# Study and engineering of gene function: mutagenesis

I. Random mutagenesis, mutant selection schemes
II. Site-directed mutagenesis, assembly of new DNA fragments
III. Rational engineering of proteins
IV. Genetic code expansion

# Readings

*1) 34 MC4 Mutagenesis*. Very short summary of types of mutagenesis techniques, with lists of specific techniques.

2) *35 MC4 Mutagenesis techniques.* Some specific mutagenesis protocols.

*3) GFP mut 1994*. A mutant of Green Fluorescent Protein with different spectral properties.

*4) Gene shuffling for glyphosate resistance 2004.*

*5) Unnatural amino acids 2005*. A review concerning systems encoding new amino acids.

# Mutations change the genetic program – does this influence biomolecular function?

- What is the biological importance of a protein or non-coding RNA – what happens to the organism when the gene is mutated or deleted?
- Which parts of a protein or RNA influence function of that biomolecule?
- How can we intentionally alter or improve protein or RNA function, using random or directed mutation?

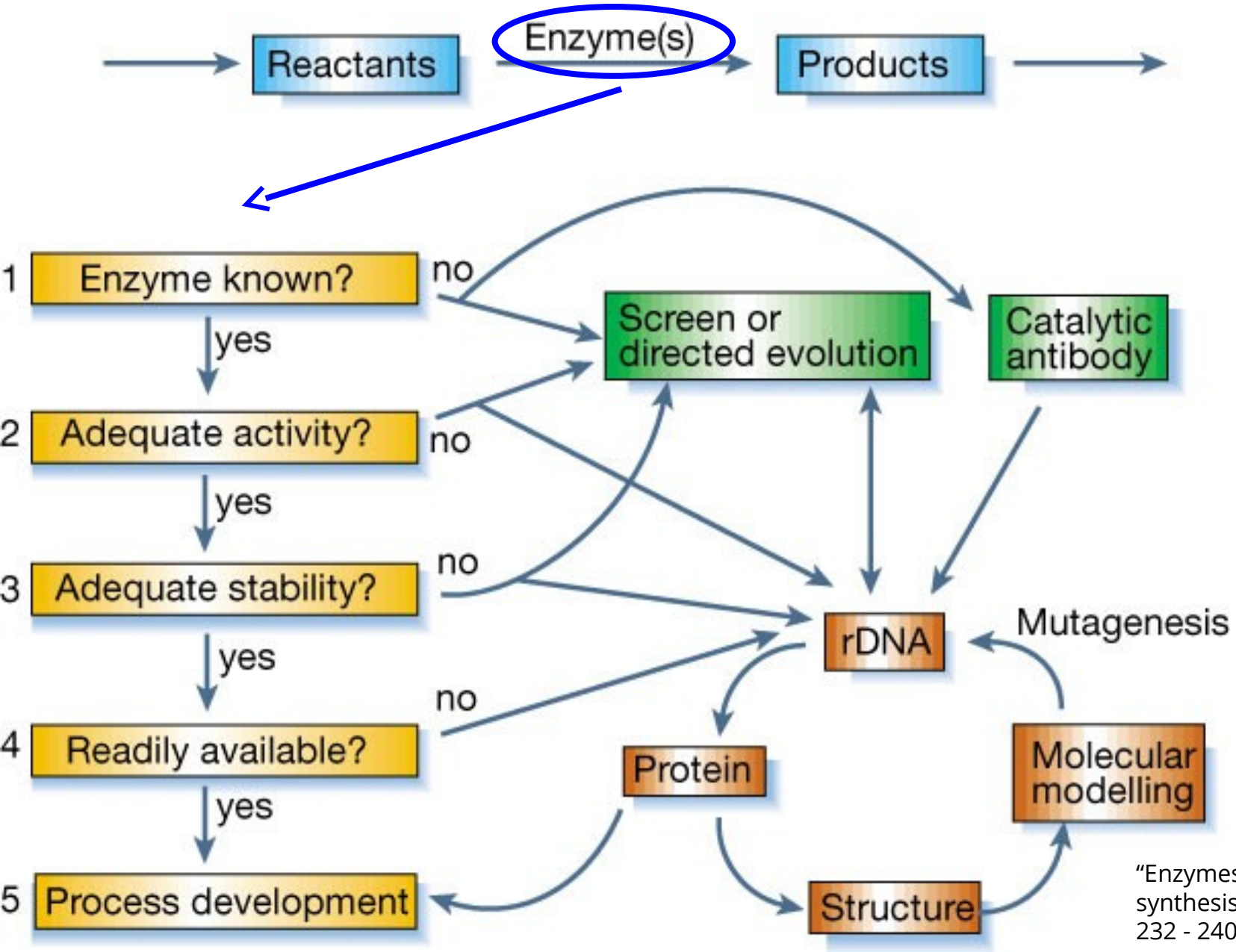
# Many enyzmes are commercially important, e.g.:

| Enzyme | Industrial use(s) |
|---|---|
| α-Amylase | Beer making, alcohol production |
| Aminoacylase | Preparation of L-amino acids |
| Bromelain | Meat tenderizer, juice clarification |
| Catalase | Antioxidant in prepared foods |
| Cellulase | Alcohol and glucose production |
| Ficin | Meat tenderizer, juice clarification |
| Glucoamylase | Beer making, alcohol production |
| Glucose isomerase | Manufacture of high-fructose syrups |
| Glucose oxidase | Antioxidant in prepared foods |
| Invertase | Sucrose inversion |
| Lactase | Whey utilization, lactose hydrolysis |
| Lipase | Cheese making, preparation of flavorings |
| Papain | Meat tenderizer, juice clarification |
| Pectinase | Clarifying fruit juices, alcohol production |
| Protease | Detergent, alcohol production |
| Rennet | Cheese making |

# Protein engineering through mutation

- Improve stability/function under new conditions
  - temperature, pH, organic/aqueous solvent, salt, etc.
- Change enzyme substrate specificity
- Enhance enzymatic rate
- Change molecular binding properties
- Create brand new enzymes

# Obtaining and improving useful enzymes




"Enzymes for chemical synthesis" , Nature 409, 232 - 240 (2001)

# Introduce random mutations followed by some kind of screen or selection: simulated evolution

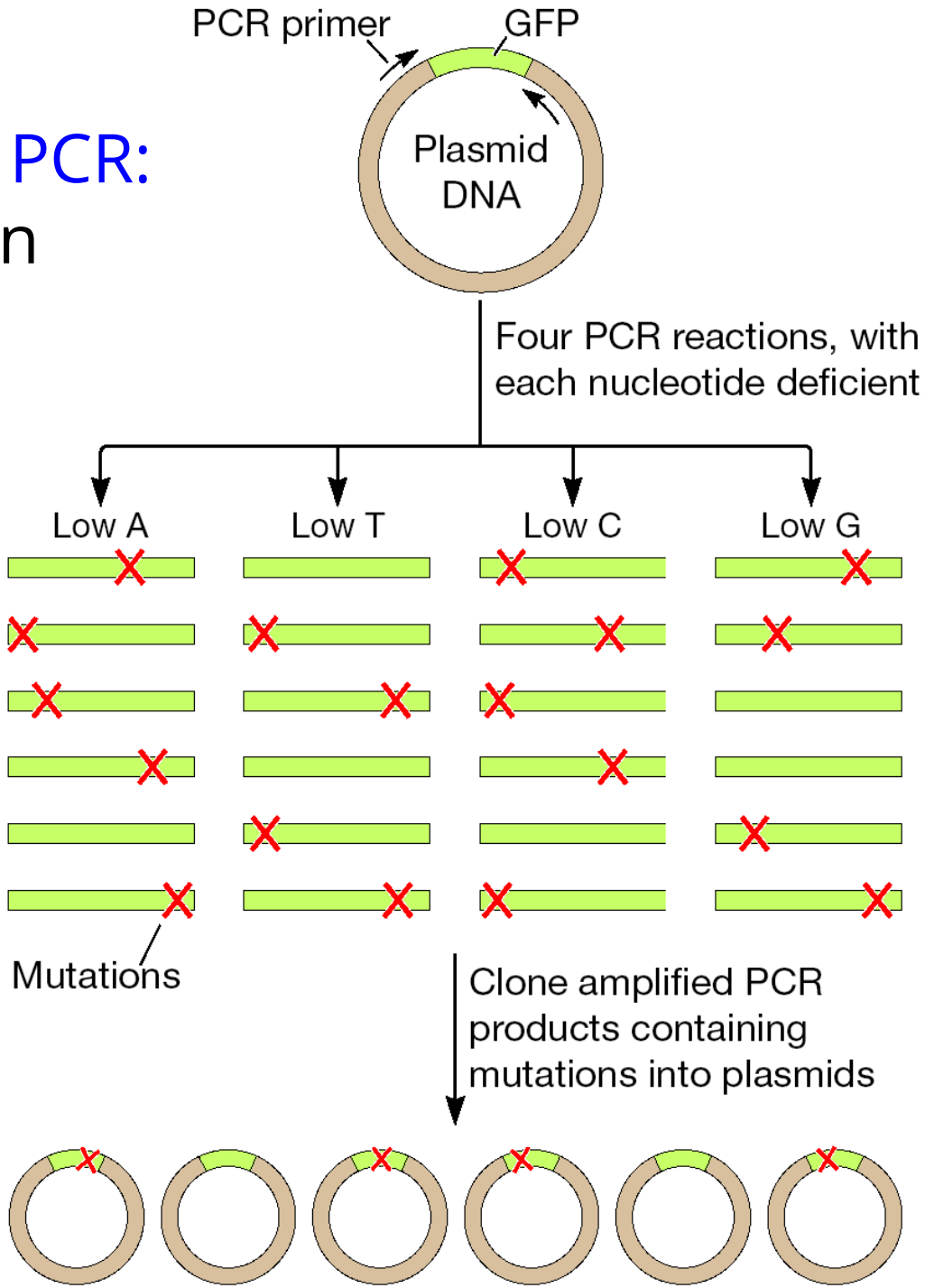
a) General mutagenesis: expose the organism or its DNA to chemical, UV, or radiation mutagen

b) PCR mutagenesis: base misincorporation
   - Include $Mn^{2+}$ in reaction
   - Reduce concentration of one dNTP

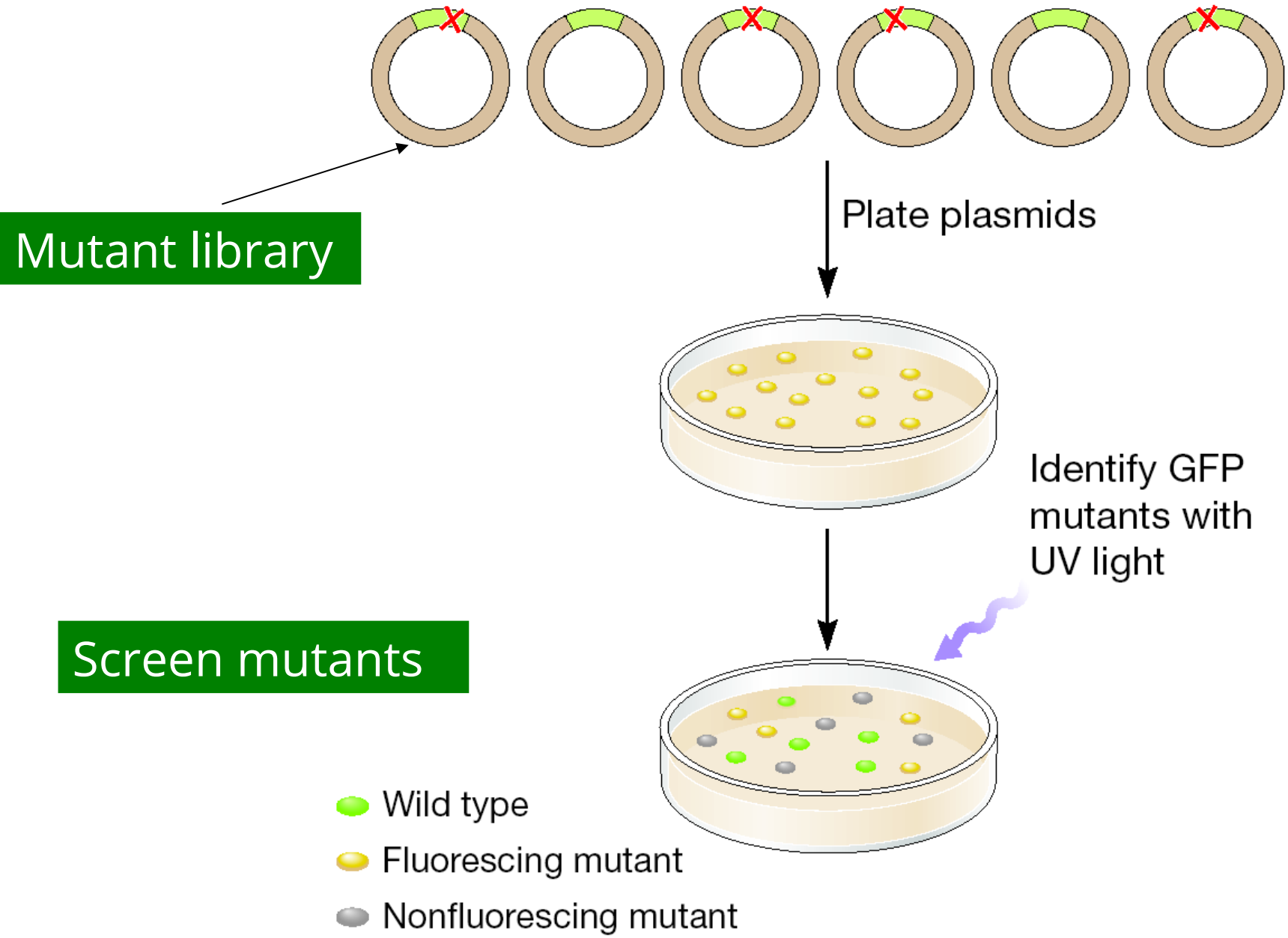
c) 'Cassette' mutagenesis
   - Partly randomized oligonucleotides used in cloning

- Many possible mutations in a LARGE library of clones
- There has to be some kind of phenotypic screen or selection for the valuable sub-population

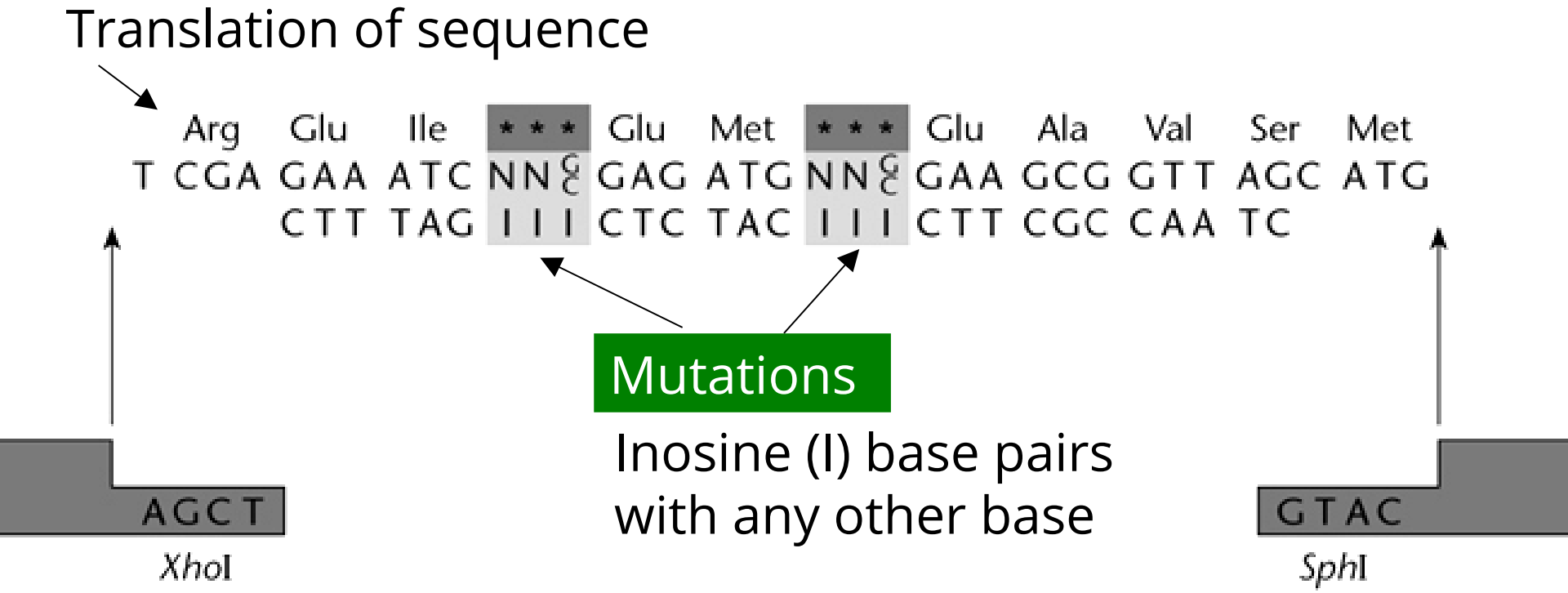
# Random mutagenesis by PCR:
Green Fluorescent Protein (GFP)



# Random mutagenesis by PCR:
the Green Fluorescent Protein



# Cassette mutagenesis (semi-random)




Synthetic cassette DNA strands are synthesized, and annealed

The cassette is ligated to make library, then screened

Allows random insertion of any amino acid at defined positions

# Random and semi-random mutagenesis: directed evolution

- Mutagenize existing protein, eg. error-prone PCR, randomized oligo cassette mutagenesis

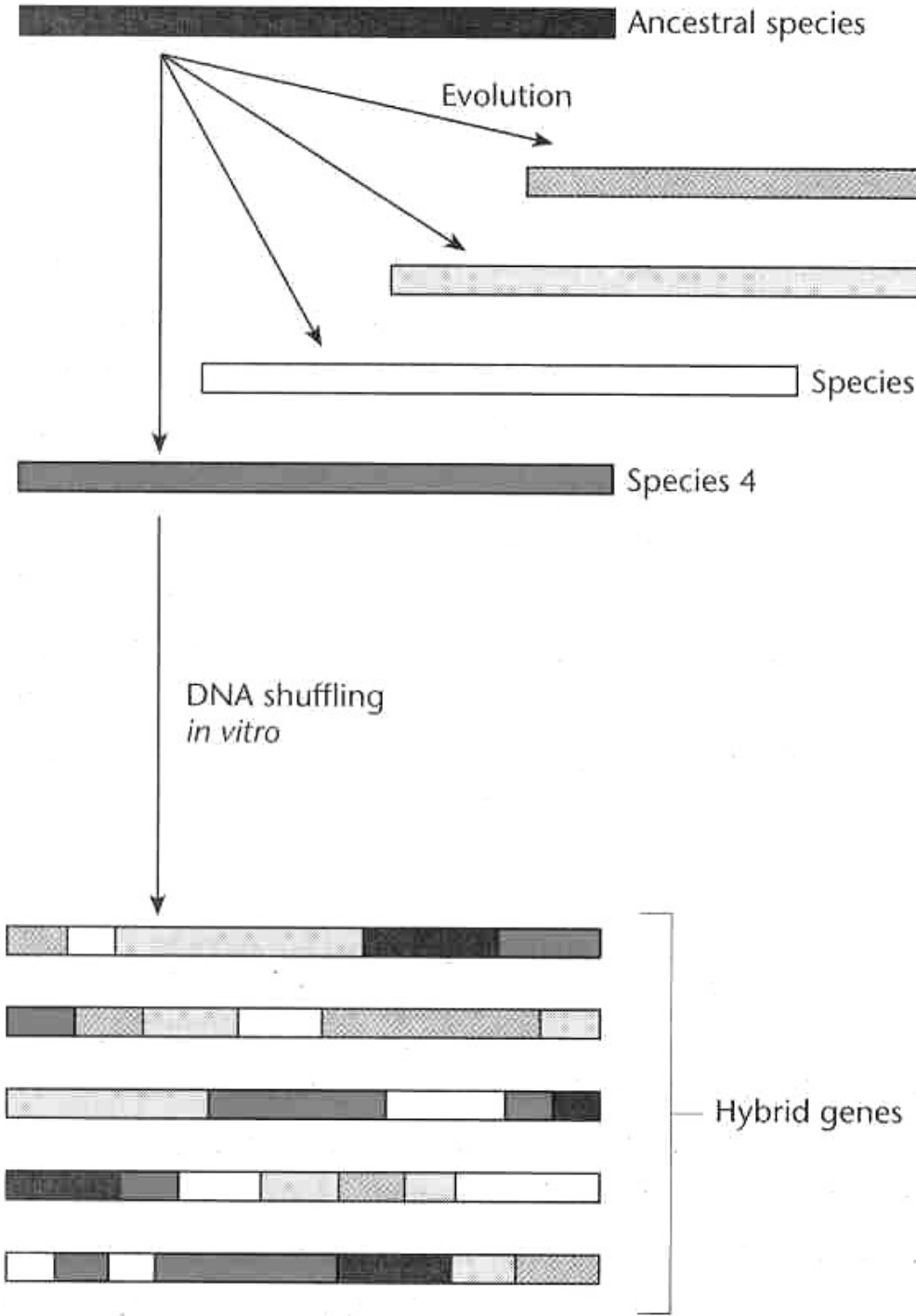
  -- and/or --

  Do “ gene shuffling”

- Create library

- Screen library of mutations for proteins with altered properties
  - Standard plate screen: 10,000-100,000 mutants
  - (Phage display: $10^9$ mutants)

# PCR shuffling – like meiotic recombination



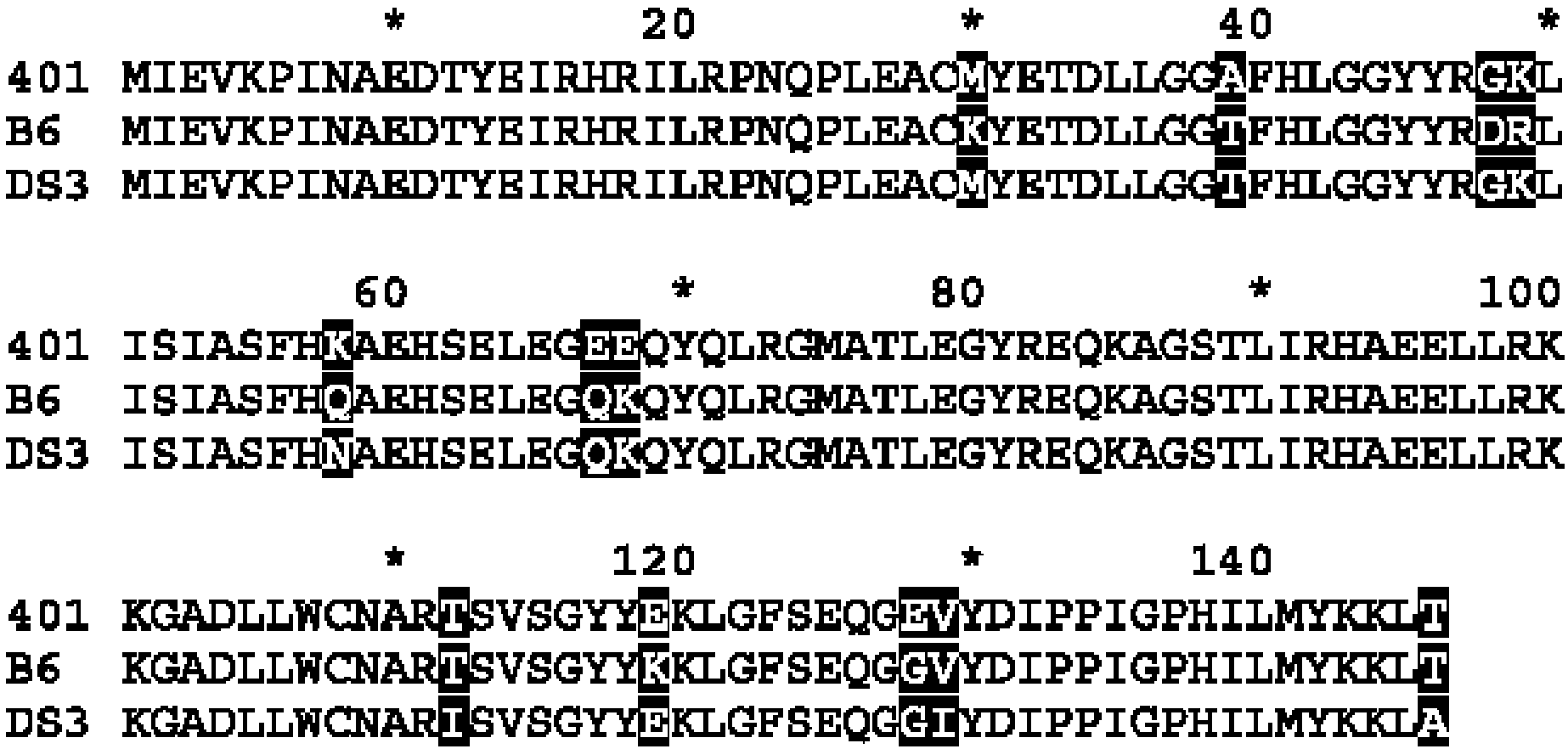

For gene shuffling protocols, have related genes in original pool:
1) evolutionary variants, or
2) variants mutated in vitro

Shuffling allows rapid scanning through sequence space:
faster than doing multiple rounds of random mutagenesis and screening

# GAT: an example for gene shuffling

- GAT gene encodes glyphosate N-acetyl transferase
- Background: glyphosate resistance (in “roundup ready” crops) can be conferred to plants with a microbial EPSPS gene, which is indifferent to glyphosate while the plant version is susceptible.
- EPSPS doesn’t modify glyphosate, allowing the glyphosate to accumulate in the plant, potentially causing problems with crop yields
- GAT acetylates glyphosate, making N-acetyl glyphosate which is not herbicidal
- Several GAT genes were isolated from bacteria by a biotech firm – they provide an alternative to EPSPS
- *The genes weren’t very active, so shuffling was done*

# Three GAT genes from *Bacillus licheniformis*




Sites of natural variation are shaded

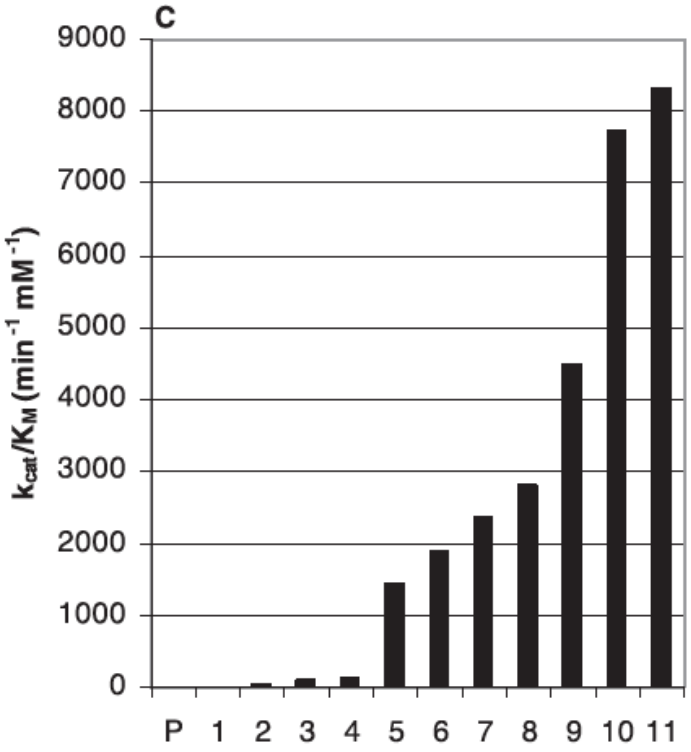
Discovery and Directed Evolution of a Glyphosate Tolerance Gene

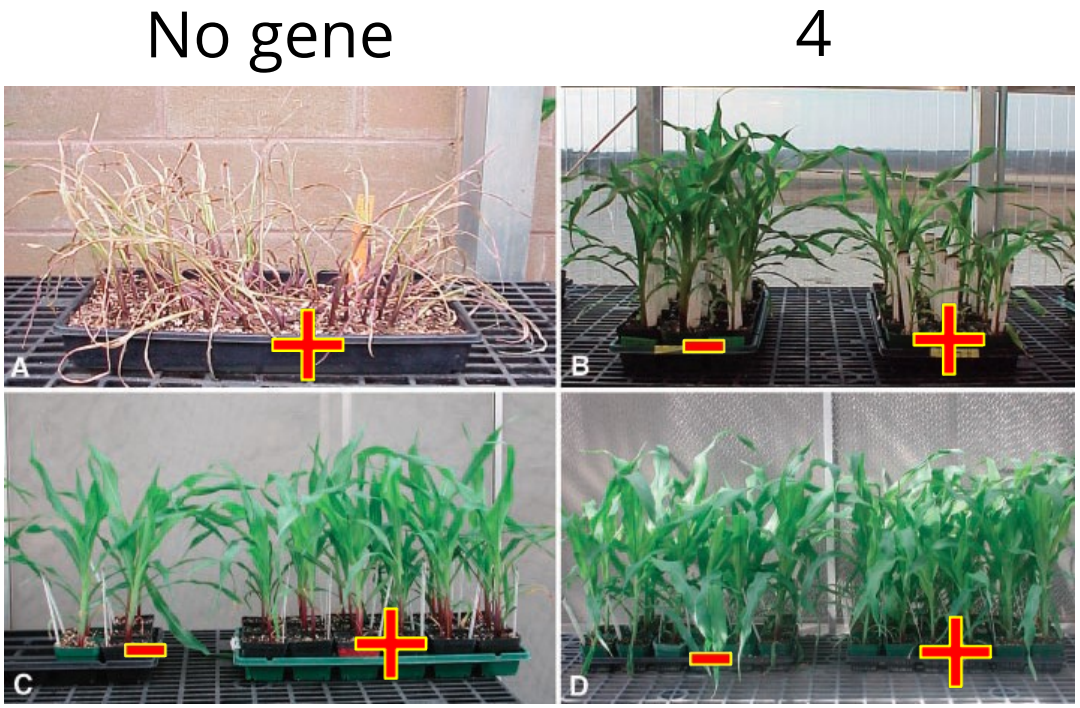
Linda A. Castle,[1]* Daniel L. Siehl,[1] Rebecca Gorton,[1] Phillip A. Patten,[2] Yong Hong Chen,[2] Sean Bertain,[1] Hyeon-Je Cho,[1] Nicholas Duck,[3]† James Wong,[3] Donglong Liu,[3] Michael W. Lassner[1]

# Mutant creation, screening, and assays

- The three GAT genes were fragmented and shuffled by PCR over 11 rounds
- Additional mutations were added at rounds 4 and 8, based on sequences of related genes in other *Bacillus* species
- After each shuffle, libraries were created
- 5000 gene variants were screened for GAT activity
- Top performing mutants were selected for further shuffling in each round

# Top performer (Kcat/Km) for each round



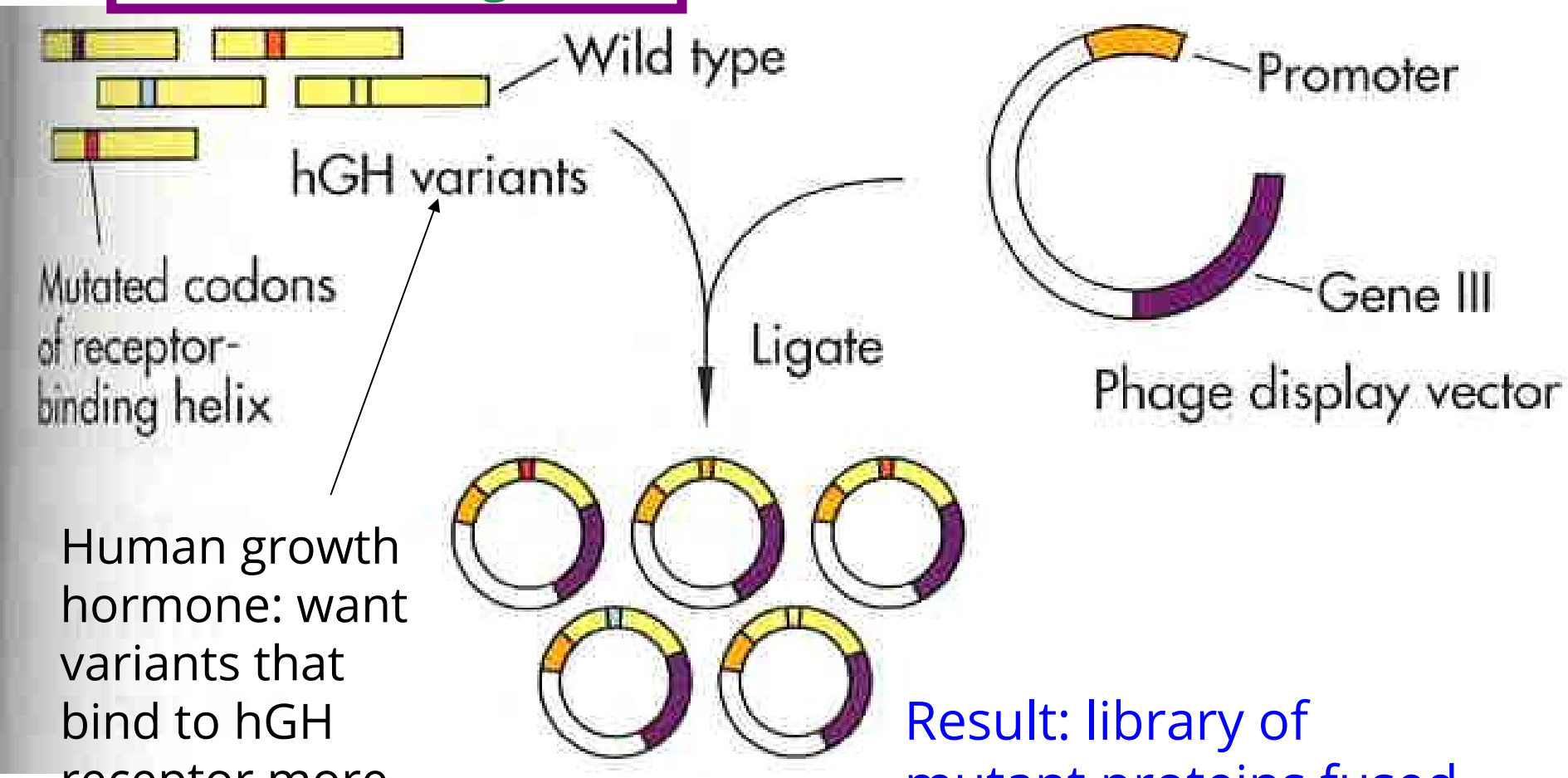

7 11

```
*           *           *           *           *          60
MIEVKPINAEDTYDLRHRVLRPNQPIEACMFESDLTRSAFHLGGFYGGKLISVASFHQAE

*           *           *           *           *         120
HSELQGKKQYQLRGVATLEGYREQKAGSSLVKHAEEILRKRGADMIWCNARTSASGYYRK

*           *         146
LGFSEQGEVFDTPPVGPHILMYKRIT
```

Blue: original diversity, Green: added rd.4, Pink: added rd. 8, Orange: random occurrence

# Screening huge ($10^9$) libraries: phage display

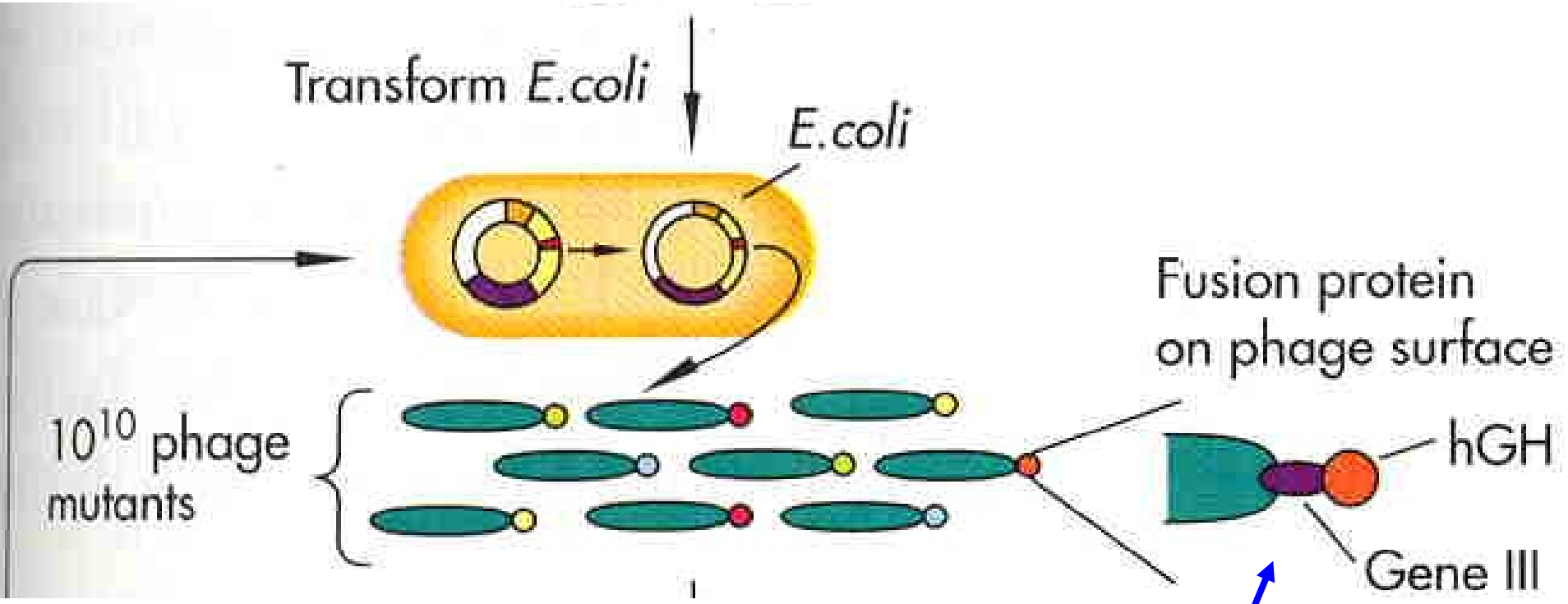
Random mutagenesis




Human growth hormone: want variants that bind to hGH receptor more tightly

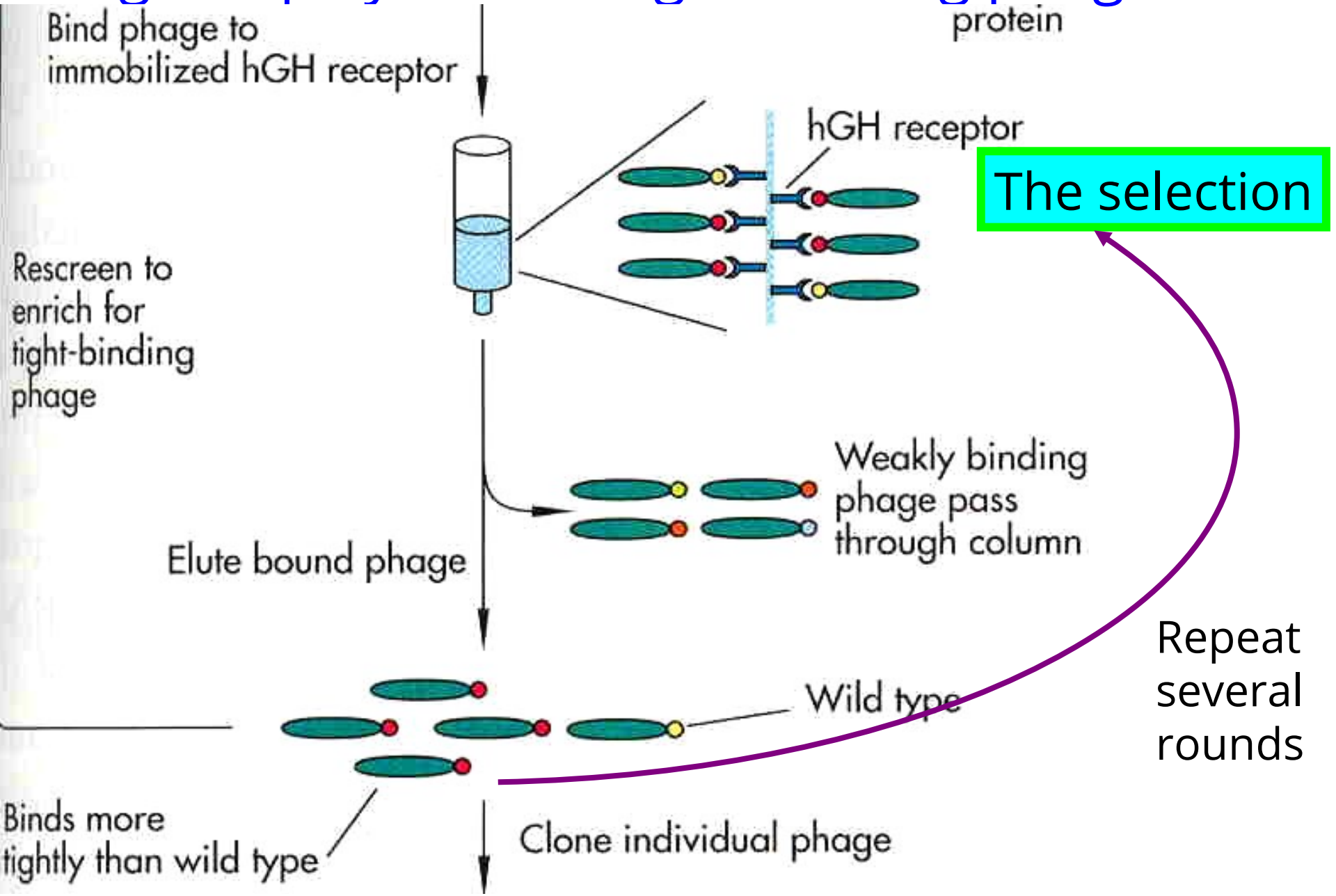
Result: library of mutant proteins fused to M13 gene III

# Phage display:
# production of recombinant phage

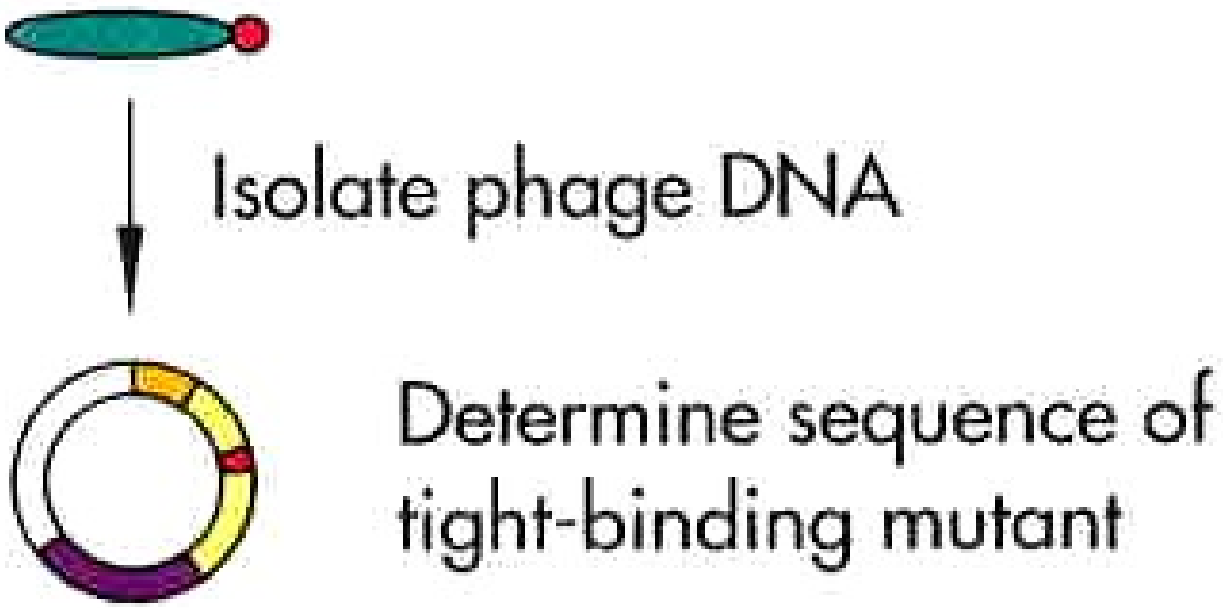



The “ display” gives an easy to select phenotype

# Phage display: collect tight-binding phage



# Which sequences were selected?

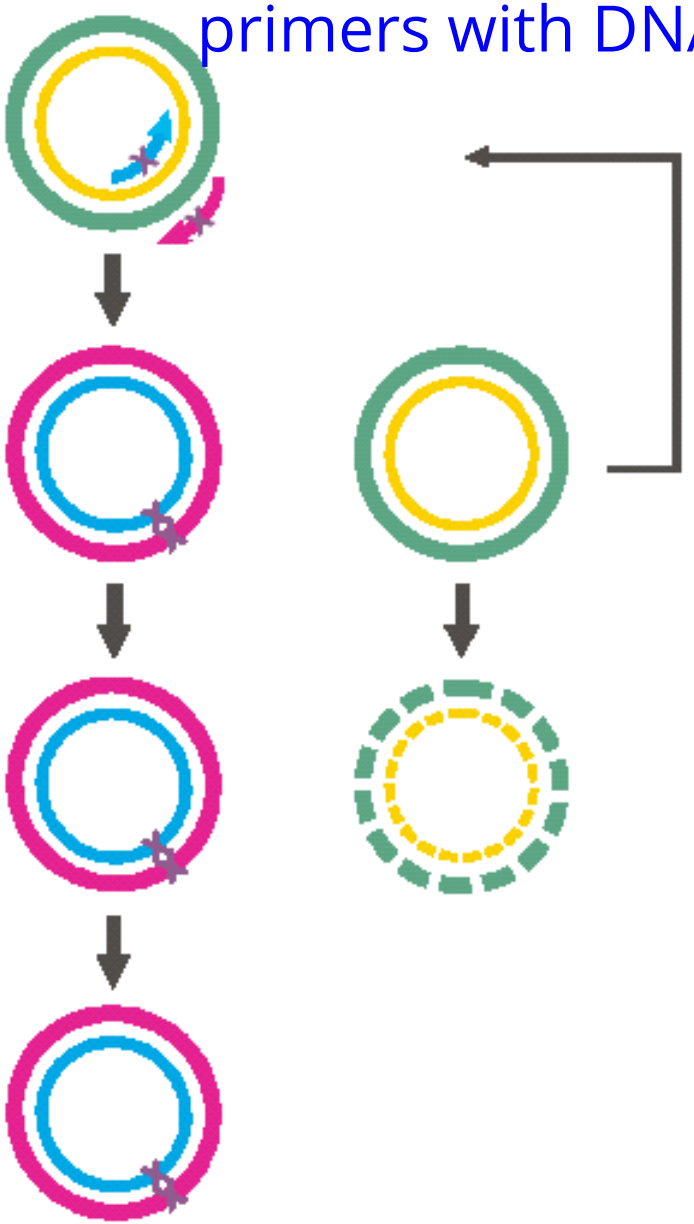


# Site directed mutagenesis: when you know exactly what you want to make

- Changes to amino acids that are likely to be important for function
- New combinations of protein domains
- Rational re-design of protein structure/function

- Addition of new amino acids to the genetic code

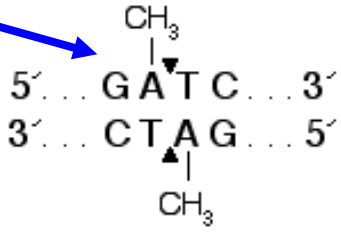
# Site-directed mutagenesis: extension of 2 mutagenic primers with DNA polymerase: the PCR/*Dpn* I protocol



**Mutant Strand Synthesis**
Perform thermal cycling to:
1) Denature DNA template
2) Anneal mutagenic primers containing desired mutation
3) Extend primers with *PfuUltra* DNA polymerase

Template DNA is destroyed, only mutated DNA remains

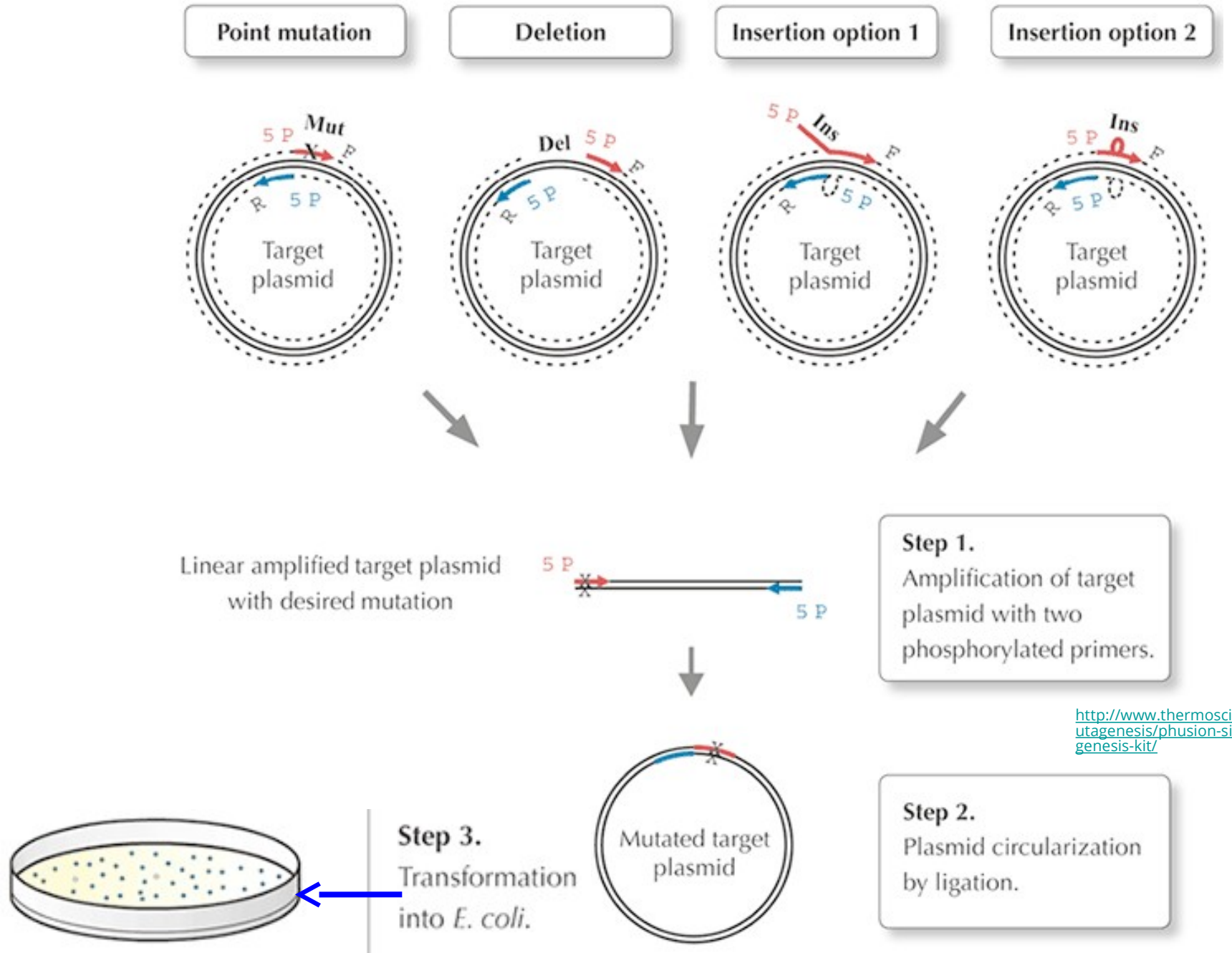
***Dpn* I Digestion of Template**
Digest parental methylated and hemimethylated DNA with *Dpn* I




**Transformation**
Transform mutated molecule into competent cells for nick repair

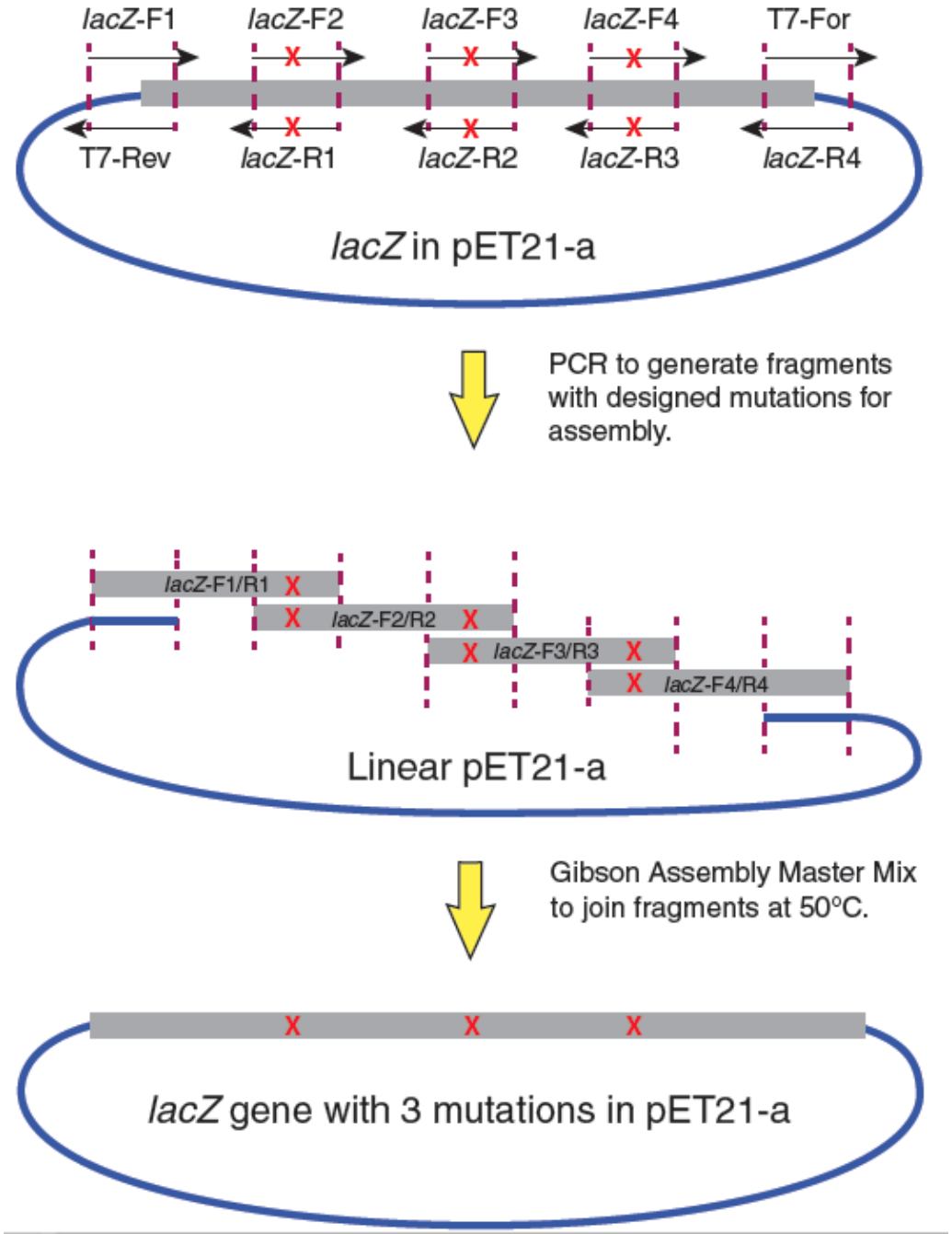
Uses double-stranded plasmid DNA

# Site directed mutagenesis: plasmid PCR/ligation/transformation




http://www.thermoscientificbio.com/mutagenesis/phusion-site-directed-mutagenesis-kit/

# Site directed mutagenesis: Gibson assembly




Multiple, overlapping DNA fragments, each with a mutation

Gibson assembly stitches the fragments together

From: New England Biolabs/Synthetic Genomics

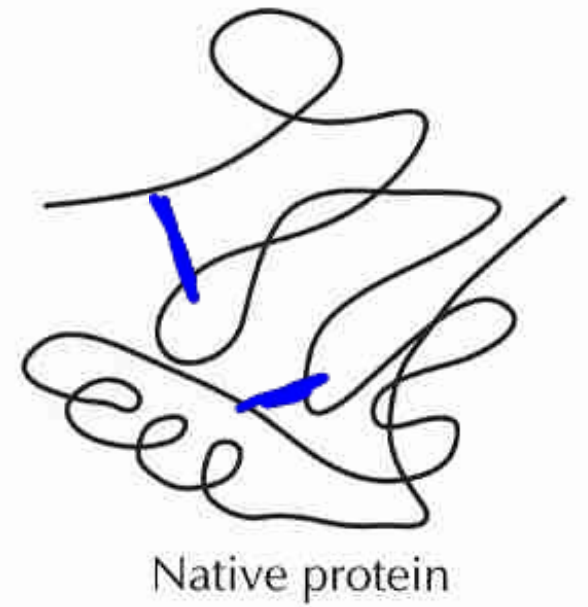
# Rational site-directed mutagenesis

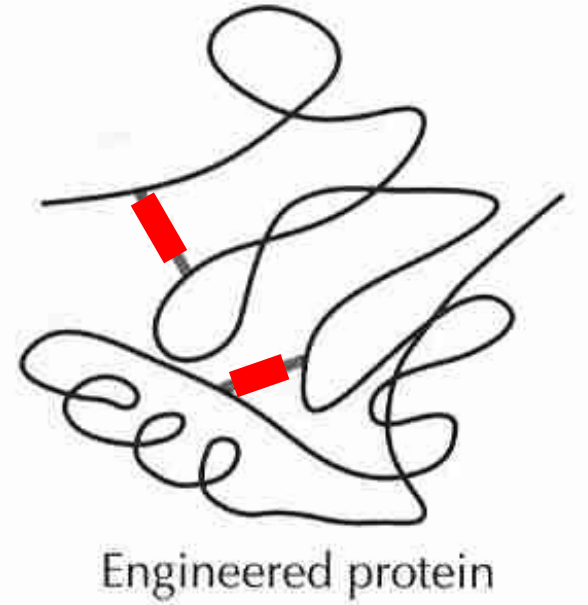
Re-program a protein by making deliberate changes in amino acid sequence, based on structures or other prior knowledge

- If structure is known, mutate amino acids in any part of protein thought to influence activity, interactions, stability, etc.

- For a protein with multiple family members: mutate desired protein in positions that bring it closer to another family member with desired properties

# Site-directed mutagenesis: T4 lysozyme



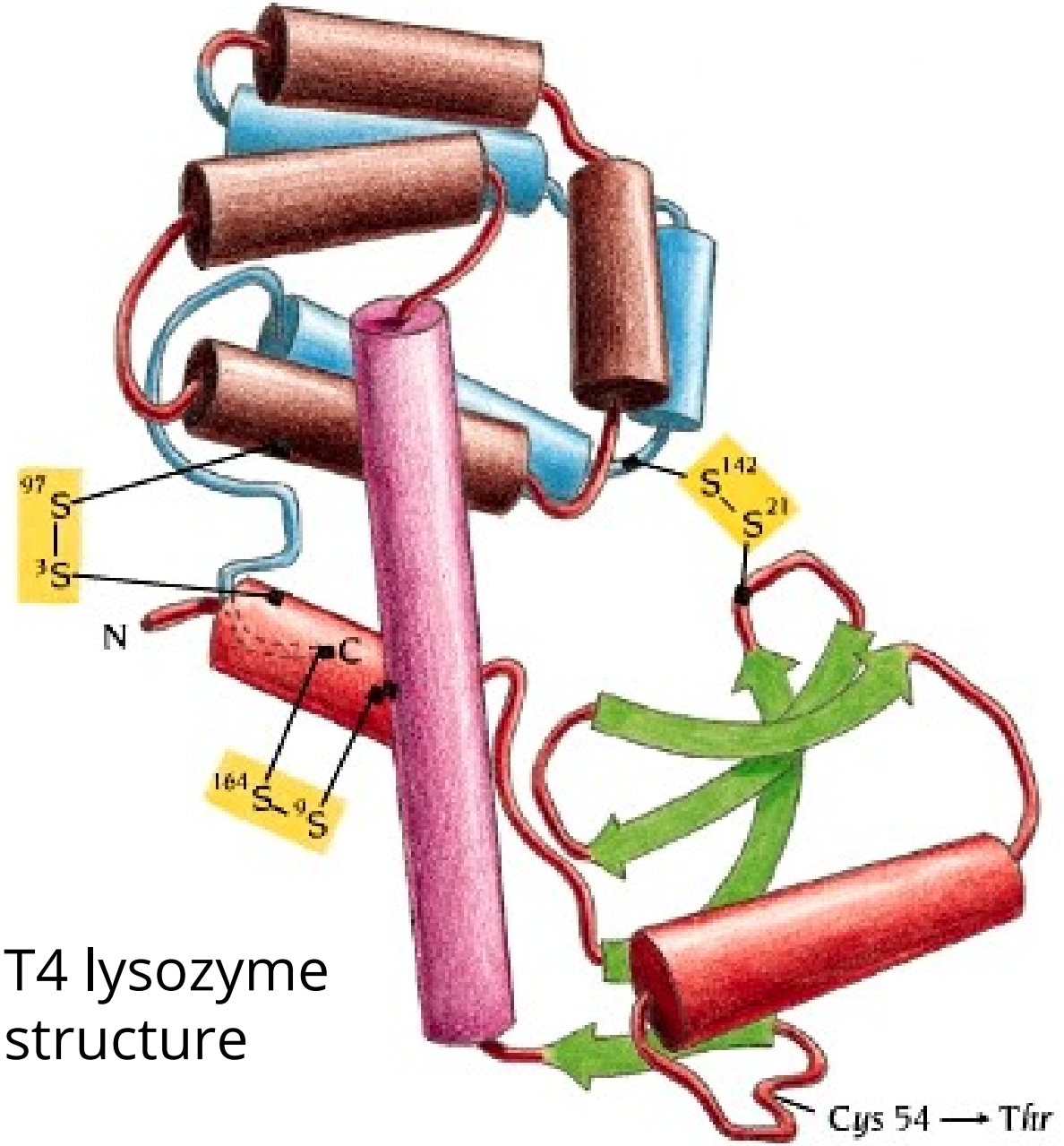




T4 lysozyme: structure known

Can it be made more stable by the addition of pairs of cysteine residues (allowing disulfide bridges to form)

Does this affect the activity of the protein?

# T4 lysozyme: a model for protein stability




T4 lysozyme structure

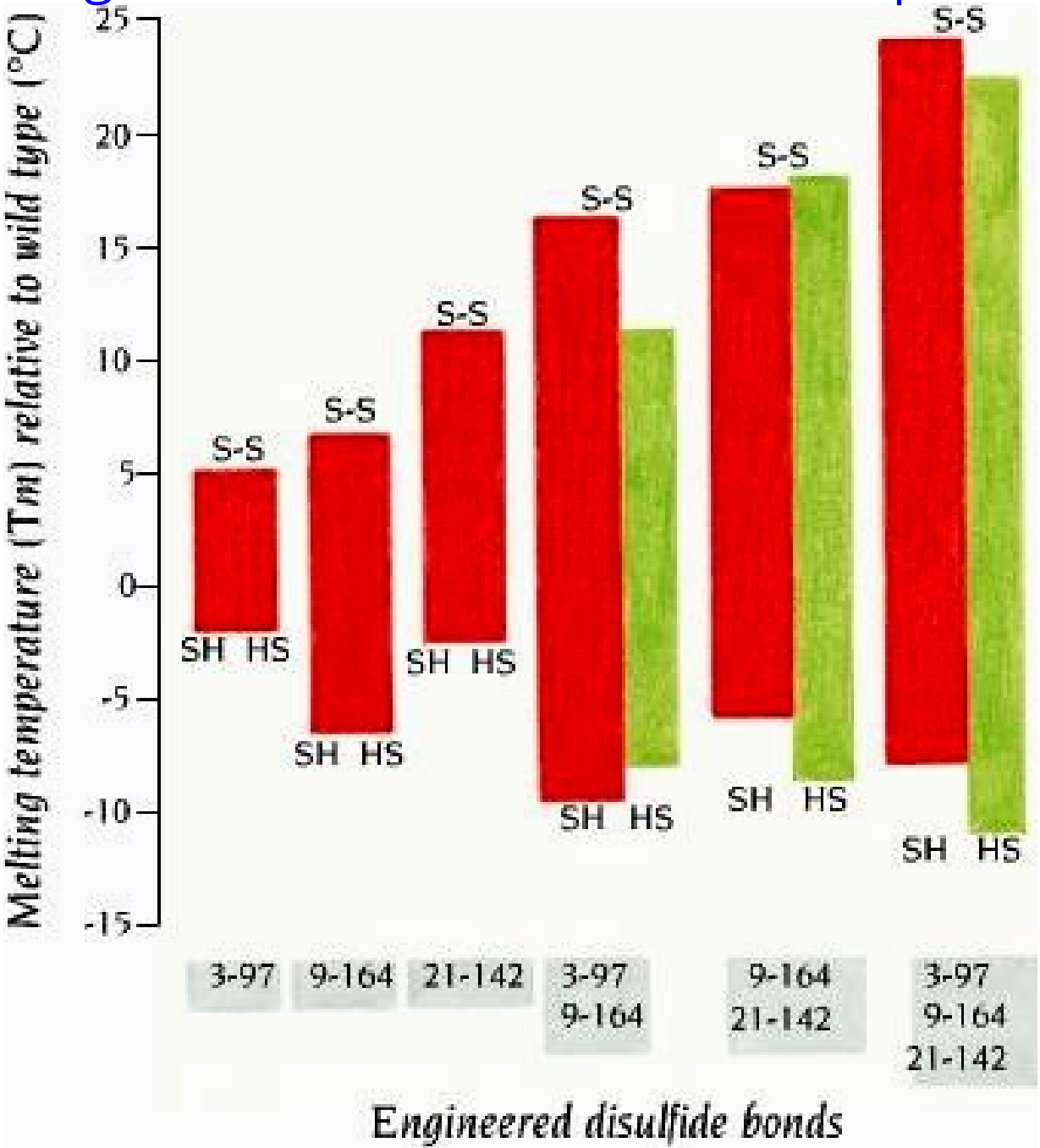
Gene was mutagenized to add cysteine residues in close proximity

Can disulfide bridges be deliberately added?

Does this enhance protein stability?

Does this affect enzyme activity?

# Engineered disulfides stabilize the protein to high temps




Bottom of bar:
melting temp in reducing conditions

Top of bar:
Melting temp in oxidizing conditions

Green bars: if the individual mutation effects were added together

# Phenotypic trade-off: Increased stability can reduce enyzme activity

| Enzyme | Amino acid at position: | | | | | | | No. of -S-S- | % Activity | $T_m$ (°C) |
|---|---|---|---|---|---|---|---|---|---|---|
| | 3 | 9 | 21 | 54 | 97 | 142 | 164 | | | |
| wt | Ile | Ile | Thr | Cys | Cys | Thr | Leu | 0 | 100 | 41.9 |
| pwt | Ile | Ile | Thr | Thr | Ala | Thr | Leu | 0 | 100 | 41.9 |
| A | Cys | Ile | Thr | Thr | Cys | Thr | Leu | 1 | 96 | 46.7 |
| B | Ile | Cys | Thr | Thr | Ala | Thr | Cys | 1 | 106 | 48.3 |
| C | Ile | Ile | Cys | Thr | Ala | Cys | Leu | 1 | 0 | 52.9 |
| D | Cys | Cys | Thr | Thr | Cys | Thr | Cys | 2 | 95 | 57.6 |
| E | Ile | Cys | Cys | Thr | Ala | Cys | Cys | 2 | 0 | 58.9 |
| F | Cys | Cys | Cys | Thr | Cys | Cys | Cys | 3 | 0 | 65.5 |

Adapted from Matsumura et al., *Nature* **342**:291–293, 1989.

wt, wild-type T4 lysozyme; pwt, pseudo-wild-type enzyme; A through F, six engineered cysteine variants; -S-S-, disulfide bonds; $T_m$, "melting" temperature (a measure of thermostability).

**F**: 3 S-S bonds, highest stability, but no activity

**D**: 2 S-S bonds, increased stability and nearly 100% active

# New amino acids in proteins?

- New amino acid = new functional group
- Alter or enhance protein function (rational design)
- Chemically modify protein following synthesis (chemical derivitization)
  - Probe protein structure, function
  - Modify protein *in vivo*, add labels and monitor protein localization, movement, dynamics in living cells

# Mutation by altering the genetic code

- 61 sense codons, 3 non-sense (stop) codons
- 20 amino acids (plus selenocysteine & pyrrolysine)

- Other amino acids exist, some in the cell (as precursors to the 20 amino acids), but have not been added to the genetic code in a living system (as far as we know)

- Can other amino acids be added to biological systems?

Xie and Schultz (2005) " A chemical toolkit for proteins -- an expanded genetic code" *Nat Rev Mol Cell Biol* 7, p. 775.

# How to modify genetic code?

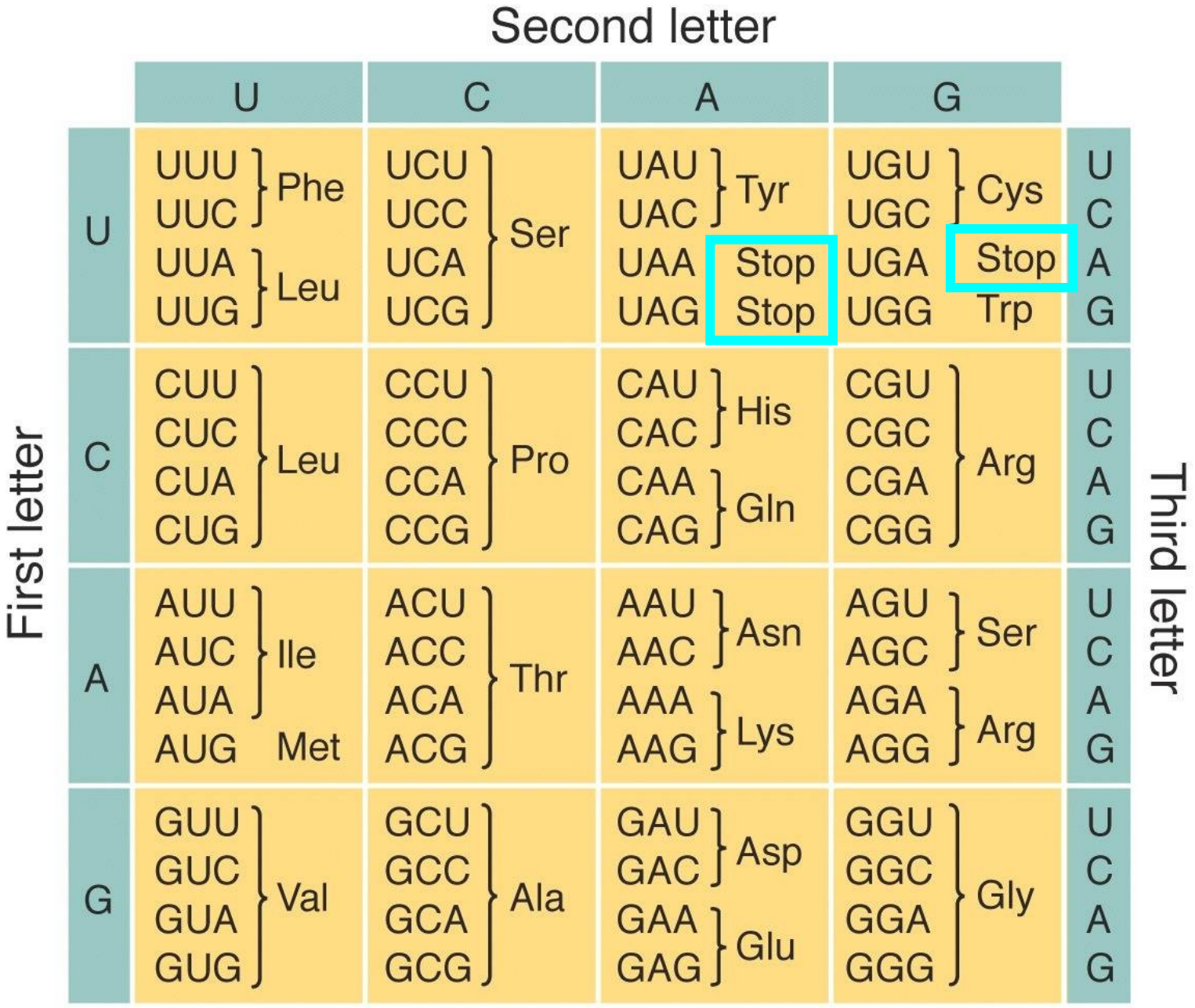
Adding new amino acids to the code--must bypass the fidelity mechanisms that have evolved to prevent this from occurring

2 key mechanisms of fidelity

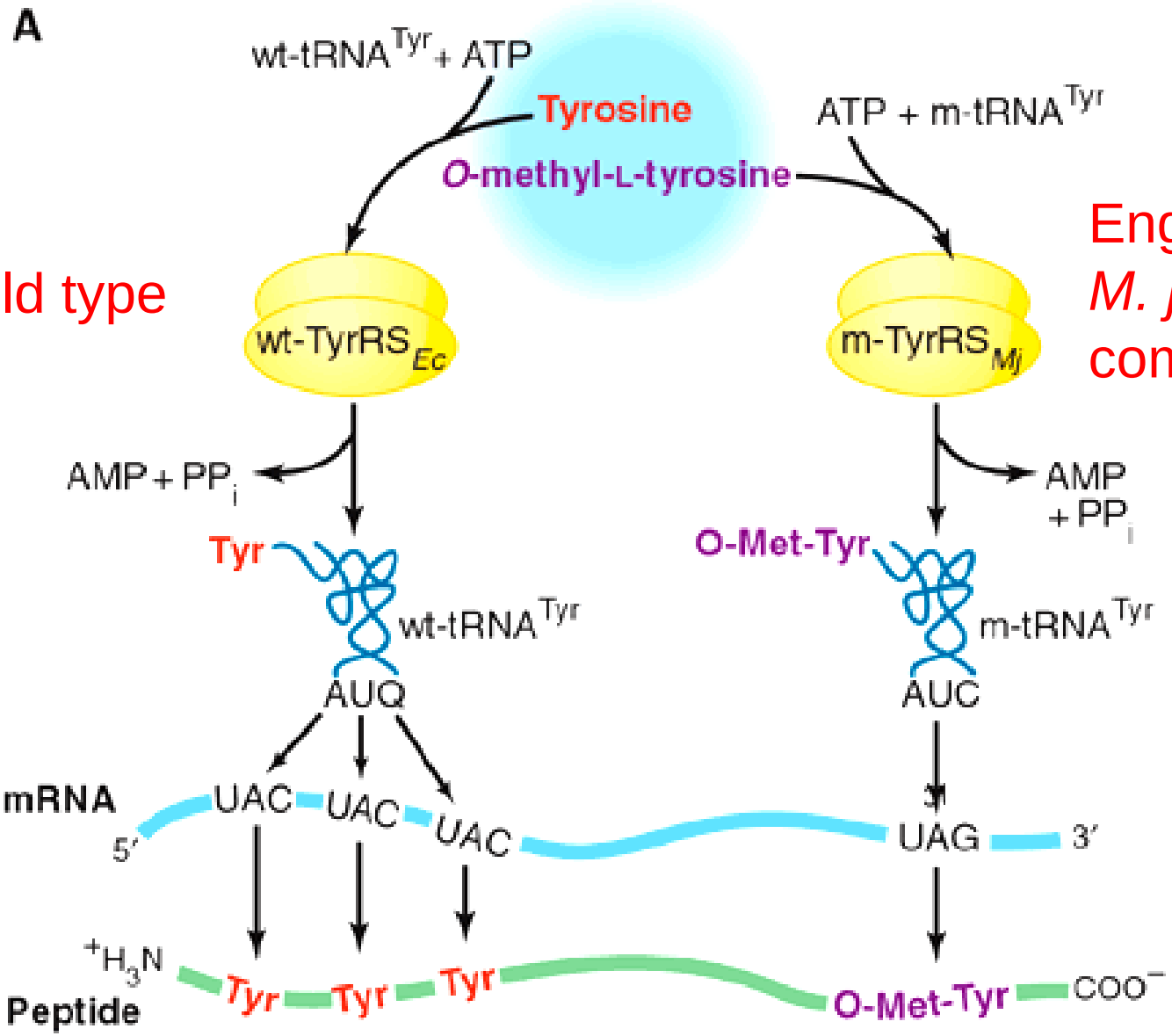
- Correct amino acid inserted by ribosome through interactions between tRNA anti-codon and mRNA codon of the mRNA in the ribosome
- Specific tRNA charged with correct amino acid because of high specificity of tRNA synthetase interaction

- Add new tRNA, add new tRNA synthetase to make a new amino acid available

# Altering the genetic code



# Site-specific insertion of unnatural amino acids




Wang et al. (2001) *Science* **292**, p. 498.
Böck (2001) *Science* **292**, p. 453.

# What can you do with a tRNA that recognizes stop codons, and that gets charged with a new amino acid?

You can program an mRNA that directs specific insertion of new amino acid

- Design protein to have UAG stop codon where you' d like the new amino acid to go
- Transform engineered *E. coli* with plasmid containing the engineered gene
- Feed cells the unnatural amino acid to get synthesis of full length gene

# Some questions:

- What are the consequences for the cell with an expanded code?

- Do new amino acids confer any kind of evolutionary advantage to organisms that have them? (assuming they get a ready supply of the new amino acid…)

- Why do cells have/need 3 stop codons????

# A system for site-specific insertion of new (unnatural) amino acids into proteins: summary

Three components were engineered:

1) new “ codon” (use amber, the rarest stop codon: UAG)
2) tRNA recognizing UAG codon, and charged with unnatural amino acid
3) aminoacyl tRNA synthetase (aaRS) to add new amino acid to the UAG tRNA

Unique proteins can be created

# What's next: Quadruplet-encoding ribosomes

Mutant ribosomes have been isolated that can decode a quadruplet base sequence (as opposed to the standard triplet)

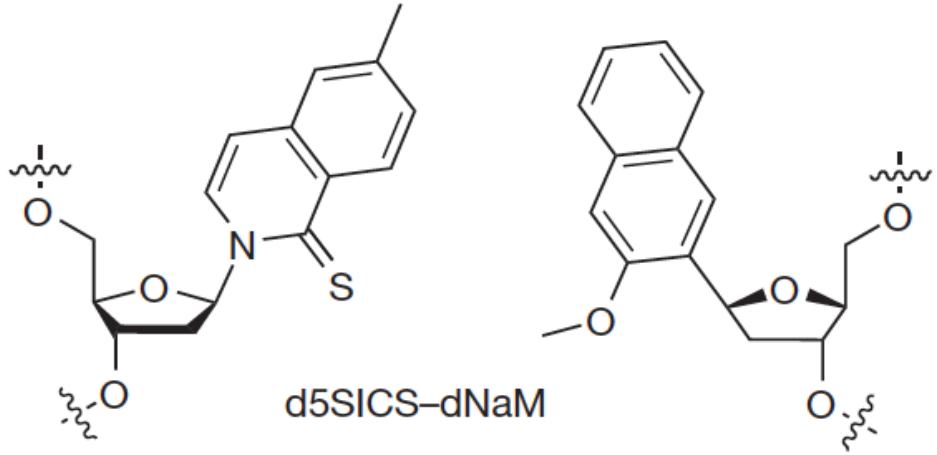
Quadruplet tRNAs charged with unnatural amino acids cause insertion of the unnatural amino acid at the quadruplet

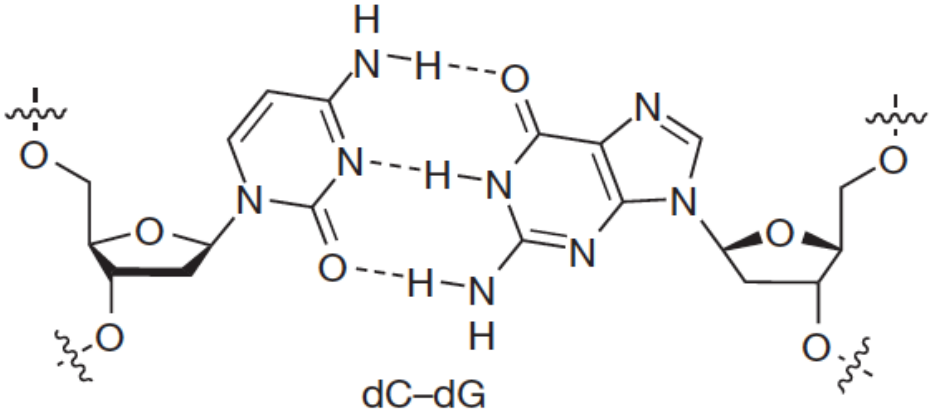
These "ribo-Q1" ribosomes are specific to only specific (orthogonal) mRNA, and won't translate "normal" mRNAs

256 blank codons available (as opposed to 3 stop codons), so it could be possible to encode many different unnatural amino acids into proteins

"...foundational technologies for the encoded synthesis and synthetic evolution of unnatural polymers in cells."

# What's next: new base pairs in DNA




Hydrophobic nucleobase: d5SICS-dNaM

Number of codons with four bases:

$4^3 = 64$

Number of codons with six bases:

$6^3 = 216$

**A semi-synthetic organism with an expanded genetic alphabet**

Denis A. Malyshev[1], Kirandeep Dhami[1], Thomas Lavergne[1], Tingjian Chen[1], Nan Dai[2], Jeremy M. Foster[2], Ivan R. Corrêa Jr[2] & Floyd E. Romesberg[1]

(see also hachimoji DNA)

# Study and engineering of gene function: mutagenesis

I. Random mutagenesis, mutant selection schemes
II. Site-directed mutagenesis, assembly of new DNA fragments
III. Rational engineering of proteins
IV. Alterations in the genetic code