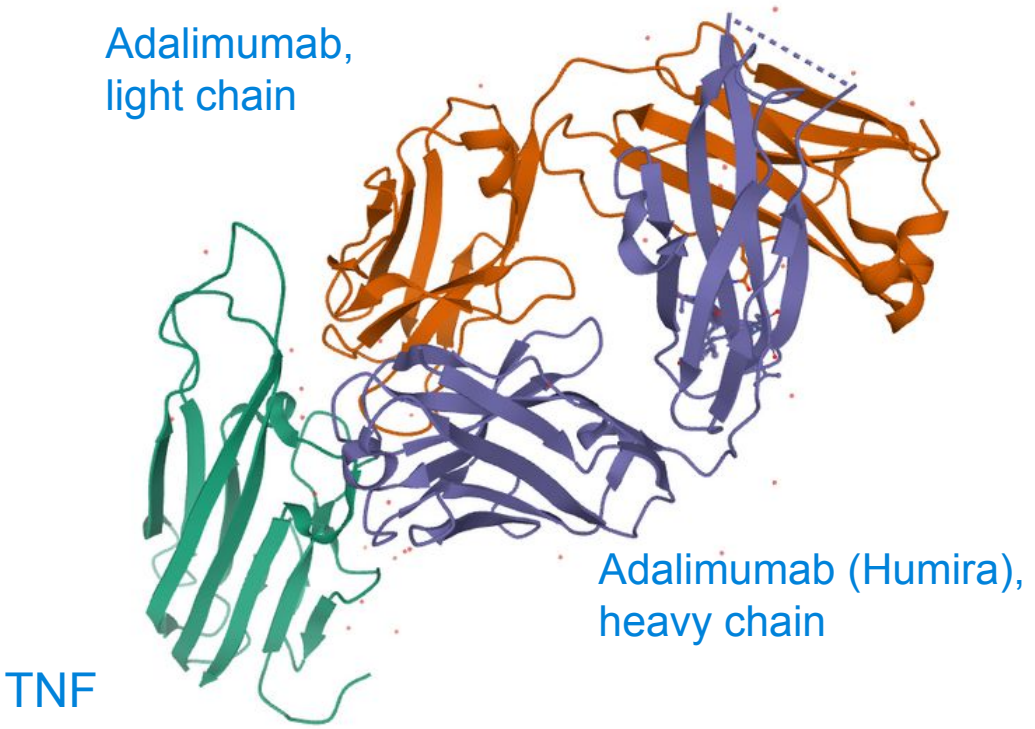


Cysteine





Mutation of glycine (G) to cysteine (M) at position 12 (green, with sulfur in yellow) in Ras protein leads to a protein that is continually activated. The structure of the oncogenic mutant (PDB ID 4ldj) reveals that the mutation modifies the interaction with GDP (magenta) and GTP, which act as the switch that turns the protein on and off.

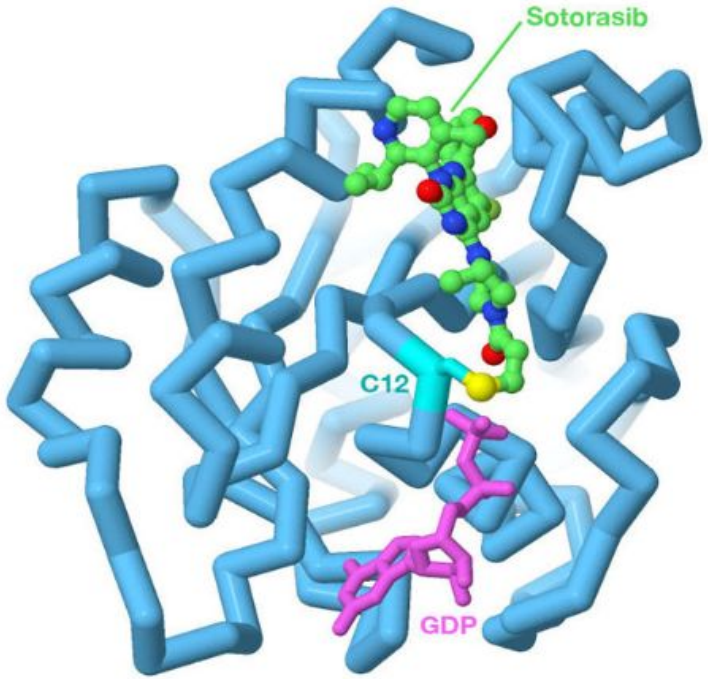


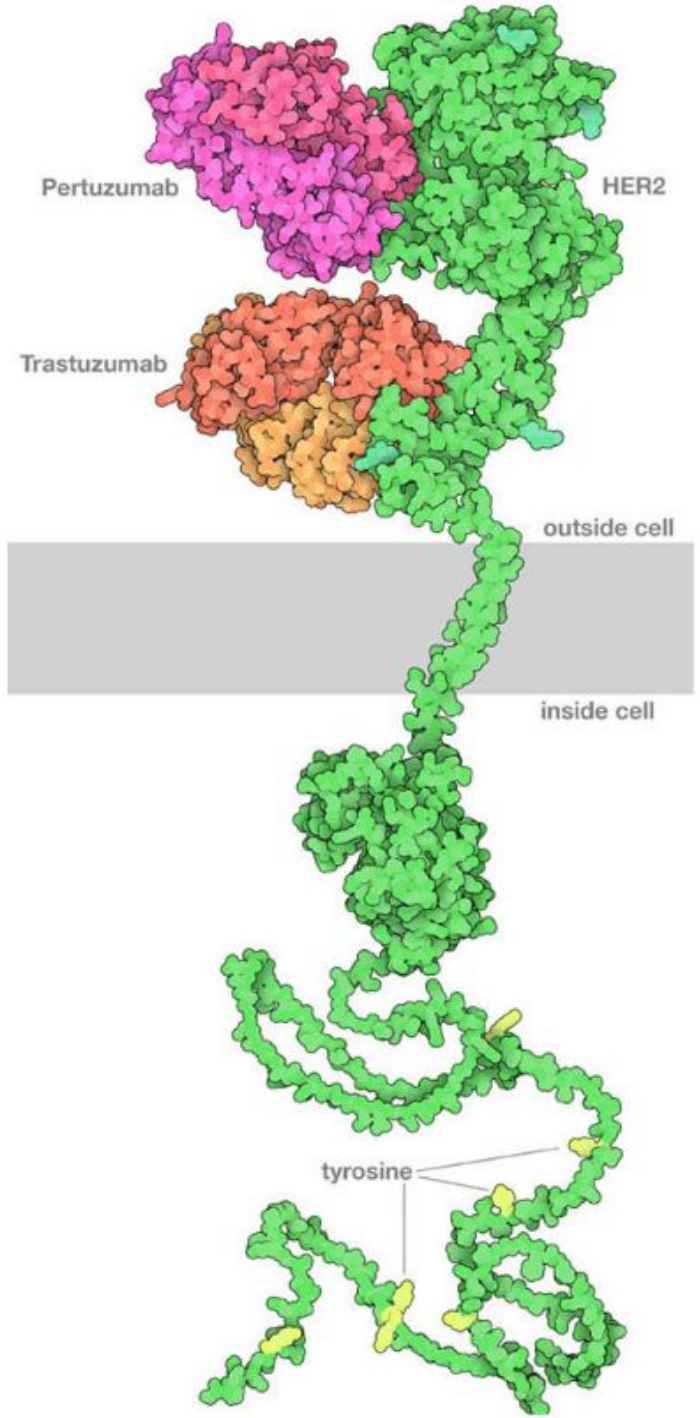
Tumor necrosis factor (TNFα) promotes the inflammatory response in autoimmune diseases, for instance inflammatory bowel disease and rheumatoid arthritis. Monoclonal antibodies against TNFα, for instance adalimumab (Humira), are used for such indications.PDB 3WD5

Many drugs are ligands of proteins and modulate protein's function

Left: The drug *sotorasib* binds covalently to the sulfur atom in cysteine 12 of the *Ras* protein, blocking its action. The drug is shown with carbon atoms in green, the cysteine sulfur is ir yellow, and GDP is in magenta. Image created in Jmol using PDB ID 6oim.

Right: The extracellular domain of HER2 bound to two therapeutic antibodies: *pertuzumab* and *trastuzumab*. The antibodies block the formation of active dimers of the receptor, thus blocking the growth signal (PDB 6ogi). The transmembrane domain is from PDB 2ksi. The kinase domain inside the cell is from PDB ID 3pp0, and the unstructured tail at bottom is from *AlphaFold2*.





Ill. Niklas Elmehed © Nobel Prize Outreach

John J. Hopfield

Ill. Niklas Elmehed © Nobel Prize Outreach

Geoffrey E. Hinton

The Nobel Prize in Physics 2024: "for foundational discoveries and inventions that enable machine learning with artificial neural networks"

Ill. Niklas Elmehed © Nobel Prize Outreach

David Baker

Ill. Niklas Elmehed © Nobel Prize Outreach

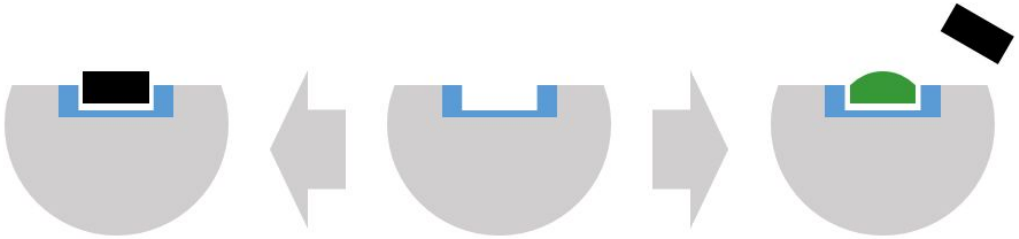
Demis Hassabis

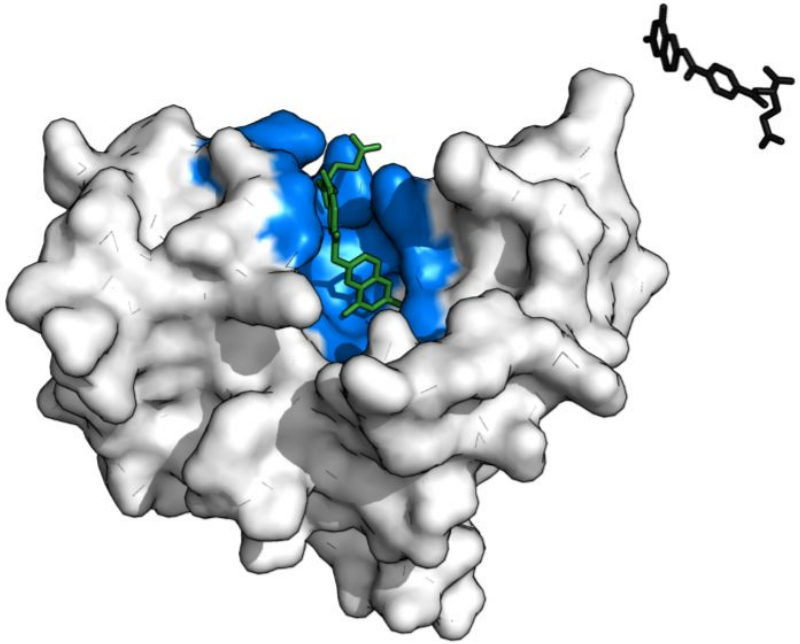
Ill. Niklas Elmehed © Nobel Prize Outreach

John M. Jumper

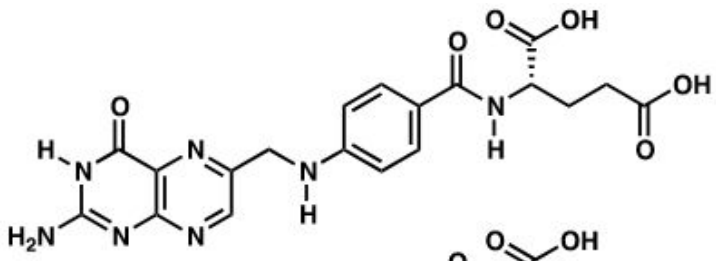
The Nobel Prize in Chemistry 2024 was divided, one half awarded to David Baker "for computational protein design", the other half jointly to Demis Hassabis and John M. Jumper "for protein structure prediction"

Drugs can compete with natural ligands

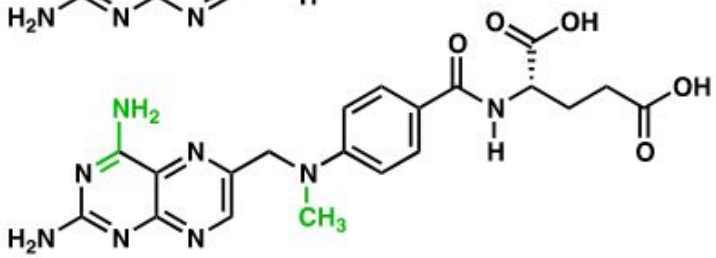




Work by Thomas Shafee, Shared under CC-AS-4.0, and work by Boghog. Based on PDB record 4QI9.



Dihydrofolic acid



MTX

The protein: Dihydrofolate reductase (DHFR) converts dihydrofolic acid into tetrahydrofolate. The process is important for cell proliferation and cell growth. DHFR is a drug target for oncology (cancer) and autoimmune diseases.

The natural substrate: Dihydrofolic acid (vitamin B9), in black. Dihydrofolic acid is the *natural ligand* of DHFR.

The drug: Methotrexate (MTX), in green, is a *synthesized ligand* of DHFR, and it is a *competitive inhibitor* of DHFR.

The binding site: where the enzyme binds its substrate and catalyses the chemical reaction, in blue.

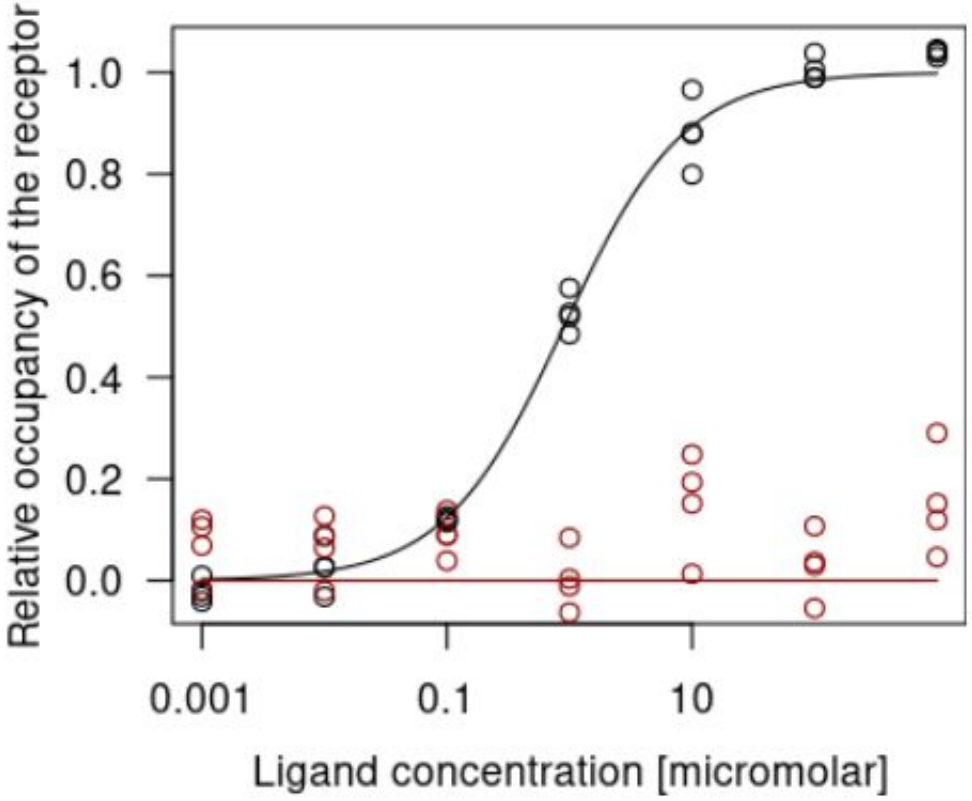
Concentration-occupancy curves characterize ligand-protein binding

X-axis: ligand concentration. Common units: molar (M), micromolar (mM, 10^{-6} M), nanomolar (nM, 10^{-9} M), picomolar (pM, 10^{-12} M).

Y-axis: relative occupancy of the receptor. Alternative values are possible, for instance response (more about that later).

Points: individual measurements. In this plot: mean value of replicates with error bars indicating variability.

Lines: fitted sigmoidal curves using the Hill function or its variants.

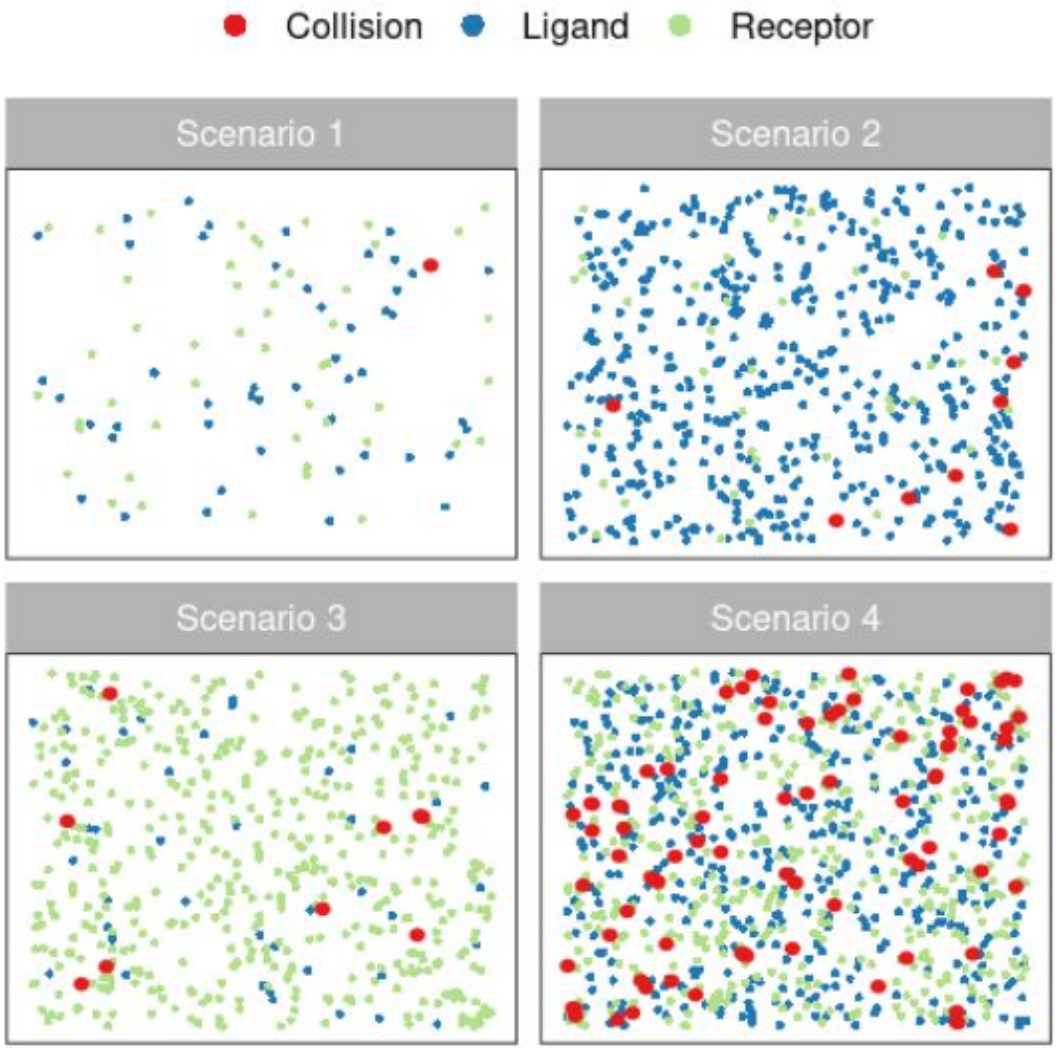


A simple mathematical model addresses a key question: how is a receptor occupied by varying concentrations of drugs?

$$\mathrm{L} + \mathrm{R} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathrm{LR}$$

$$\frac{d[LR]}{dt} = k_1[L][R] - k_2[LR]$$

- Ligand binding to receptor is a reversible reaction.
- **The law of mass action**: the rate of the chemical reaction is directly proportional to the product of the activities or concentrations of the reactants. The proposition can be derived from the *collision theory*. See the right graph for an illustration.



An ordinary differential equation (ODE) model quantifies receptor occupancy by varying concentrations of ligands

$$\mathrm{L} + \mathrm{R} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathrm{LR}$$

The law of mass action

$$\frac{d[LR]}{dt} = k_1[L][R] - k_2[LR]$$

At equilibrium, no net change of [LR]

$$k_1[L][R] = k_2[LR]$$

$$R_{total} = [R] + [LR]$$

$$k_1[L]([R_{total}] - [LR]) = k_2[LR],$$

$$[LR] = \frac{k_1[L][R_{total}]}{k_1[L] + k_2}$$

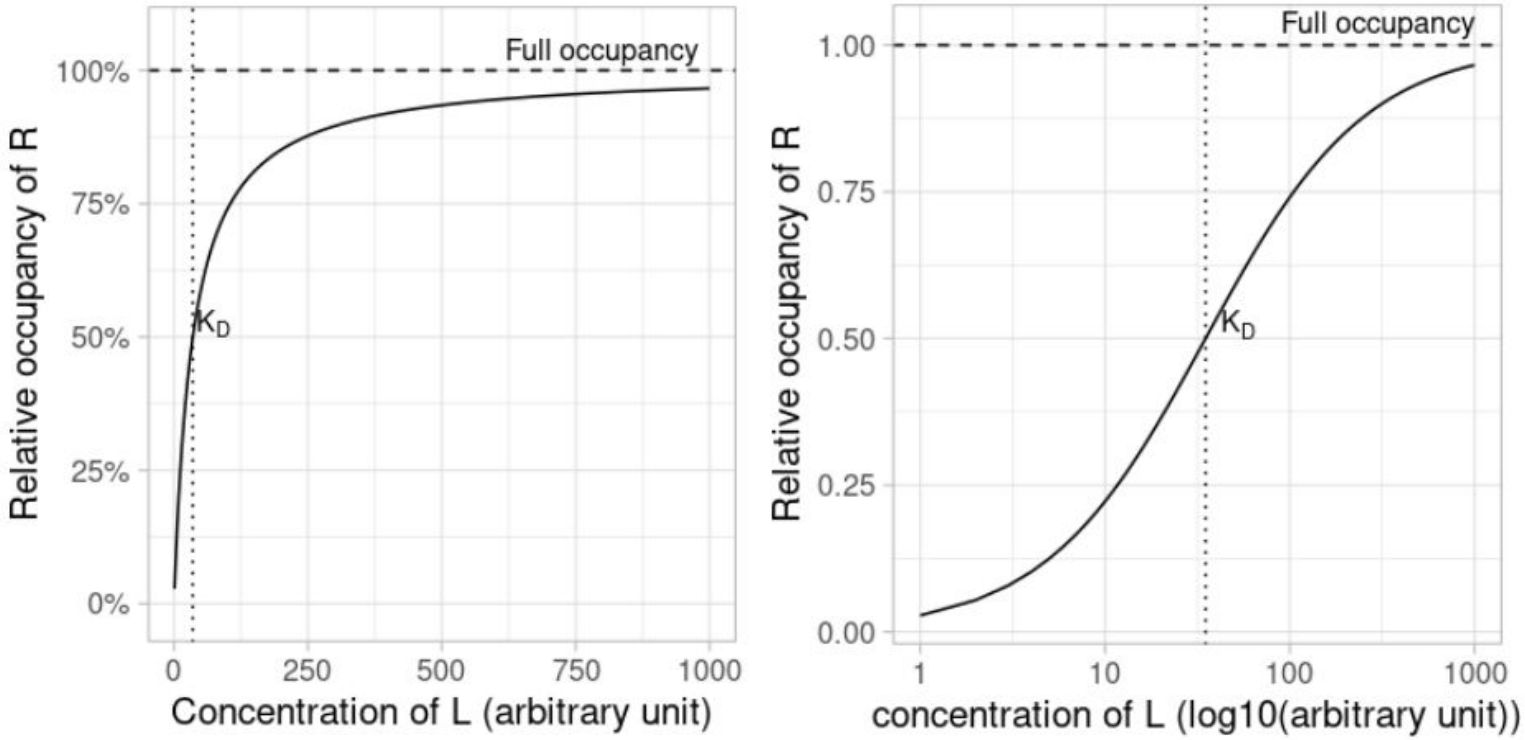
$K_D \equiv k_2/k_1$

$$[LR] = [R_{total}]\frac{[L]}{[L] + K_D}$$

The ODE model induces the simplest form of *Hill-Langmuir Equation*

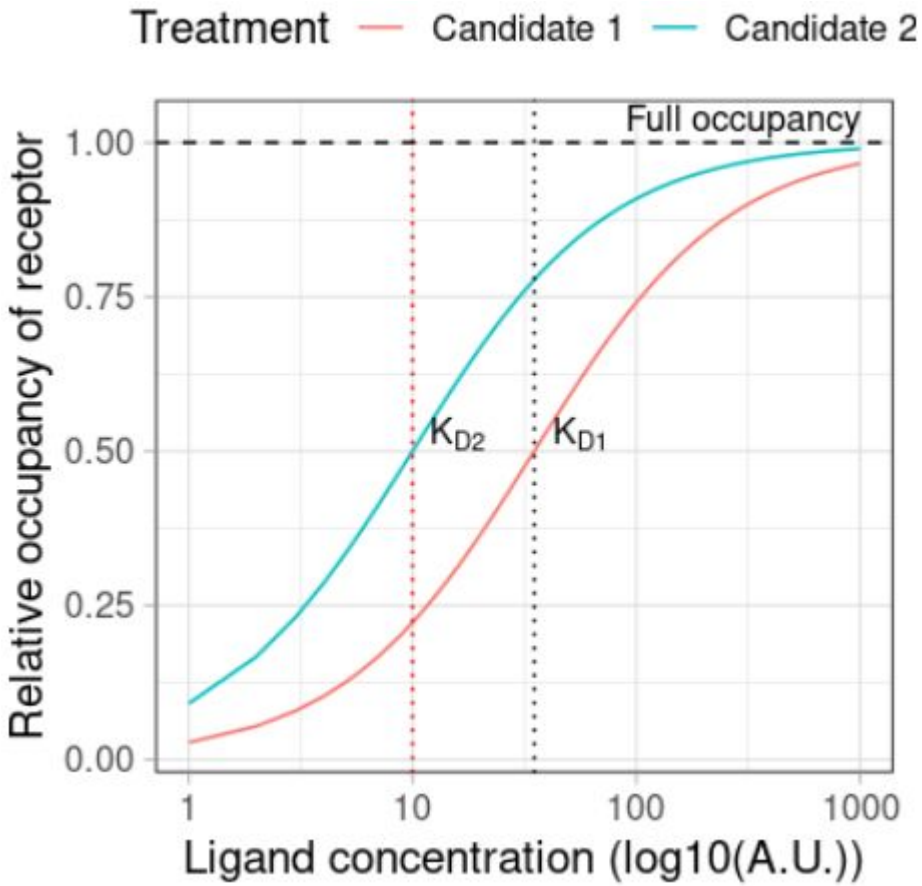
$$\mathrm{L} + \mathrm{R} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathrm{LR} \qquad K_D \equiv k_2/k_1$$

- The Hill-Langmuir function describes the occupancy of receptors by natural ligands of drugs.
- We can interpret K_D in two ways: the ratio of reaction speeds, and the concentration required to occupy half of the receptors.
- We will further discuss the Hill function again in the future.



$$[LR] = [R_{total}]\frac{[L]}{[L] + K_D}$$

Question: all other conditions the same, which drug candidate is more favorable? Why?



The Lotka-Volterra model of predator-prey relationships

- The Lotka-Volterra equations modelling predator-prey relationships.

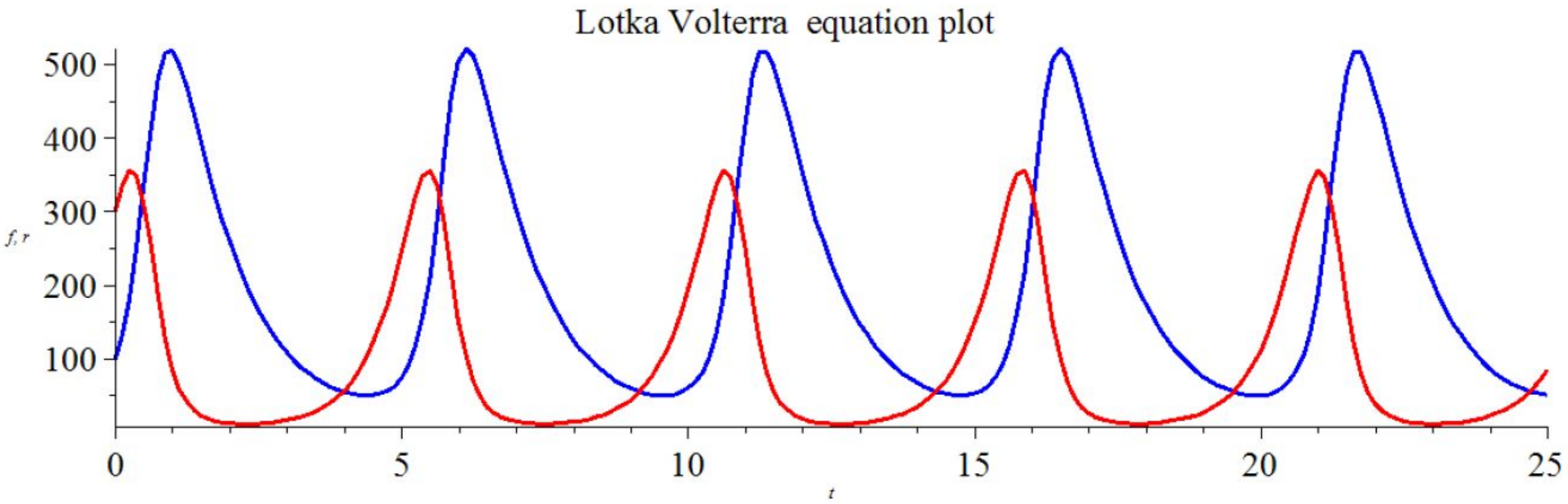
$$\frac{dx}{dt} = \alpha x - \beta xy, \tag{1}$$

$$\frac{dy}{dt} = -\gamma y + \delta xy, \tag{2}$$

where

- x is the number of prey (*e.g.* rabbits),
- y is the number of predator (*e.g.* foxes),
- $\frac{dx}{dt}$ and $\frac{dy}{dt}$ represent growth rates of the two populations,
- t represents time,
- α, β, γ, and δ are real parameters specifying the interaction of the two species.

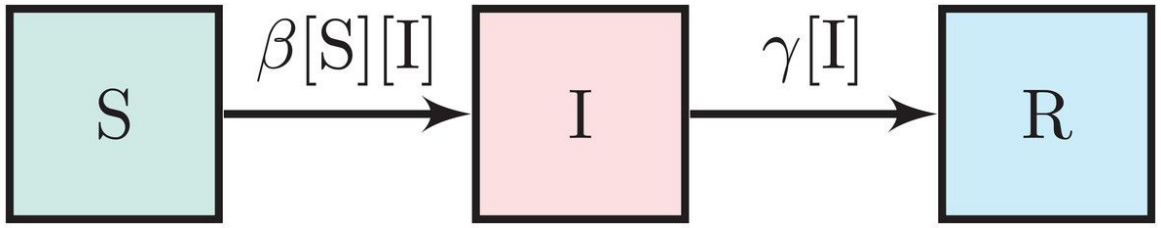
The Lotka-Volterra equations, visualized



Rabbits

Foxes

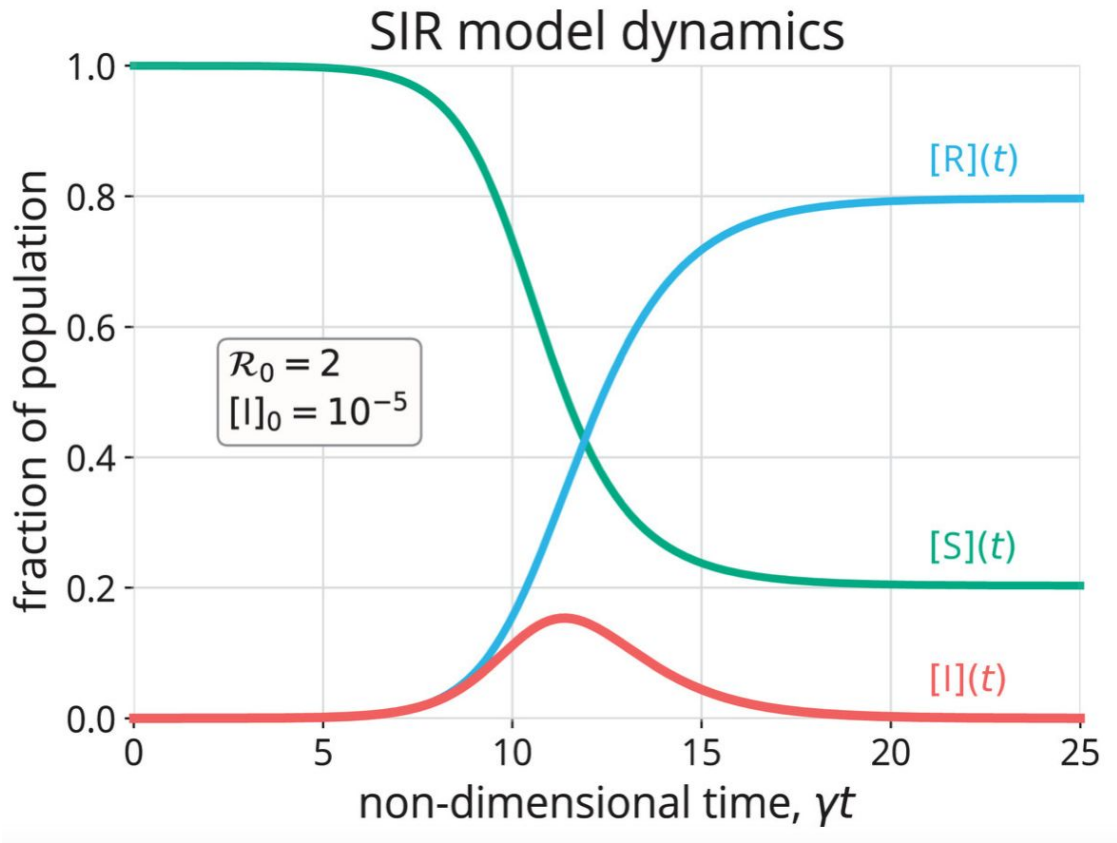
The SIR model of epidemiology models population behavior of viral infection and recovery



$$\frac{dS}{dt} = -\frac{\beta IS}{N},$$
$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I,$$
$$\frac{dR}{dt} = \gamma I$$

The SIR model of epidemiology

- S: Susceptible
- I: Infectious
- R: Removed

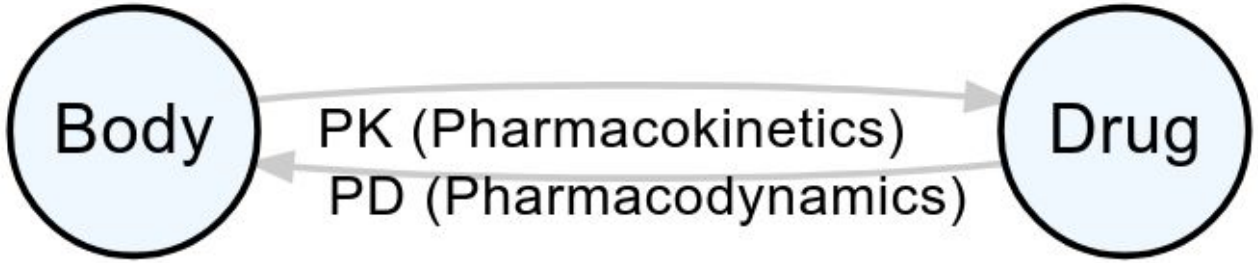


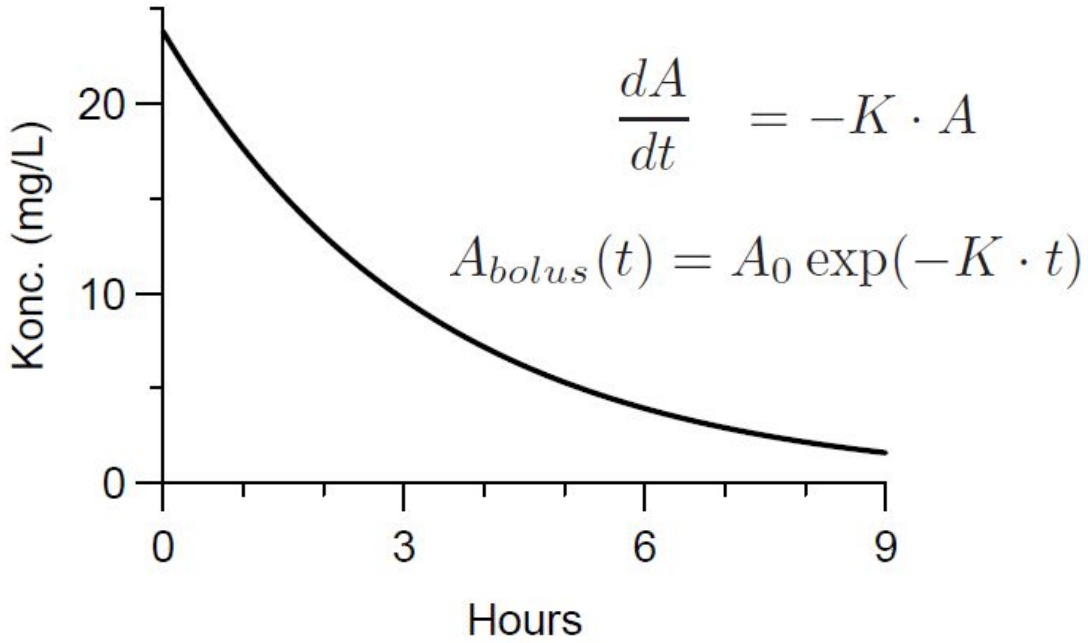
R_0, the basic reproduction number, is the number of people infected by the initial infectious individual. It is defined as β/γ.

Simon CM. 2020. The SIR dynamic model of infectious disease transmission and its analogy with chemical kinetics. PeerJ Physical Chemistry 2:e14https://doi.org/10.7717/peerj-pchem.14

ODE-based mechanistic models are often used in pharmacokinetic modelling

- Pharmacokinetics (PK) describes how the drug is absorbed, distributed, metabolised, and excreted by the body.
- Pharmacodynamics (PD) describes the effect of the drug to the body, mediated by drug-target interactions. PD is affected by PK, as well as other properties such as behaviour and genetics.
- A basic mathematical model of PK is a compartment model, i.e. one or more ordinary differential equations that describe the relationship between drug concentration and time. The simplest model is the decay model of bolus (injection).





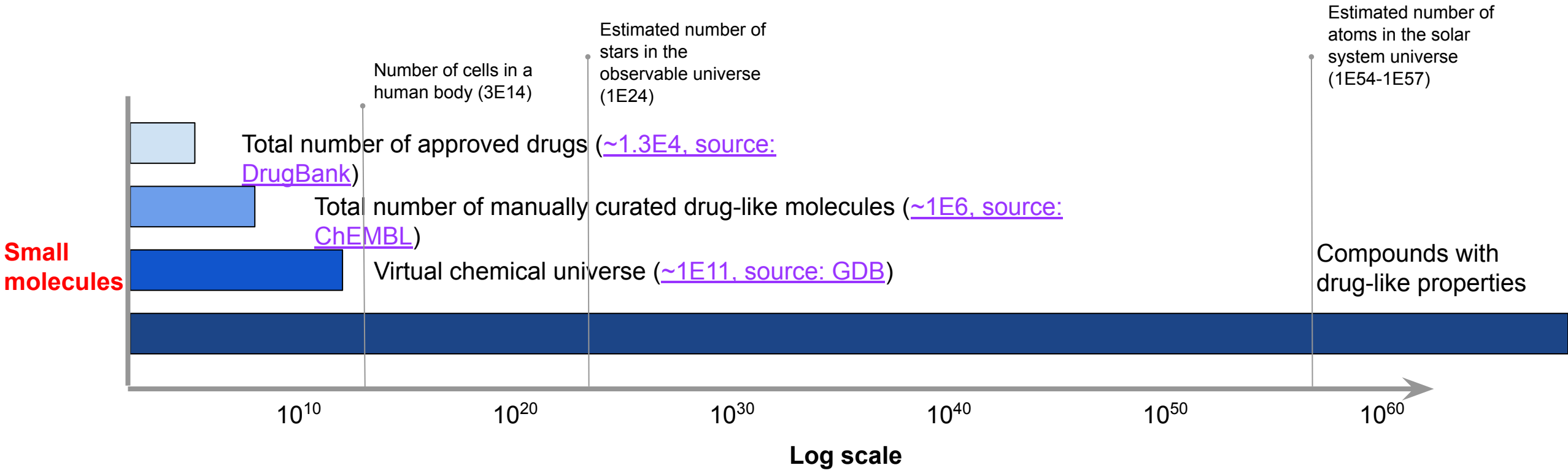
$$\frac{dA}{dt} = -K \cdot A$$

$$A_{bolus}(t) = A_0 \exp(-K \cdot t)$$

Points to consider besides occupancy

1. **Response**: Does binding of the drug inhibits the (pathological) function of the protein?
2. **Disease relevance**: does inhibition of the drug slows, stops, or inverses disease progression?
3. **Safety**: is the drug safe for patients?
4. **Specificity:** Does the drug bind to other proteins, or other molecules (RNA, DNA, …)?
5. **Exposure**: can the drug reach the protein target if it is injected or orally taken?
6. **Biomarker**: what can we measure to determine whether the drug has binds to the protein?
7. **Patients**: which patients are likely to respond to the treatment?
8. **Commercial:** does it pay off to discover, develop, and commercialize the drug?
9. …

Why drug *discovery*? 1. The chemical space is huge

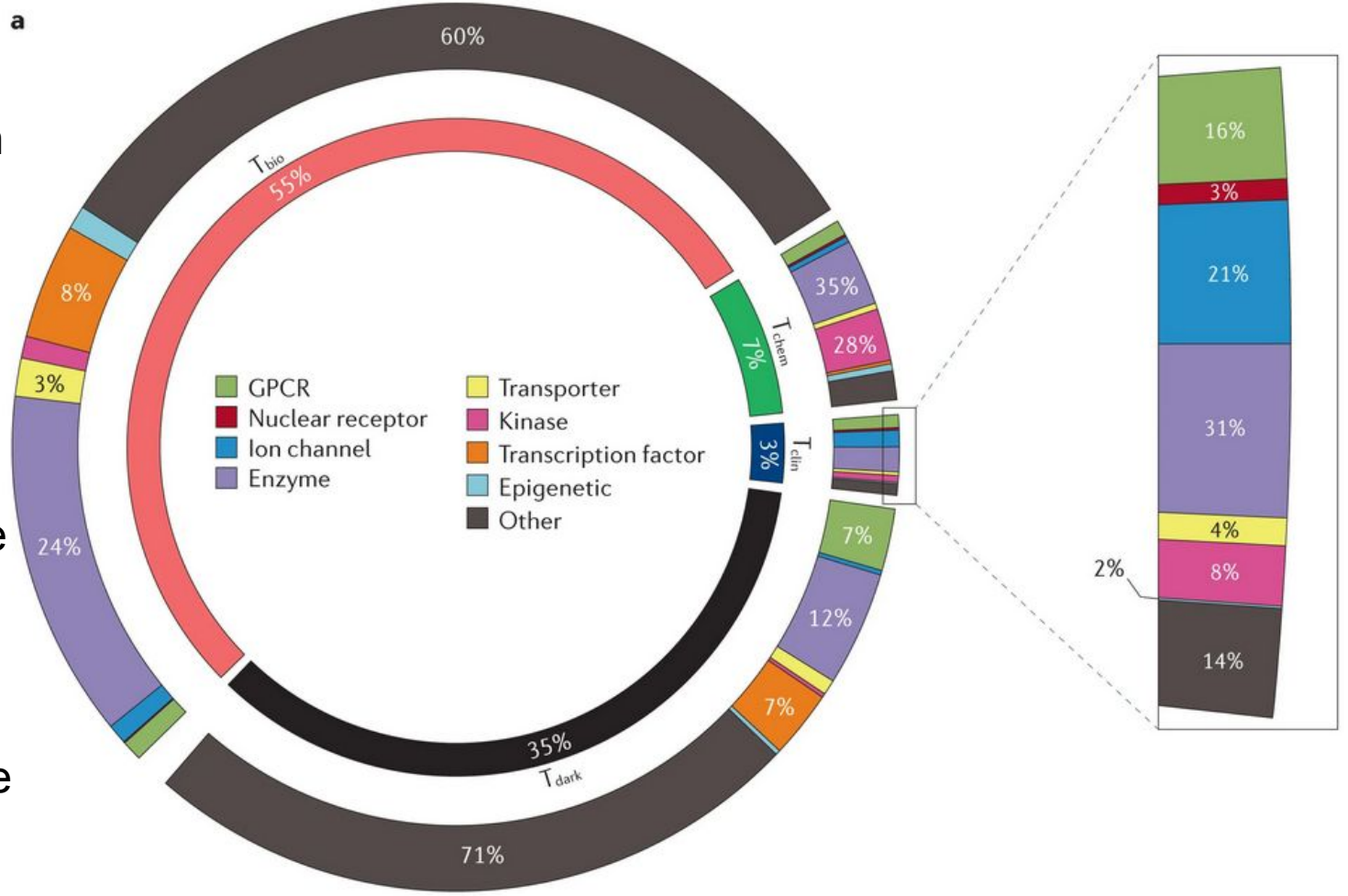


Why drug *discovery*? 2. The druggable proteome is huge - even excluding mutations, transcriptome, genome, …

There are about 20,000 proteins encoded by the human genome. We can classify them by (1) our knowledge of them, and (2) whether we have reliable chemical tools, biological tools, or even drugs to manipulate them.

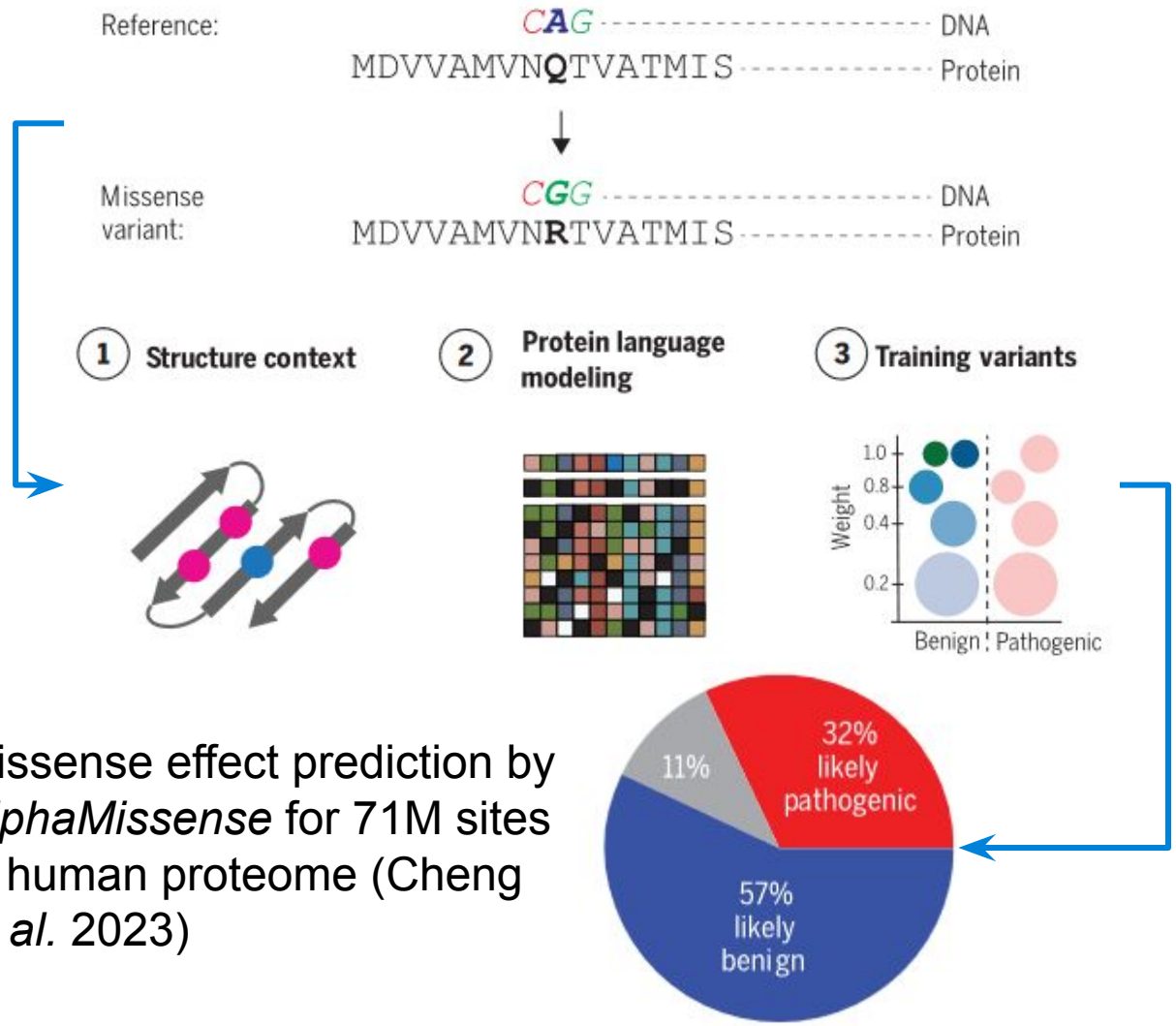
Inner ring: percentages of the whole proteome, classified by whether we have drugs (T_{clin}), whether we have chemical tool compounds (T_{chem}), whether we have biological compounds (T_{Bio}), or we are in the dark (T_{Dark}). Currently, we have only drugs for a few hundred proteins.

Outer ring: protein families.

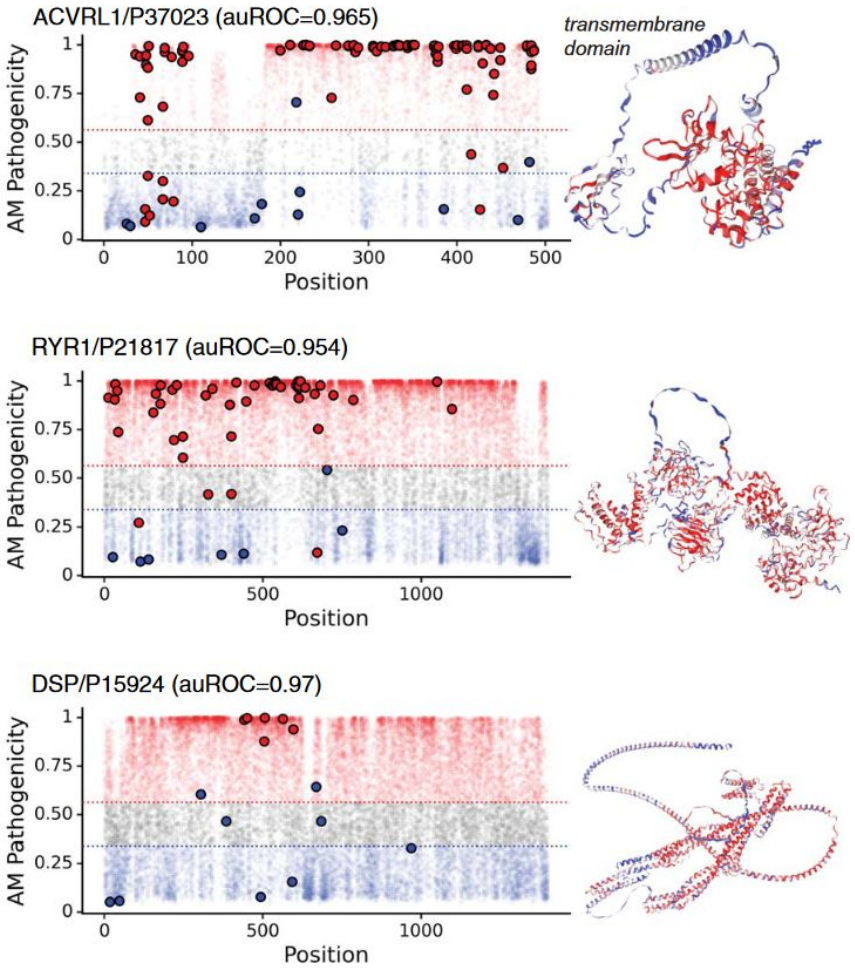


Oprea, et al. "Unexplored Therapeutic Opportunities in the Human Genome." Nature Reviews Drug Discovery 17 (February 23, 2018): 317–32.

Why drug *discovery*? 2. The druggable proteome is huge - now consider the mutations with predicted pathogenicity

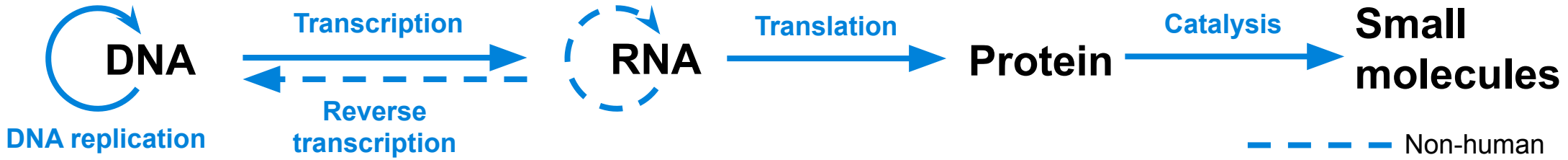


Missense effect prediction by *AlphaMissense* for 71M sites in human proteome (Cheng *et al.* 2023)



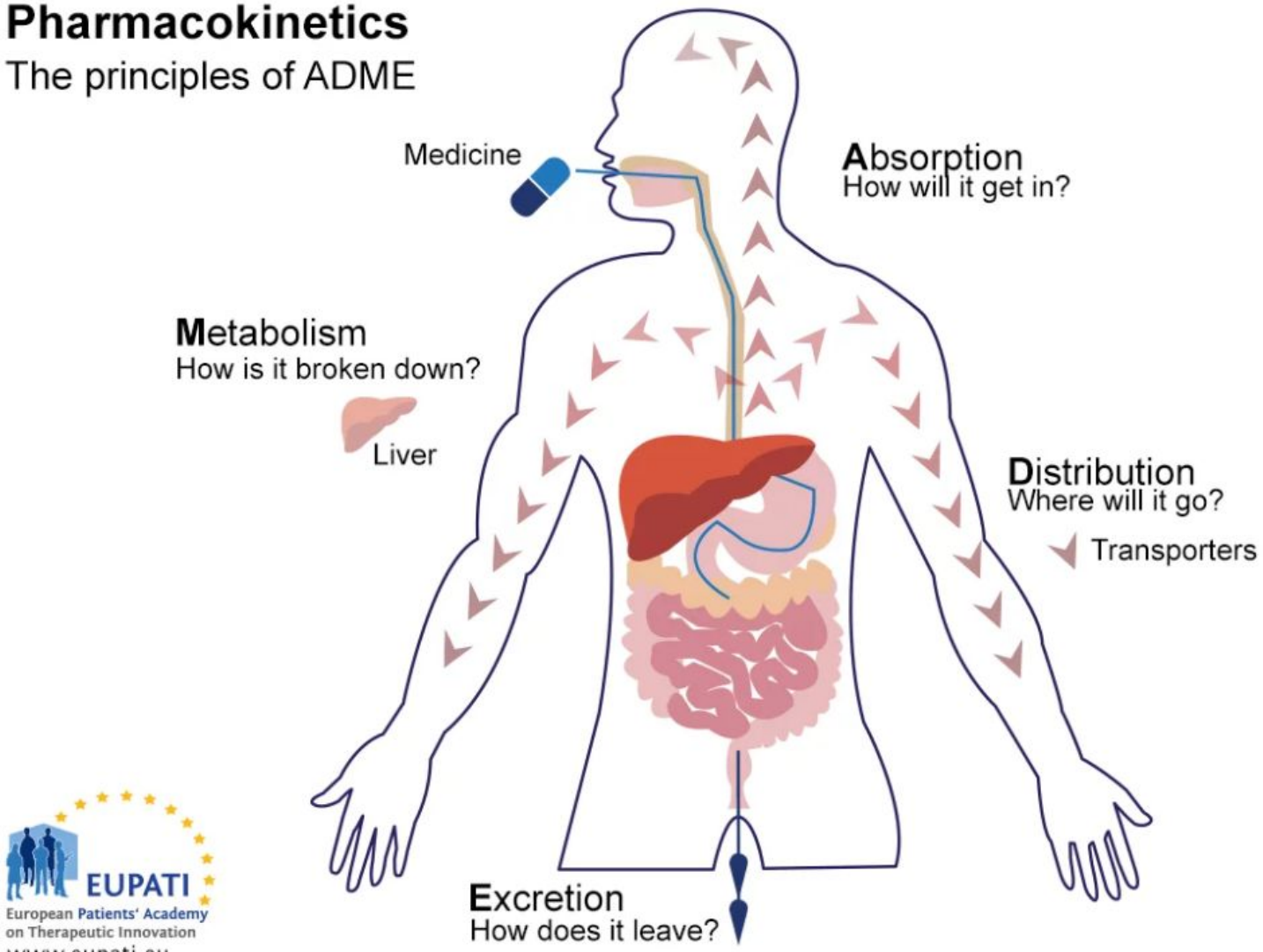
Example proteins chosen from ACMG clinically actionable genes

Why drug *discovery*? 3. The central dogma as an information channel: nodes and edges can all be targeted by drugs



Target	Example drugs
Small molecules	Dietary supplements
Catalysis	Enzyme inhibitors
Protein	Receptor agonists/antagonists, ion channel blockers, antibodies
Translation	Antimicrobial protein synthesis inhibitors
RNA	Antisense oligonucleotides (ASO), vaccines
Transcription	Antimicrobials (e.g. actinomycin D and α-Amanitin), splicing modifiers (e.g. Risdiplam/Evrysdi)
Reverse transcription	Antivirals (e.g. reverse transcriptase inhibitors AZT/Zidovudine)
DNA	Gene therapies (e.g. chimeric activated receptors in T-cells, CAR-T)
DNA replication	Topoisomerase inhibitors (e.g. quinolones) and chemotherapy agents

Why drug *discovery*? 4. The drug have to be absorbed and distributed in order to have systemic and organ-specific effects

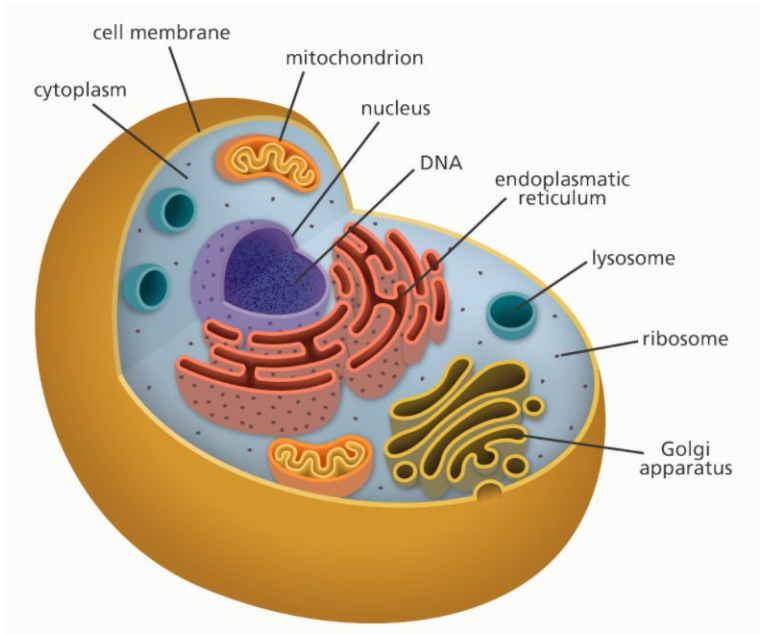


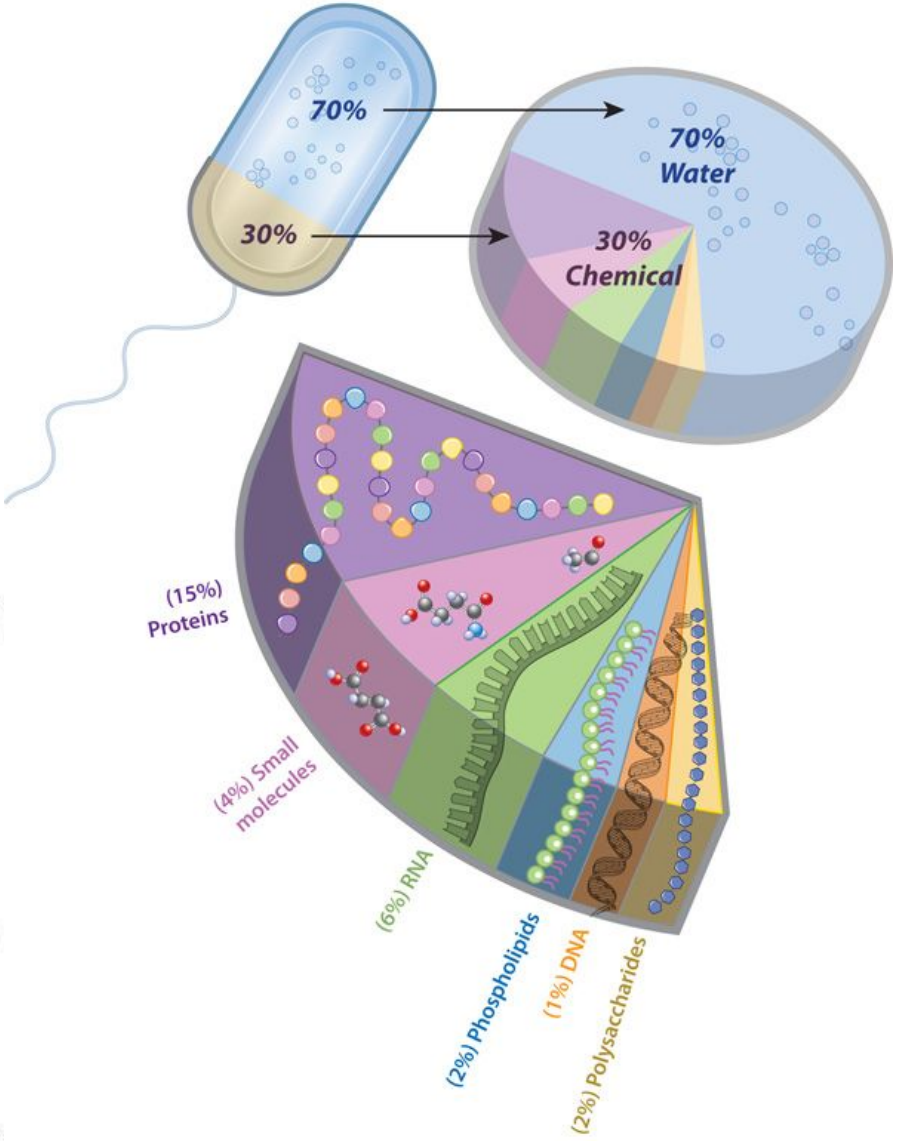
Why drug *discovery*? 5. Drugs have to reach the targets - despite physical barriers

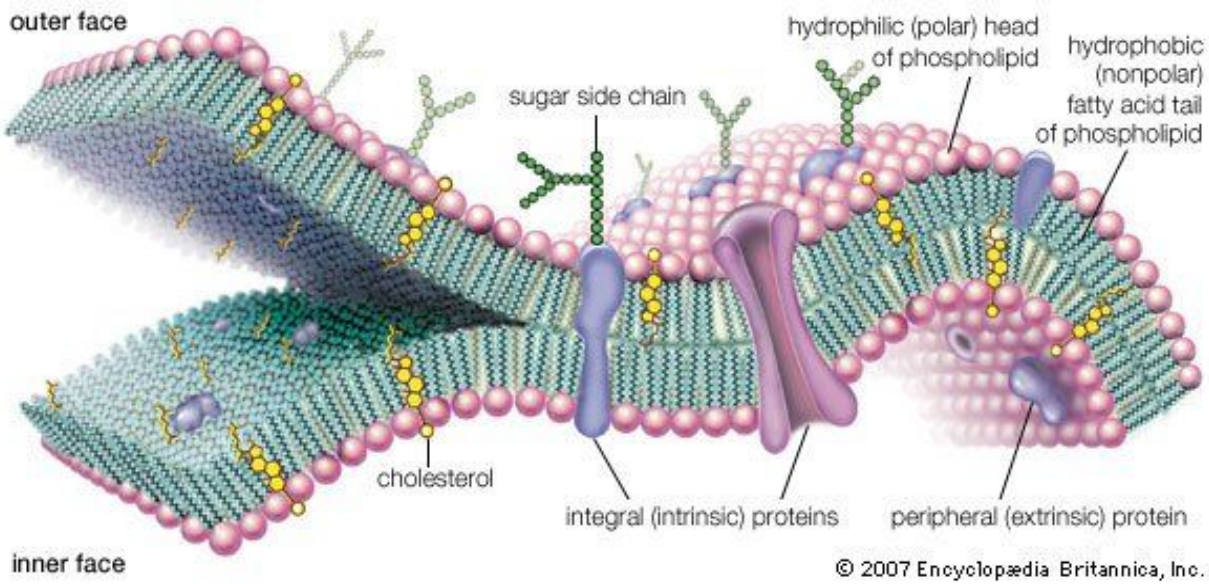
Bottom: Cell membrane, copyright of Encylopedia Britannica, Inc.

Top: Figure from The Human Protein Atlas

Right: Chemical composition of a human cell, by Scitable Nature Education.



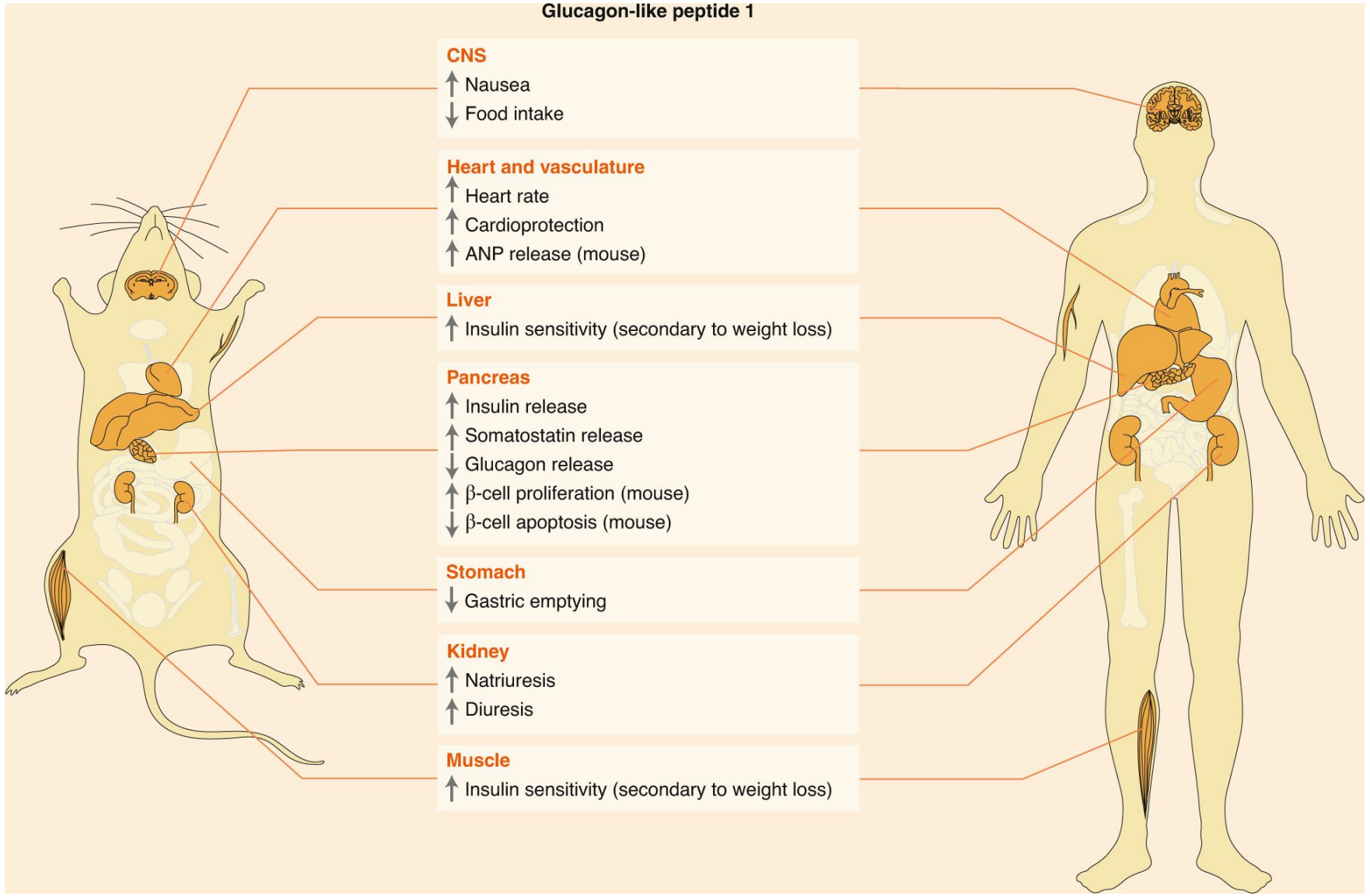




Why drug *discovery*? 6. The drug can have organ-specific and systemic effects, causing either benefits or risks

Direct effects of Glucagon-like peptide (GLP-1) and GLP1 receptor agonists (GLP1-RA) like semaglutide.

Gribble, Fiona M., and Frank Reimann. "Metabolic Messengers: Glucagon-like Peptide 1." Nature Metabolism 3, no. 2 (February 2021): 142–48.



Why drug *discovery*? 7. Do all patients benefit from the drug, or only some of them? Learn from the story of Herceptin

Link to the video

Questions for the video

1. What is the **indication** of *Herceptin*? What is its generic (USAN, or United States Adopted Name) name?
2. What is the **gene target** of Herceptin?
3. Which class best describes the target: Enzyme, Ion channel, Receptor and Kinase, or Structural protein?
4. In which year was the **target** of Herceptin described? When was Herceptin **approved**?
5. What was the **improvement** of Herceptin compared with earlier antibodies?
6. Why does a **biomarker** matter besides developing drugs?
7. In the clinical trial of *Herceptin* for **metastatic breast cancer**, how much improvement in the **median survival** did Herceptin achieve? And how much improvement is in the **adjuvant setting** (Herceptin applied directly after operation)?

Questions for further thinking

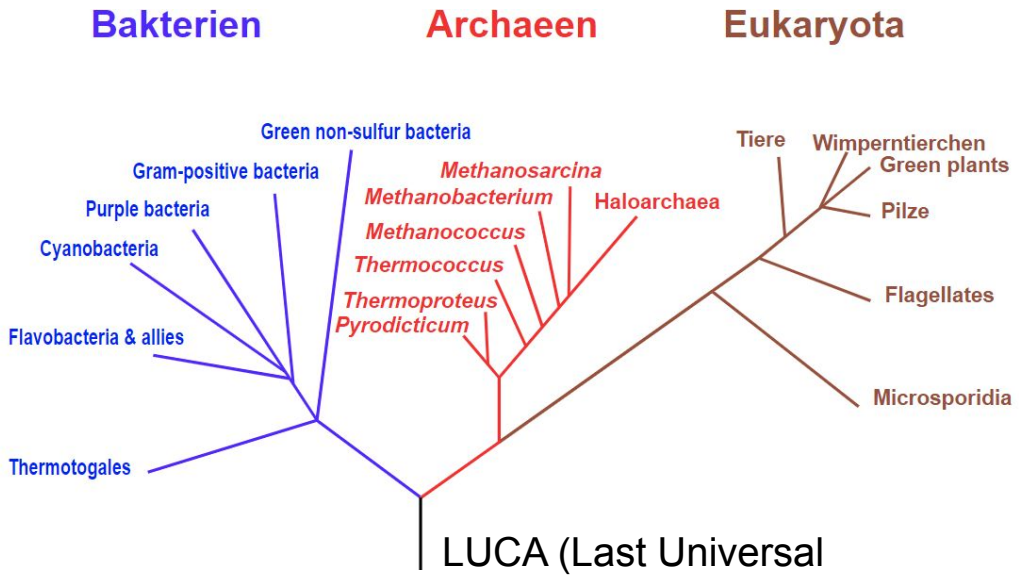
- Susan Desmond-Hellmann summarizes successful drug development in four aspects: (1) having a deep understanding of the basic science and the characteristics of the drug, (2) targeting the right patients, (3) setting a high bar in the clinic, and (4) working effectively with key regulatory decision makers. Where do you think mathematics and computer science play a crucial role?
- She emphasized the importance of collaboration. What skill sets do we need for that?
- How do you like her presentation? Anything that you can learn from her about presentation and storytelling?

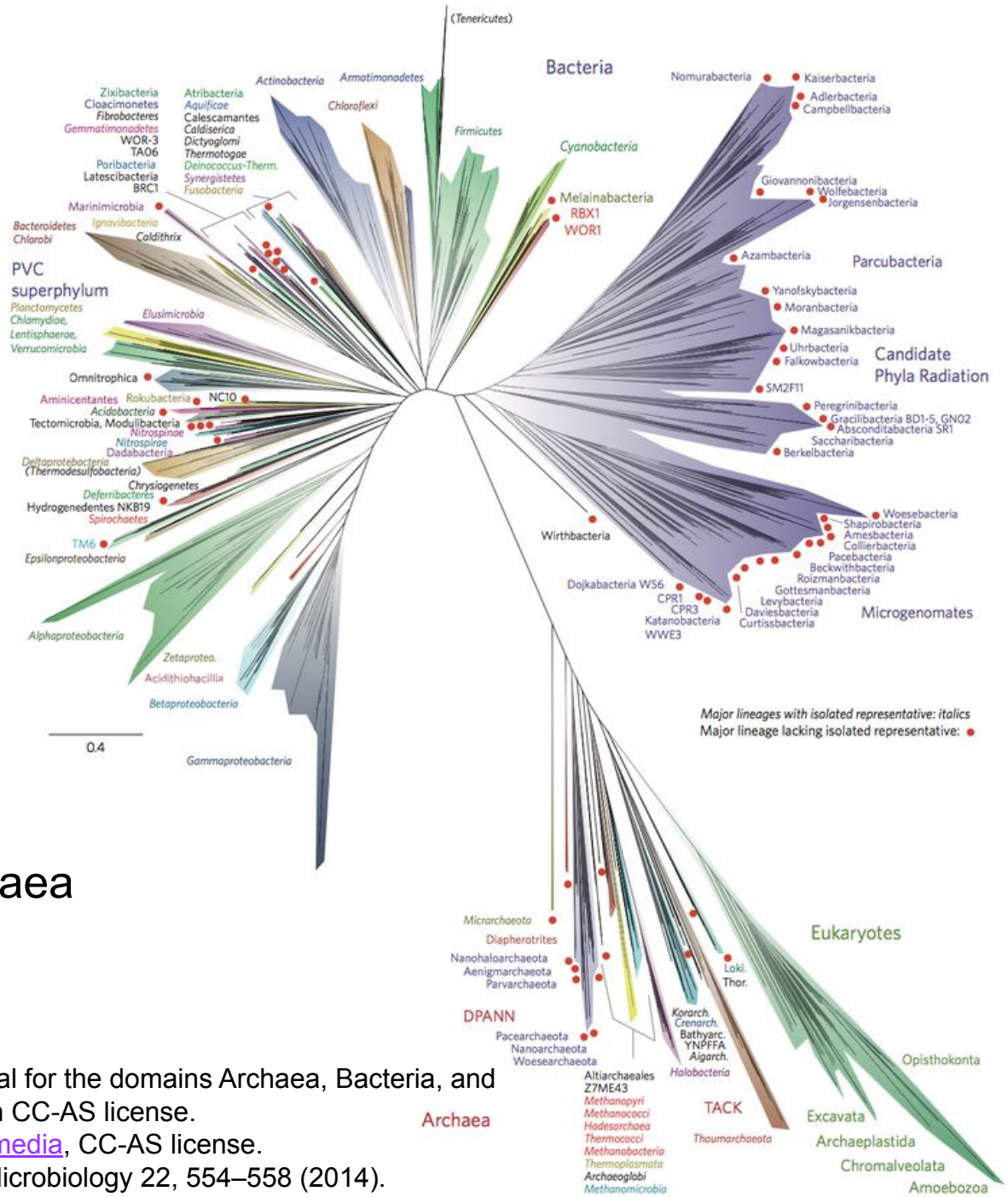
Conclusions and outlook

- We reviewed the central dogma from the drug discovery's perspective.
- We learned examples of ODE-based mechanistic models.
- We considered key aspects to consider for a drug to work.
- Next time, we shall continue learning statistic and causal models.

Backup slides

Virus is evolutionarily special





- The three-domain model of *cellular* life: (eu-)bacteria, archaebacteria, and eukaryotes.
- The two-domain model: bacteria as one branch, archaea and eukaryotes as the other.
- Virus fits in no domain of neither models.

1. Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A 87, 4576–4579 (1990). Figure from Wikimedia, reused with CC-AS license.
1. Hug, L. A. et al. A new view of the tree of life. Nat Microbiol 1, 1–6 (2016). Figure from Wikimedia, CC-AS license.
1. Forterre, P., Krupovic, M. & Prangishvili, D. Cellular domains and viral lineages. Trends in Microbiology 22, 554–558 (2014).