UC San Diego

# Lecture 7 - Protein Engineering Approaches (Rational and AI-driven)

BENG168

Instructor: Adam M. Feist, Assistant Professor, Shu Chien - Gene Lay Department of Bioengineering

# In-Class Announcements & Follow Up

- Update to previous lecture slides
- In-class announcements

# Medical Innovation Conference 2026

*MIC is designed to inspire undergraduates by showcasing a wide range of career paths in translational medicine, from graduate research to industry roles. With panels, keynote talks, and company representatives, the event celebrates the future of medicine while connecting students with experts who are shaping it.*


# Biomanufacturing Seminar – UCSD Student Group Led

📅 **Dates:** January 8 to March 13 every Wednesday via zoom

⏰ **Time: 9:30 to 10:30 am**

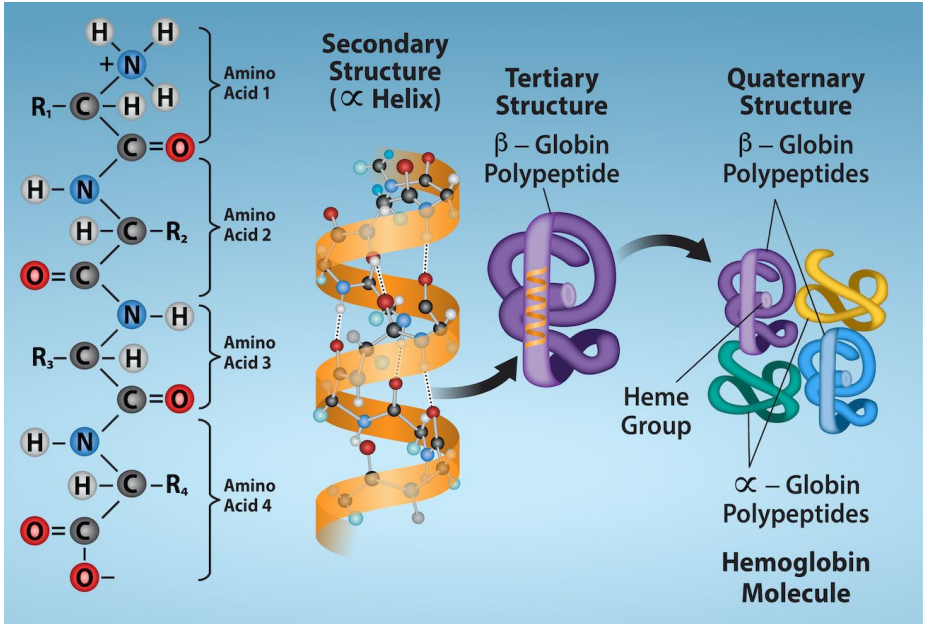
📍 Location: Zoom link as follows: https://urldefense.com/v3/__https://ucsd.zoom.us/j/97526231316__;!!Mih3wA!Fez537p1t7reo2QVvH4Bx8b-hqCXNjmpBNKOFcQcpGsErxZerZ26BUbhtaOO7QIH39RTO6HtAbIet1th$

🎯 **Who Should Attend:** Students in biology, chemical engineering, **bioengineering,** biochemistry, biotechnology, and related fields.

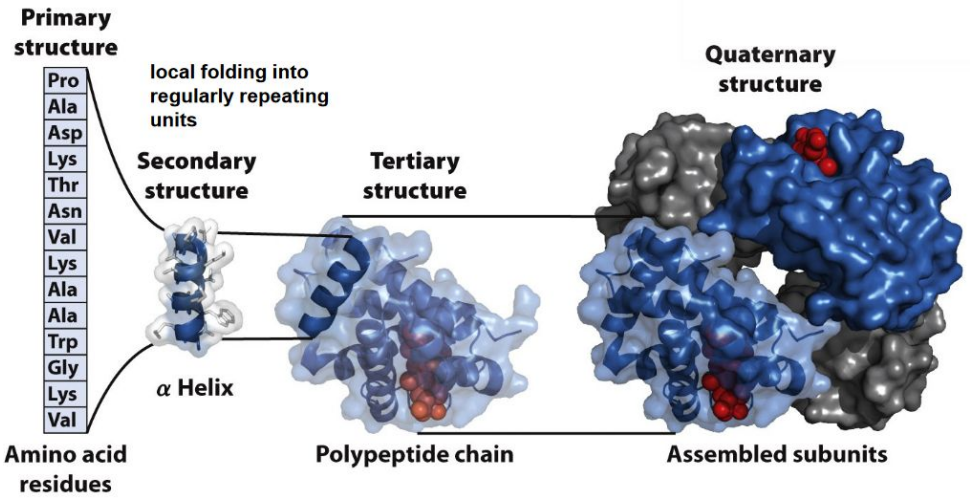
💡 **Why Attend:** Gain exposure to real-world biotechnology applications, expand your professional network, and explore career opportunities in bioprocessing and biomanufacturing.

**4 -1/28/2026** Biologics Drug Development
Gayle Derfus @ Oceanside Gilead Sciences Executive Director, Drug Substance
**5 -2/4/2026** Role of Manufacturing Science and Technology: MAbs
Scott Rosenthal @ Oceanside Genentech Executive Director MSAT
**6 -Date to be announced** Role of Manufacturing Science and Technology: Biosimilars
Lalit Saxena @ South Korea Samsung Biologics Senior Director MSAT
**7 -2/11/2026** Bioengineer Perspective on Delivering Monoclonal Antibodies
Eric Fallon @ San Diego Genentech/Vir Biotech/Neurocrine Process Development/MSAT /CMC
**8 -2/18/2026** Manufacturing viral vectors for gene and cell therapy
Daniel Gibbs @ San Diego Cirsium Biosciences CEO
**9 -2/25/2026** Bioengineer Organs
Emily Beck @ Minneapolis Miromatrix Director Upstream R&D
**10 -3/4/2026** Automated Experimentation and Evolutionary Engineering of Microbes for Industrial Biotechnology Adam Feist @ UCSD UCSD Bioengineering Assistant Professor
**11 -3/11/2026** Industrial Chemicals Bioprocessing: Scaling Technologies
Seth Levine @ San Diego Genomatica Engineering Fellow

# Protein Structure Refresher – 4 levels of protein structure



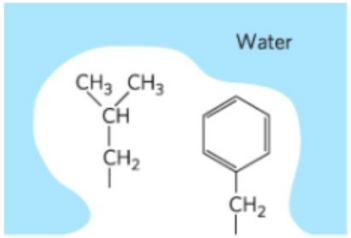

**Primary, secondary, tertiary, and quaternary structure of hemoglobin.** The primary structure of a hemoglobin is its amino acid sequence. It secondary structure is entirely α helices. Its tertiary structure is globular. Four protein chains come together to form the quaternary structure that is the functional hemoglobin protein. (credit: Rao, A., Ryan, K., & Tag, A. Department of Biology, Texas A&M University)



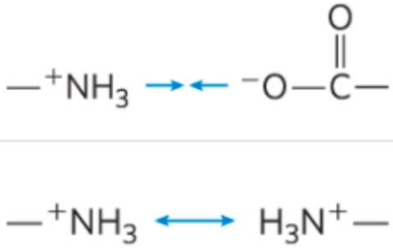

**Level of Structure in proteins.** The primary structure of a protein is its sequence of amino acids. The secondary structure of a protein refers to its basic patterns of hydrogen bonding. In other words, the secondary structure is concerned with the three-dimensional structures formed via the side chain's intermolecular interactions (e.g., α-helices and β-sheets). The tertiary structure of a protein is the overall three-dimensional shape of the protein. In proteins, this level of structure is also roughly described as a type of protein fold. The quaternary structure deals with the interactions between multiple polymer chains of a protein (e.g., the structure of a multi-chain assembly). From Nelson, D. and Cox, M., Lehninger Principles of Biochemistry, 8th ed.

# Protein Structure Refresher – Noncovalent interactions are responsible for protein folding

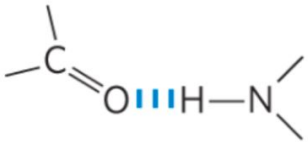
1) Hydrophobic effect:




2) Ionic attractive or repulsive forces:




3) Hydrogen bonds:




4) Van der Waals interactions/London dispersion forces:

**TABLE 2-3 van der Waals Radii and Covalent (Single-Bond) Radii of Some Elements**

| Element | van der Waals radius (nm) | Covalent radius for single bond (nm) |
|---|---|---|
| H | 0.11 | 0.030 |
| O | 0.15 | 0.066 |
| N | 0.15 | 0.070 |

From Nelson, D. and Cox, M., Lehninger Principles of Biochemistry, 8th ed.

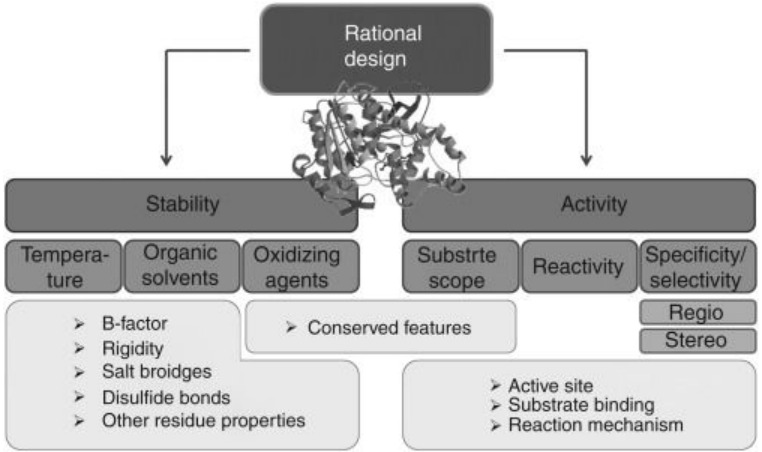
# Module 1: Site-Directed Protein Engineering (Rational Design) (Source Pages: Chapter 3: 156–160)

- Targeting specific amino acid sites.
- Using PCR for precise mutations.
- Incorporate unique nonstandard residues.

# Rational Design Strategy

- **Key Residues:** Identify exactly which amino acids are responsible for **activity** or **stability** using high-resolution structural data.
- **Knowledge-Based:** Requires high-resolution structural data (e.g., X-ray crystallography) to predict which residues affect function.
- **Targeted Alteration:** Swaps specific codons to improve properties like thermal tolerance, pH stability, or substrate specificity.
- **Iterative Process:** Engineered versions must be expressed and tested to verify if the mutation produced the desired effect.

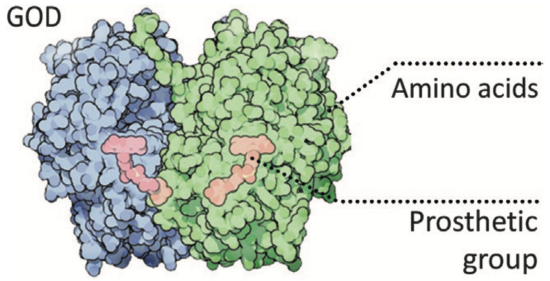
**Targets for rational design depending on the property desired to be altered.**




Isabel Oroz-Guinea, Katja Zorn, Henrike Brundiek,
doi.org/10.1016/B978-0-12-813167-1.00002-5

# Directed Mutagenesis Principles

- **Specific Targeting:** Changes individual amino acids or short sequences at a precisely defined site in the gene.
- **Trial-and-Error:** Even with structures, predicting exact outcomes is difficult, often requiring several rounds of testing.
- **Distance Effects:** Mutations may target residues far apart in primary sequence but juxtaposed in 3D space after folding.




***Enzymes** are biocatalytic proteins. Shown here is **glucose oxidase (GOD)**, which is a dimer molecule consisting of 2 × 256 amino-acid components. FAD (flavin adenine dinucleotide) acts as a prosthetic group in the active center (more detail in Chapter 3).*

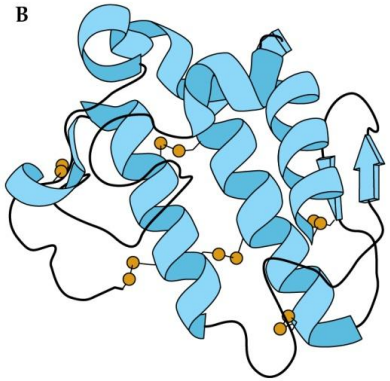
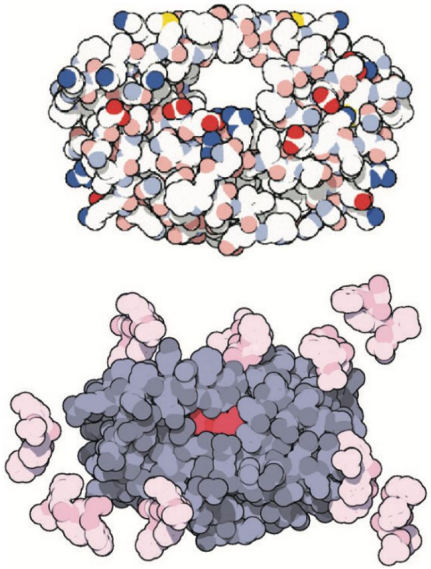



**Figure 3.9** Disulfide bond in a protein. **(A)** A covalent, disulfide bond forms by oxidation of sulfhydryl (SH) groups on cysteines. **(B)** Disulfide bonds between cysteines (represented by brown circles) within a polypeptide (ribbon diagram) contribute to the structural stability of the protein.

# Approaches to Determine Edits for Rational Design

- **Structural Foundations:** Utilize X-ray crystallography or NMR to determine the exact 3D coordinates of residues within functional sites.
- **Computational Prediction:** Apply molecular docking and homology modeling to simulate interactions and predict how specific changes affect biophysics
- **Hypothesis-Driven Strategy:** Propose targeted modifications based on mechanistic understanding, such as swapping active-site residues to alter substrate specificity



*Computer-aided drug design: Drugs can be designed and tested on computers. Automated docking methods are used to find the best docking sites on a biomolecule. If the predicted bond is strong enough, the molecule can be synthesized and its activity tested. The best site for saquinavir is shown in red.*

# Implementation: Overlap Extension PCR

- **Primer Design:** Utilize **4 primers**, where the **two internal primers** carry the specific mismatched nucleotides intended to rewrite.
- **Fragment Synthesis:** Amplify the "left" and "right" fragments of the gene separately in two initial PCR reactions.
- **Gene Reassembly:** Denature and re-anneal the fragments so they overlap at the mutation site, then extend them to form a full-length mutant.
- Recall **DNA polymerase** and **restriction enzymes** from - Lecture 1, Module 2 content.

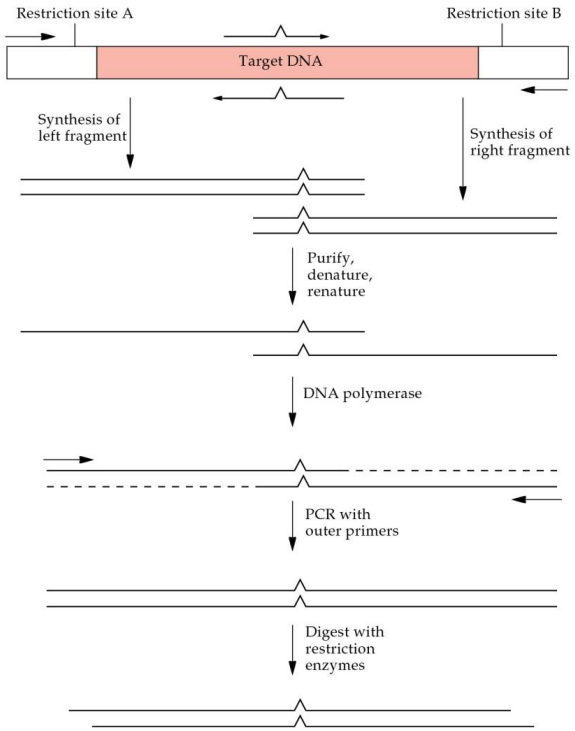



**Figure 3.51** Site-directed mutagenesis by overlap extension PCR. The left and right portions of the target DNA are amplified separately by PCR. The primer pairs are shown by horizontal arrows. Oligonucleotides that carry the mutations are depicted as a line with a spike; a spike denotes a position that contains a nucleotide that is not found in the native gene. The amplified fragments are purified, denatured to make them single stranded, and then reannealed. Regions of overlap are formed between complementary mutation-producing oligonucleotides. The single-stranded regions are made double stranded with DNA polymerase, and then the entire fragment is amplified by PCR. The resultant product is digested with restriction endonucleases to facilitate cloning into a vector that has been digested with the same enzymes.

# Alternative: Inverse PCR Mutagenesis

- **Circular Templates:** Uses a plasmid as the starting material rather than linear DNA.
- **Divergent Primers:** Primers fold 'backwards' around the circle, incorporating the mutation at the start of the new strand.
- **Ligation:** The **linear PCR product** is treated with T4 polynucleotide kinase and ligated to restore the circular plasmid.
- **High Efficiency Edits:** Introduce point mutations, deletions, or insertions with high precision, yielding a frequency of correct mutants that is so high it requires the screening of only a few clones.

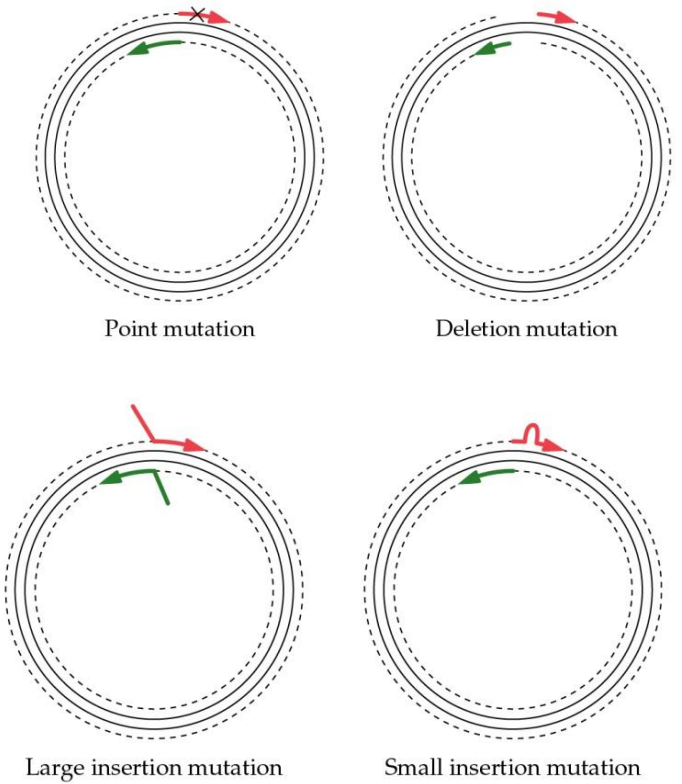



**Figure 3.52** Overview of the basic methodology to introduce point mutations, insertions, or deletions into DNA cloned into a plasmid. The forward and reverse primers are shown in red and green, respectively. The solid circles represent tem-plate DNA. The dotted lines represent newly synthesized DNA. The X indicates an altered nucleotide(s).

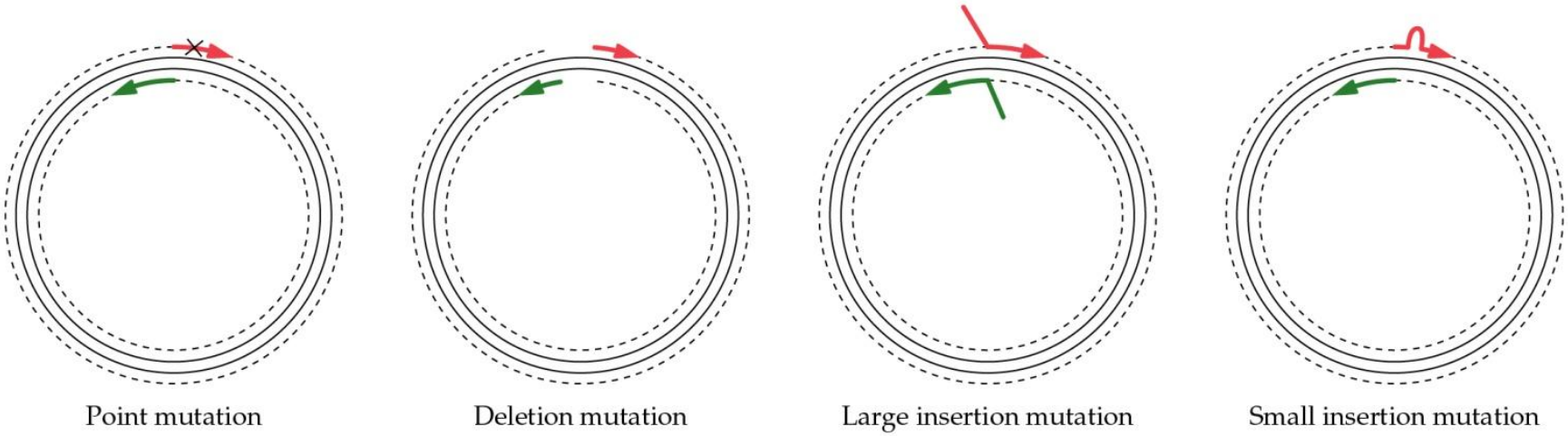



**Figure 3.52** Overview of the basic methodology to introduce point mutations, insertions, or deletions into DNA cloned into a plasmid. The forward and reverse primers are shown in red and green, respectively. The solid circles represent tem-plate DNA. The dotted lines represent newly synthesized DNA. The X indicates an altered nucleotide(s).

# Stabilizing Receptor-Associated Protein (RAP)

- **Natural Function:** RAP acts as a molecular chaperone in the Golgi to help process and transport LRP1 receptors to the cell surface.
- **Acid Sensitivity:** Wild-type RAP is unstable and denatures in the acidic environment of endosomes, limiting its use as a hemophilia treatment.
- **Clinical Goal:** Engineer a more stable version of RAP that remains bound to LRP1 at low pH to prevent protein clearance.

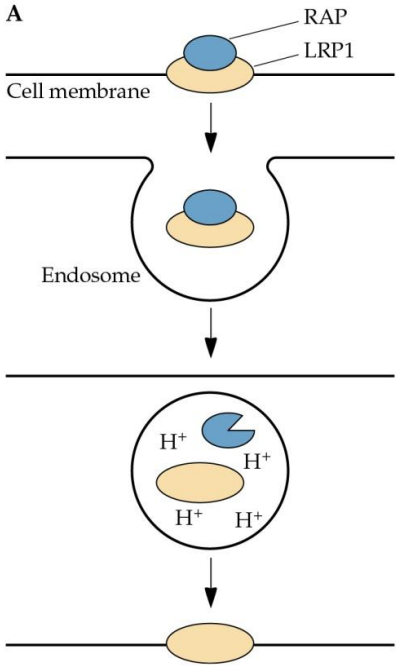



**Figure 3.60** Directed mutagenesis of receptor-associated protein (RAP) to increase acid stability. **(A)** Following binding of exogenous RAP to low-density lipoprotein receptor-related protein 1 (LRP1) in the cell membrane, the protein complex is taken into the cell by endocytosis. Acid-sensitive wild-type RAP is denatured in the acidic endosome and releases LRP1, which is recycled back to the cell membrane. **(B)** To

# Stabilizing Receptor-Associated Protein (RAP)

- **Covalent Bridges:** Introduce a new disulfide bond between residues Y260C and T297C to lock the protein's domain in place.
- **Residue Swapping:** Replace four histidine residues with phenylalanine to prevent protonation and unfolding in acidic conditions.
- **Predictive Modeling:** Use computer simulations to identify the exact residues that cause the protein to unfold at pH 5.5.

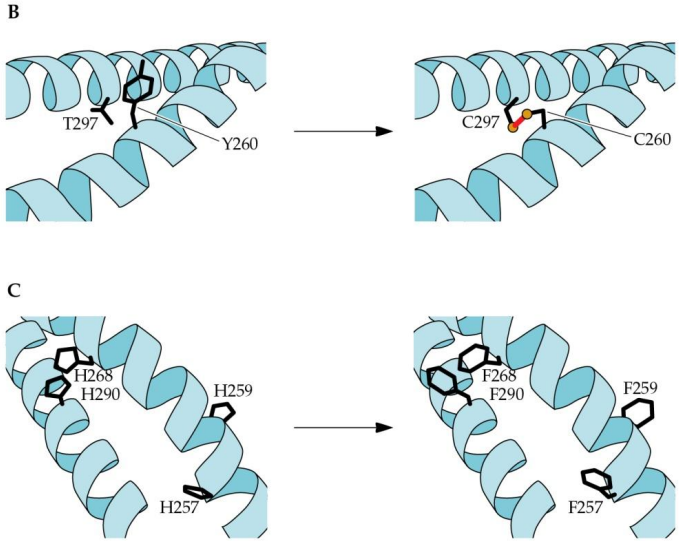



**Figure 3.60** Directed mutagenesis of receptor-associated protein (RAP) to increase acid stability. **(A)** Following binding of exogenous RAP to low-density lipoprotein receptor-related protein 1 (LRP1) in the cell membrane, the protein complex is taken into the cell by endocytosis. Acid-sensitive wild-type RAP is denatured in the acidic endosome and releases LRP1, which is recycled back to the cell membrane. **(B)** To increase the stability of RAP, a disulfide bond was introduced into domain D3. Tyrosine at position 260 (Y260) and threonine at position 297 (T297) were changed to cysteines (C260 and C297) by site-directed mutagenesis. **(C)** To further increase acid stability, four histidines (H257, H259, H268, H290) that are protonated at low pH were changed to phenylalanine (F257, F259, F268, F290). Data from Prasad et al., *J. Biol. Chem.* **290:**17262, 2015.

# Results: Stabilizing Receptor-Associated Protein (RAP)

- **Dissociation Constant (KD):** Measures how easily a protein "lets go" of its target; a lower KD indicates tighter binding.
- **Fold Change Calculation:** The ratio of KD in acid (pH 5.5) vs. neutral (pH 7.4). This measures how much "grip" is lost in the endosome.
- **Wild-Type (WT) Results:** KD jumps from 0.68 to 123 nM in acid—a 181-fold loss of strength.
- **Engineered Success:** Combined mutations (disulfide bonds + histidine removal) resulted in a fold change of only 22.
- **Outcome:** The engineered protein stays bound to its receptor in acidic environments, creating a much more potent therapeutic inhibitor.

**Table 3.17** Binding affinity of LRP1 for mutant RAP[a,b]

| RAP | Mutations | KD at pH 7.4 | KD at pH 5.5 | Fold change in KD |
|---|---|---|---|---|
| Wild-type | None | 0.68 | 123 | 181 |
| RAP disulfide | Y260C, T297C | 1.67 | 58.2 | 35 |
| RAP quad | H:F H257F, H259F, H268F, H290F | 1.38 | 72.7 | 53 |
| RAP combined | Y260C, T297C, H257F, H259F, H268F, H290F | 0.96 | 21.5 | 22 |

[a]Data from Prasad et al., *J. Biol. Chem.* **290:**17262–17268, 2015.
[b]KD, dissociation constant (nM) fold change = KD at pH 5.5/KD at pH 7.4.

# Beyond the Standard 20 Amino Acids

- **Synthetic Side Chains:** Incorporate over 200 non-natural amino acids to add **unique chemical functions** to proteins.
- **Orthogonal Machinery:** Requires a unique tRNA/aminoacyl-tRNA synthetase pair from a different organism (e.g., *Methanococcus*).
- **Chemical Novelty:** This technique enables the creation of proteins with enhanced catalytic efficiency, specialized binding affinities, or entirely new biological functions.

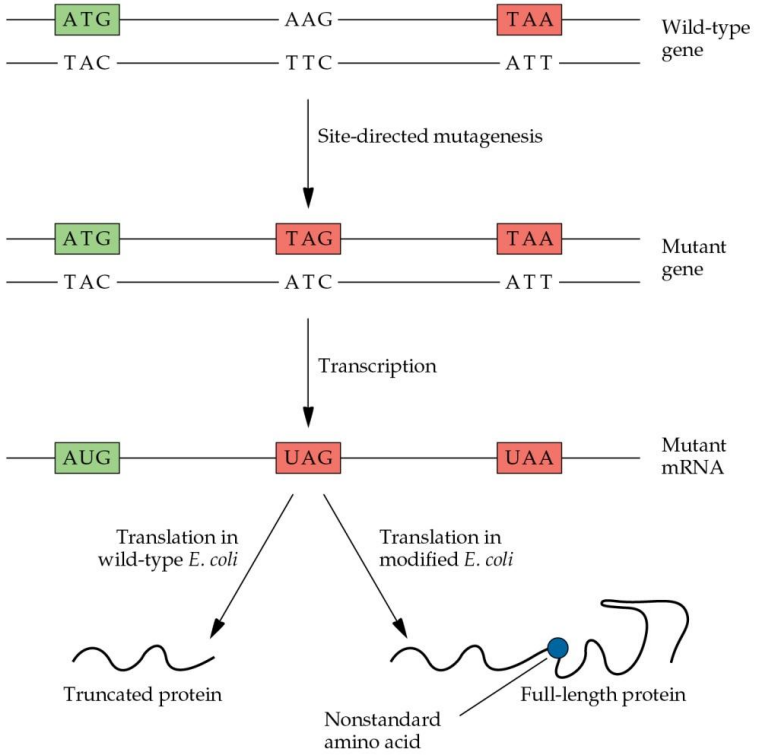
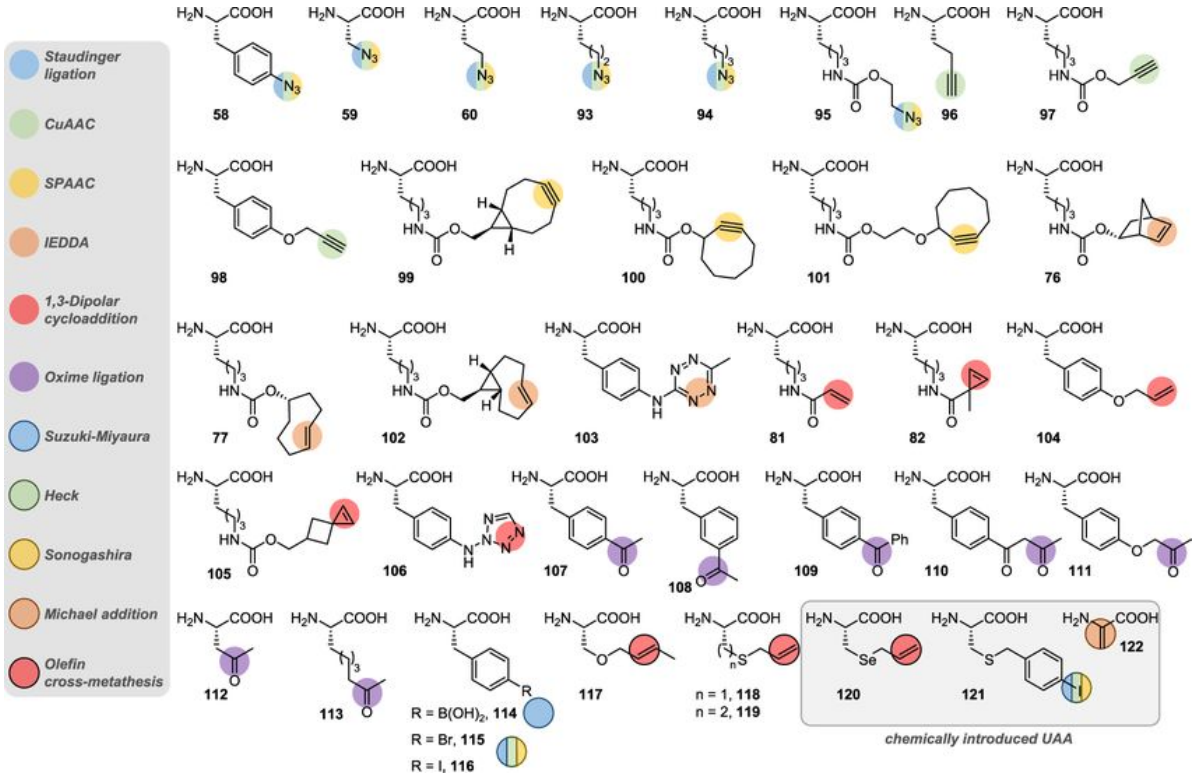



**Figure 3.53** Production of a protein containing a nonstandard amino acid. The start codon is highlighted in green, and the stop codons are in red. The inserted nonstandard amino acid is shown in blue.

# Beyond the Standard 20 Amino Acids

- **Chemical Novelty:** This technique enables the creation of proteins with enhanced catalytic efficiency, specialized binding affinities, or entirely new biological functions.




Sornay et al. (2022). 10.1098/rsos.211563.

# Reassigning the Genetic Code – Stop Codon

- **Amber Suppression:** Target the **UAG (Amber)** stop codon for reassignment because it is the least-utilized termination signal in *E. coli*, which minimizes the disruption host protein synthesis.
- **CUA Anticodon:** Engineer a mutant tRNA with a **CUA anticodon** that recognizes the **UAG** triplet as a specific instruction to add a nonstandard residue rather than stopping.
- **Full-Length Synthesis:** This prevents the production of a truncated or incomplete protein, allowing for the stable synthesis of a **functional recombinant protein** containing the modified amino acid at a precise location.

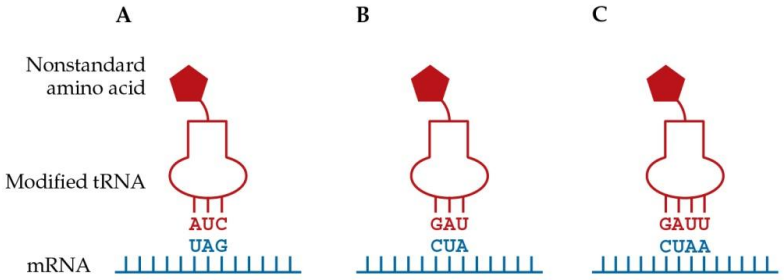



**Figure 3.54** Codons in mRNA may be altered by site-directed mutagenesis of target genes to encode a nonstandard amino acid. Rarely used stop codons **(A)** or redun-dant codons **(B)** may be reassigned to encode a nonstandard amino acid that is cova-lently bound to a modified tRNA. Extended codons of four or more bases **(C)** may be recognized by modified tRNAs. Note that in the example shown here, recognition of the first three bases (CUA) of the four-base codon by a natural tRNA (leucine-tRNA) would result in a reading frame shift; however, this codon is rarely used by *E. coli*, and therefore proteins with the incorrect amino acids would be rare.

# Expanding the Genetic Code – Synthetic Biology

- **Extended Codons:** Use four- or five-base codons to increase the number of possible unique instructions within the genetic code.
- **Reading Frame Control:** These extended codons are recognized by specialized tRNAs to prevent shifts that would otherwise ruin the protein sequence.
- **Novel Properties:** By adding more "letters" to the code, we can tailor the physical and biological properties of recombinant proteins with extreme precision.

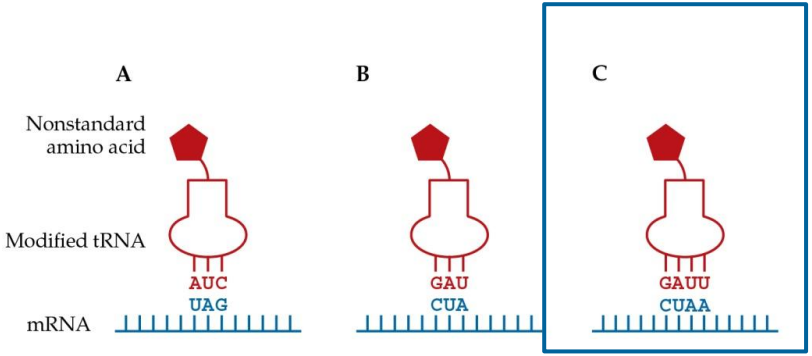
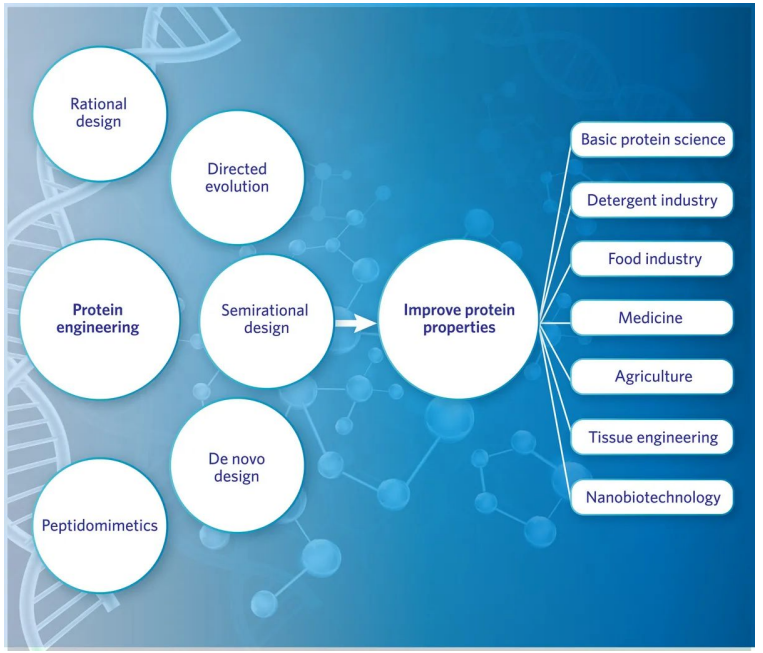



**Figure 3.54** Codons in mRNA may be altered by site-directed mutagenesis of target genes to encode a nonstandard amino acid. Rarely used stop codons **(A)** or redun-dant codons **(B)** may be reassigned to encode a nonstandard amino acid that is cova-lently bound to a modified tRNA. Extended codons of four or more bases **(C)** may be recognized by modified tRNAs. Note that in the example shown here, recognition of the first three bases (CUA) of the four-base codon by a natural tRNA (leucine-tRNA) would result in a reading frame shift; however, this codon is rarely used by *E. coli*, and therefore proteins with the incorrect amino acids would be rare.

# Summary: The Rational Designer

- **Dependency:** Success is tied directly to the availability of high-resolution structural blueprints.
- **Precision:** Allows for 'surgical' edits that can dramatically improve stability or specificity without random guessing.
- **Limitation:** If the 3D structure is unknown, these methods become difficult 'trial-and-error' experiments. This is where AI can step in - Module 2!




Insights Into Protein Engineering: Methods and Applications - Priyom Bose. The Scientist.

# Video - Concepts in the module or a demonstration

- Unleashing the Potential of Genetic Code Expansion - Stop at 3:20
- Expanding the Genetic Code: How Constructive Bio Is Rewriting Biology
- Site-Directed Mutagenesis: Genetic Engineering (2min26s)

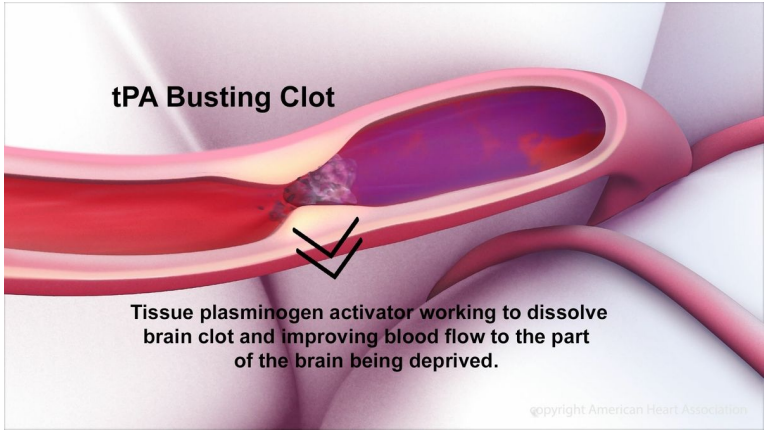
# Module 2: Design Case Studies and AI-driven Protein Design

## (Source Pages: Chapter 3: 165–167, Review Article - Koh et al., 2025: AI-driven protein design doi: 10.1038/s44222-025-00349-8)

- Examine therapeutic protein redesign
- Analyze industrial bioprocessing optimization
- Emerging Impact of AI-driven design

# Case Study: Improving tPA Specificity – Tissue Plasminogen Activator (tPA)

- **Clinical Need:** Tissue plasminogen activator (tPA) dissolves blood clots but is cleared rapidly and can cause nonspecific bleeding.
- **Engineering Goals:** Increase the enzyme's half-life in plasma and improve its targeting specificity for fibrin in clots.
- **Rational Strategy:** Use structural data to identify residues involved in clearance and substrate binding for targeted modification.




American Heart Association

**Fig. 9.8** *Vampire bats have tPA, the model for genetically engineered rtPA.*

# Implementation: The Triple Mutant tPA

- **Persistence:** Changing Thr-103 to Asn allows the drug to persist in plasma 10 times longer than the native form.
- **Targeting:** Mutating a specific Lys-His-Arg-Arg sequence to four Alanines significantly increases "clot-only" targeting.
- **Outcome:** The resulting triple mutant, **Tenecteplase**, is highly effective and produced in CHO cells for proper eukaryotic processing.

**Table 3.18** Stabilities and activities of various modified versions of tPA[a,b]

| tPA variant | Modification(s) | Stability in plasma | Fibrin binding | Activity in plasma | Activity vs clots |
|---|---|---|---|---|---|
| 1 | Thr(103) → Asn | 10 | 0.34 | 0.68 | 0.56 |
| 2 | LysHisArgArg(296–299) → AlaAlaAlaAla | 0.85 | 0.93 | 0.13 | 1.01 |
| 3 | Thr(103) → Asn, LysHisArgArg(296–299) → AlaAlaAlaAla | 5.3 | 0.33 | 0.13 | 0.65 |
| 4 | Thr(103) → Asn, Asn(117) → Gln | 3.4 | 1.0 | 1.13 | 1.17 |
| 5 | LysHisArgArg(296–299) → AlaAlaAlaAla, Asn(117) → Gln | 1.2 | 1.33 | 0.16 | 1.38 |
| 6 | Thr(103) → Asn, LysHisArgArg(296–299) → AlaAlaAlaAla, Asn(117) → Gln | 8.3 | 0.87 | 0.06 | 0.85 |

[a]Data from Keyt et al., *Proc. Natl. Acad. Sci. USA* **91**:3670–3674, 1994.

[b]All of the values shown are normalized to the wild type. Plasma stability is the reciprocal of the time it takes for plasma clearance; larger numbers indicate a more stable derivative. Fibrin specificity is reflected by a high activity versus clots and a low activity in plasma.

# Case Study – Metabolic Waste: The Lactate Bottleneck

- **Oxygen Limitation:** Low-oxygen conditions in large bioreactors force cells to shift toward inefficient anaerobic metabolism.
- **Lactate Accumulation:** Pyruvate is converted to lactate by dehydrogenase instead of entering the productive TCA cycle.
- **Growth Inhibition:** Acidification of the culture medium by lactate reduces cell viability and final protein yields.

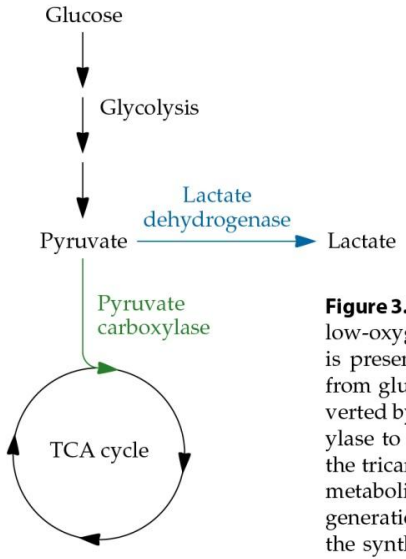



**Figure 3.48** Production of lactate under low-oxygen conditions. When oxygen is present, pyruvate, which is formed from glucose during glycolysis, is converted by the enzyme pyruvate carboxylase to an intermediate compound in the tricarboxylic acid (TCA) cycle. This metabolic pathway is important for the generation of cellular energy and for the synthesis of biomolecules required for cell proliferation. However, under low-oxygen conditions, such as those found in large bioreactors, pyruvate carboxylase has a low level of activity. Under these conditions, lactate dehydrogenase converts pyruvate into lactate, which yields a lower level of energy. Cultured cells secrete lactate, thereby acidifying the medium.

# Case Study: Solving the Lactate Problem Approach 1 — Rational Overexpression

- **The Rational Edit:** Overexpress the human **pyruvate carboxylase** gene targeted to the mitochondria.
- **Mechanism:** This forces pyruvate into the TCA cycle by converting it to oxaloacetate, bypassing the lactate "waste" route.
- **Results:** Achieved a 27% reduction in the rate of lactate production and a 15% decrease in overall lactate yield.

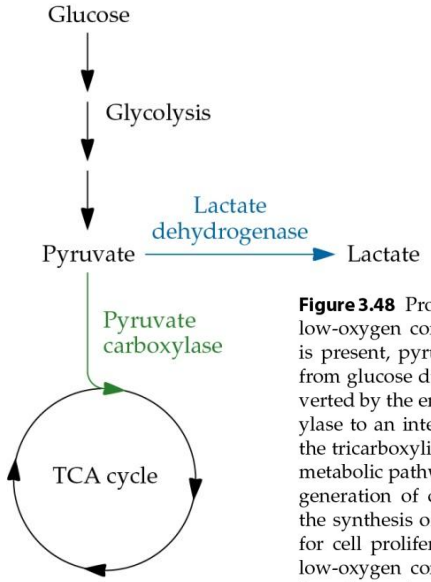



**Figure 3.48** Production of lactate under low-oxygen conditions. When oxygen is present, pyruvate, which is formed from glucose during glycolysis, is converted by the enzyme pyruvate carboxylase to an intermediate compound in the tricarboxylic acid (TCA) cycle. This metabolic pathway is important for the generation of cellular energy and for the synthesis of biomolecules required for cell proliferation. However, under low-oxygen conditions, such as those found in large bioreactors, pyruvate carboxylase has a low level of activity. Under these conditions, lactate dehydrogenase converts pyruvate into lactate, which yields a lower level of energy. Cultured cells secrete lactate, thereby acidifying the medium.

# Case Study: Solving the Lactate Problem Approach 2 — Systems metabolic engineering

- **The Logic:** Previous attempts to simply knock out *Ldha* were lethal because cells couldn't maintain redox balance (NAD+ regeneration).
- **The Genomics Edit:** Multiplex CRISPR-Cas9 knockout of *Ldha* **AND** all four Pyruvate Dehydrogenase Kinases (*Pdk1–4*).
- **Mechanism:** Knocking out *Pdks* removes the "locks" (inhibition) on the *Pdh* complex, allowing uninhibited flow of pyruvate into the mitochondria even at high speeds.
- **Industrial Impact:** Produced negligible lactate, maintained high growth rates, and reached titers of ~3 g/L without needing pH control.

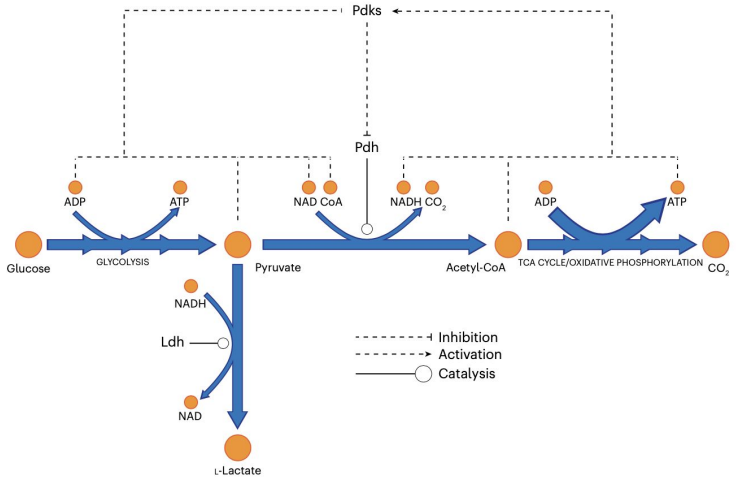
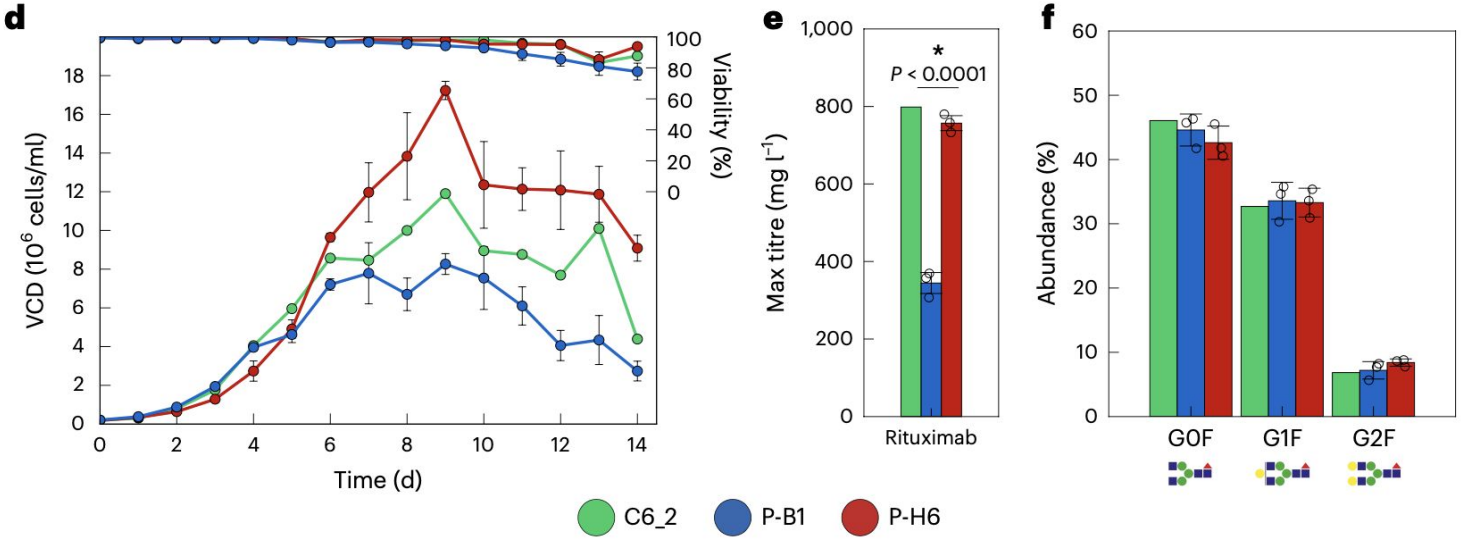



**Fig. 1 | The Warburg effect is influenced by a regulatory circuit involving multiple metabolites and proteins and can be eliminated by multiplex knockout. a**, Pyruvate sits at a branch point between fermentation through Ldh or oxidative metabolism starting with the Pdh complex. Pdks are regulated by the products and substrates of the Pdh reaction, forming a negative feedback loop that reinforces an increase in lactate secretion when glycolytic flux is high.

Hefzi et al. Nature Metabolism, 10.1038/s42255-024-01193-7

# Case Study: Solving the Lactate Problem
# Approach 2 — Systems metabolic engineering



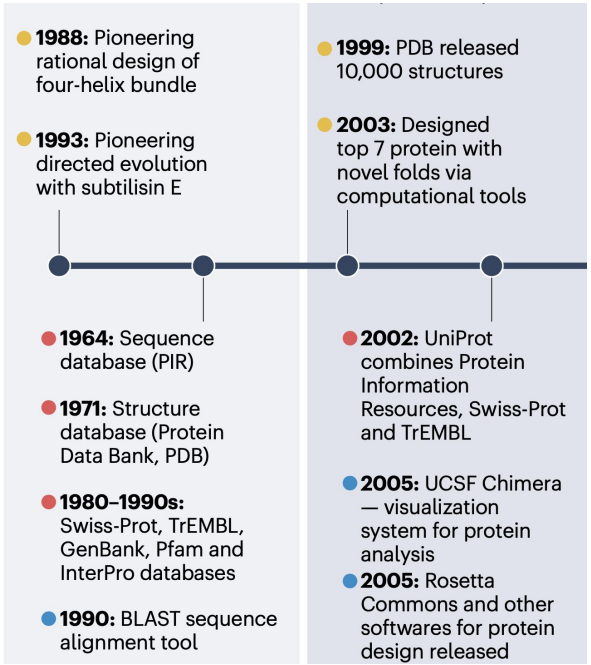

**d**–**f**, Separately, parental (C6_2, green, $n = 1$ bioreactor), mock control (P-B1, blue, $n = 3$ bioreactors) and Warburg-null (P-H6, red, $n = 3$ bioreactors) cells producing rituximab were grown in fed-batch culture. **d**, Knockout cells showed a prolonged period of exponential growth compared to both the parental and control lines (see Supplementary Fig. 1a for maximum lactate concentrations for all clones). **e**, Knockout cells were able to maintain (versus parental) or improve (versus control) product titre (*$P < 0.05$ as determined by a two-sample, two-tailed $t$-test between mock and Warburg-null clones). **f**, Warburg-null clones maintained comparable glycosylation of rituximab at day 14 of the culture. Data are shown

Hefzi et al. Nature Metabolism, 10.1038/s42255-024-01193-7

# Why AI Succeeds in Shape Prediction – Great Data!

- **Data Foundations:** Decades of experimental data in the **Protein Data Bank (PDB)** provide high-resolution "Rosetta Stones" linking sequence to structure
- **Pattern Recognition:** Supervised learning models infer the hidden **biophysical 'grammar'** and residue interactions that traditional rule-based software often misses
- **Evolutionary Insights:** AI leverages **Multiple Sequence Alignments (MSAs)** to identify co-evolving residues that are physically touching in 3D space.

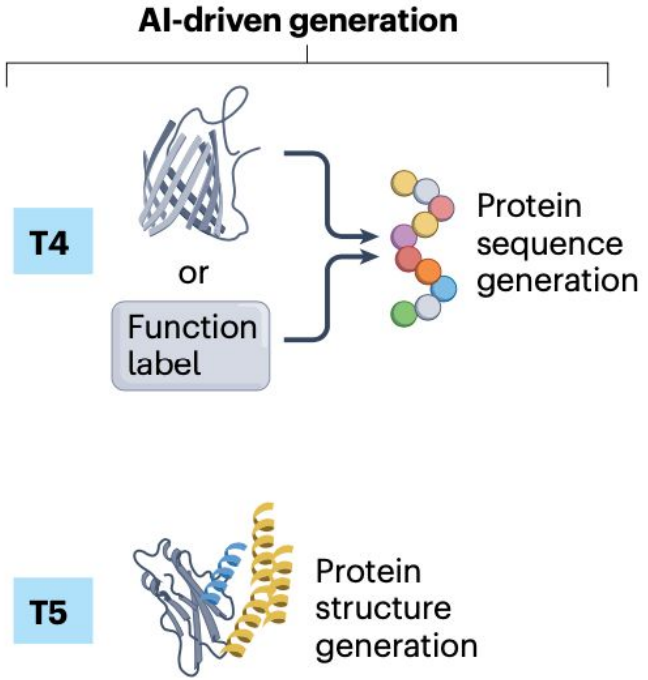
Fig. 1 | Historical development of artificial intelligence-driven protein design tools.




Koh et al., 2025: AI-driven protein design doi.org/10.1038/s44222-025-00349-8

# The AI Shift: Predictive Design

- **Astronomical Space:** A standard protein has approx $10^{455}$ possible sequences, making exhaustive wet-lab testing physically impossible.
- **Predictive Discipline:** AI transforms design from trial-and-error into a discipline that predicts successful structures on a computer.
- **Beyond Templates:** AI navigates this space to design ***De Novo* proteins** "from scratch" that have no counterparts in nature.




Koh et al., 2025: AI-driven protein design doi.org/10.1038/s44222-025-00349-8

# Cornerstone AI Toolkits

- **Structure Prediction: AlphaFold 2 / 3** predict 3D folds with near-experimental accuracy directly from sequence data.
- **Inverse Folding: ProteinMPNN** generates sequences that reliably fold into specific, pre-defined structural backbones with high throughput.
- **Multi-objective Design:** AI simultaneously optimizes for function, structure, and developability (solubility and yield).

**Table 1 | Overview of artificial intelligence toolkits and applications in protein design workflow**

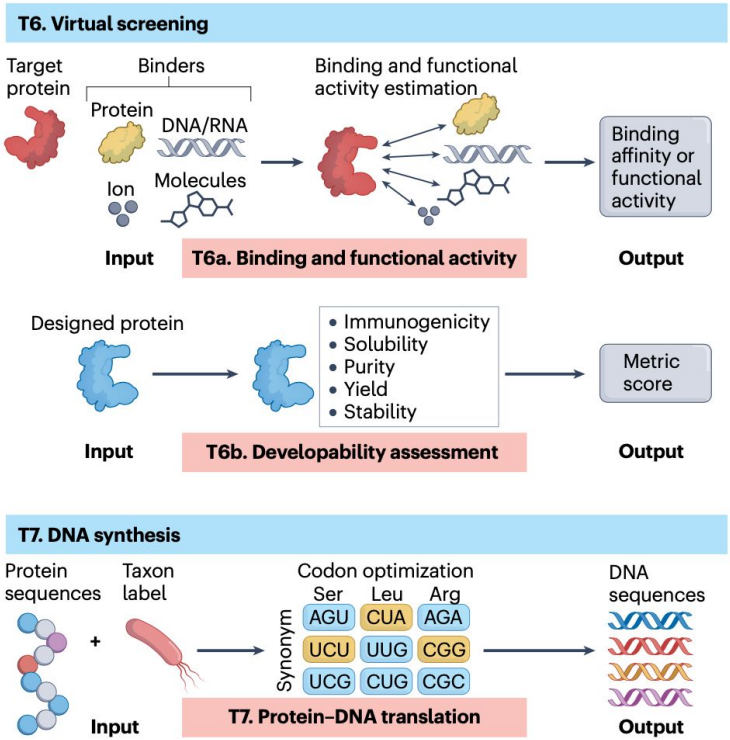
| | | |
|---|---|---|
| T2. Protein structure prediction | T2a. Protein folding | Targeting functional sites$_{(DE.2)}$<br>Sequence folding validation$_{(RD.2)}$<br>Targeting key residues$_{(RD.3)}$ |
| | T2b. Biomolecular co-folding | Targeting functional sites$_{(DE.2)}$<br>Sequence folding validation$_{(RD.2)}$<br>Targeting key residues$_{(RD.3)}$ |
| | T2c. Structure stability prediction | Targeting stability regions$_{(DE.2)}$<br>Sequence folding validation$_{(RD.2)}$<br>Targeting key residues$_{(RD.3)}$ |

Koh et al., 2025: AI-driven protein design doi.org/10.1038/s44222-025-00349-8

# De Novo Structure Generation

- **Diffusion Models:** Like AI image generators, these models (e.g., RFDiffusion) 'dream' up brand new protein shapes from noise.
- **Binder Design:** RFDiffusion can create ultra-stable proteins designed specifically to wrap around a target like the SARS-CoV-2 spike.
- **Virtual Screening:** *In silico* tools filter millions of AI designs to identify the top candidates before lab synthesis.

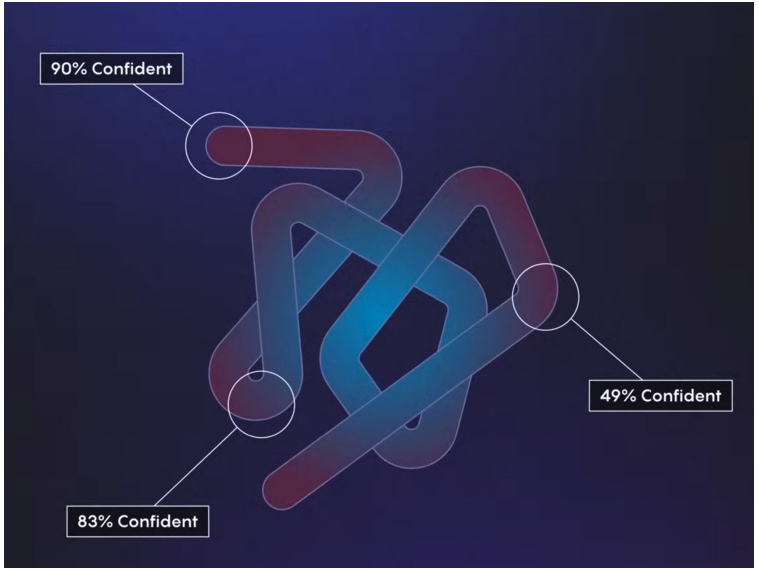
Fig. 3 | Artificial intelligence toolkits for protein design




Koh et al., 2025: AI-driven protein design
doi.org/10.1038/s44222-025-00349-8

# Summary: The Modern Bioengineer

- **Hybrid Success:** Modern design combines deep structural knowledge (Rational) with the predictive speed of machine learning (AI).
- **Cost Efficiency:** *In silico* screening reduces the burden of wet-lab validation, focusing resources only on high-potential candidates.
- **Next Step Transition:** These methods require a known structure; next, we study **Evolutionary Engineering** for when we don't.




Quanta Magazine - DeepMind's AlphaFold

# Video - Concepts in the module or a demonstration

- The New Era of AI-Powered Protein Design (4min33s)

**More videos**

- How AI Cracked the Protein Folding Code and Won a Nobel Prize
  - Many great concepts and content for learning here
    - 12:03 mins talks about AI
    - 18:24 talks about design
- David Baker explains his Nobel Prize research on protein design
  - kind of long and a general talk

# The End