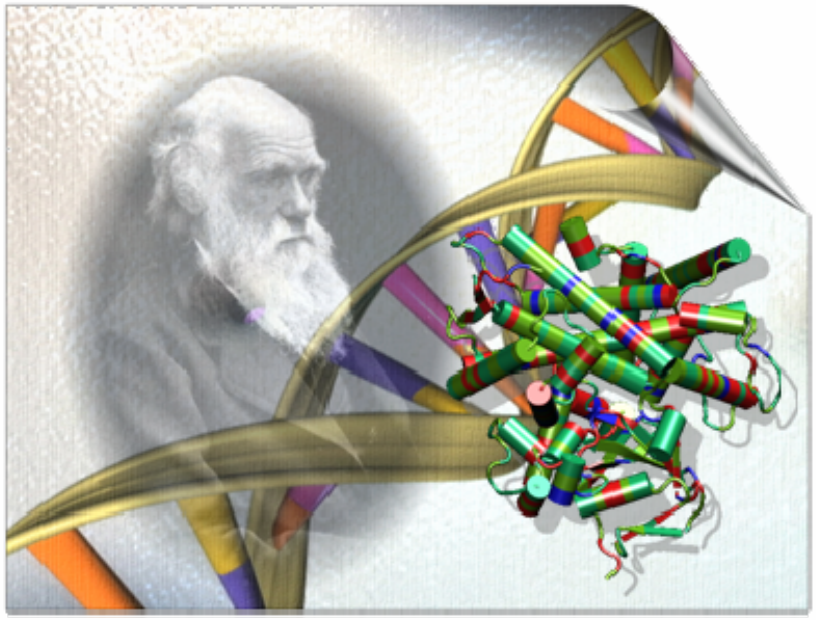
Protein engineering – Directed evolution



Roman Sobotka

EVROPSKÁ UNIE
Evropský fond pro regionální rozvoj

EUROPEAN TERRITORIAL CO-OPERATION
AUSTRIA-CZECH REPUBLIC 2007-2013
Gemeinsam mehr erreichen. Společně dosáhneme více.

Protein engineering – non-rational *versus* rational design

Protein design is the design of new protein molecules, either from scratch or by making calculated variations on a known structure.

In **rational protein design** - detailed knowledge of the structure and function of the protein is used to make desired changes

In **directed evolution** (non-rational design), random mutagenesis is applied to a protein, and a selection regime is used to pick out variants that have the desired qualities

Nobel prize in chemistry 2018

Frances H. Arnold

- *“for the directed evolution of enzymes”*

George P. Smith and **Gregory P. Winter**

- *“for the phage display of peptides and antibodies”*

Non-rational design has to be always (a bit) rational...

Library size:

20 amino-acids -> 100 residues protein -> 20^{100} (10^{130}) variants
8 position randomized to all 20 amino-acids -> 10^{10} variants

With a library size 10^{10} what is a chance to find a functional protein?

Combinatorial approaches (semi-rational design) creating gene libraries of defined composition **designed from structural and probabilistic constraints** of the encoded proteins:

- mutagenesis on of an existing protein(s), domain(s)
- polar/non-polar amino-acids pattern
- *in silico* based predictions (structure based, computer modeling) -> *hot-spot approach, grafting approach*

Directed evolution of proteins

Directed evolution does **not require prior structural knowledge** of a protein, nor it is necessary to be able to predict what effect a given mutation will have

Requires **high-throughput –** huge number of variants has to be screened

What can be altered or improved

Michaelis constant (Km), Vmax, catalytic rate constant (kcat)

Thermal tolerance, pH stability

Reactivity in different solvents

Eliminate cofactor requirement

Alter substrate binding site/specificity

Increase protease resistance

Alter allosteric regulation

Protein – protein/ligand affinity

Completely new proteins - binding agents, antigens, antibody fragments

Randomized gene library

Directed evolution of proteins relies on the construction of large libraries of variant genes (10^6 – 10^{14})

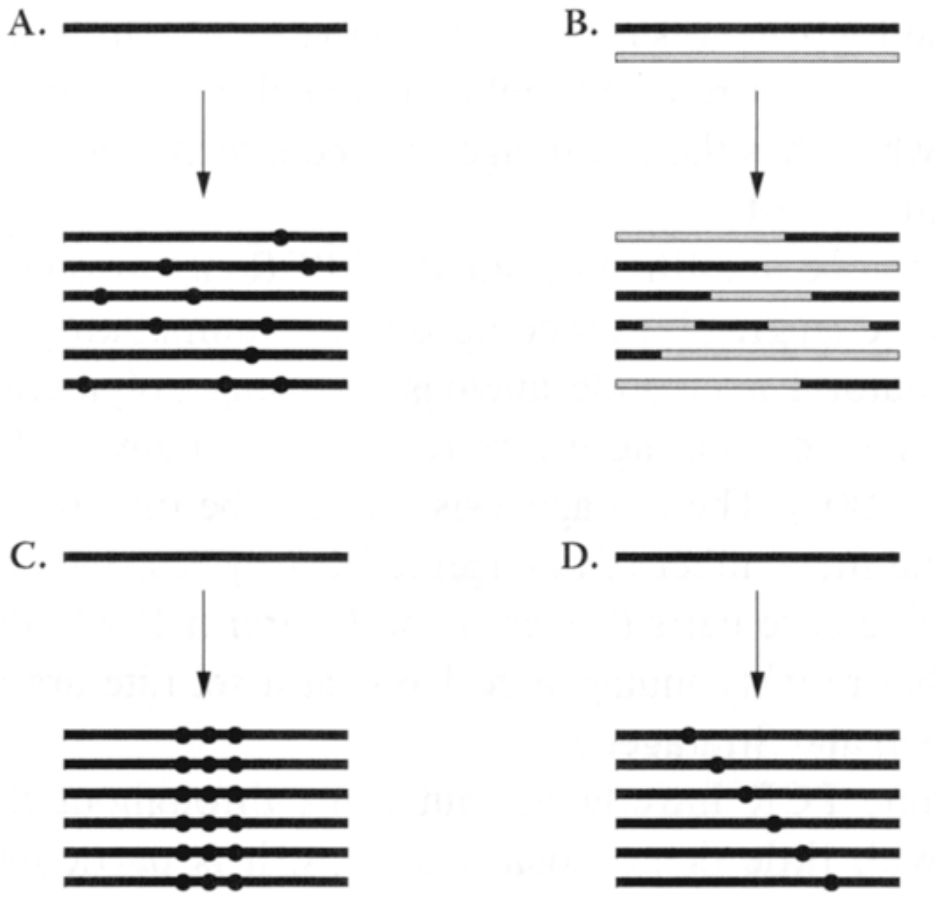
Typically based on the sequence of one or a small number of starting proteins (genes) that already have properties similar of those required

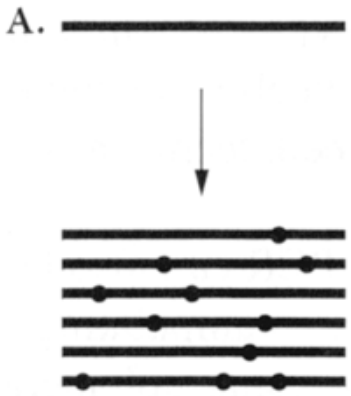
-> a library is designed and constructed with two (conflicting) goals:

- members need to be sufficiently similar to share a similar function (structure)
- members need to be sufficiently different to gain new functional property

Strategies of library construction

- Random mutagenesis (A)
- Recombination (B)
- Site-directed diversification (C)
- Scanning mutagenesis (D)





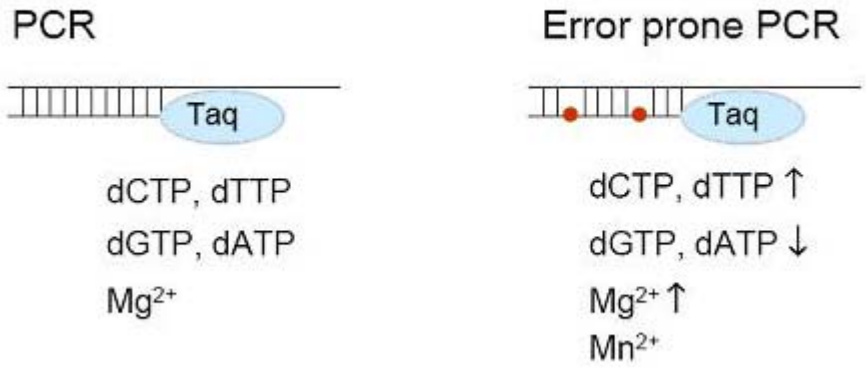
Random mutagenesis

Error-prone PCR

Mutator *E. coli* strains - deficiency in DNA repair pathways

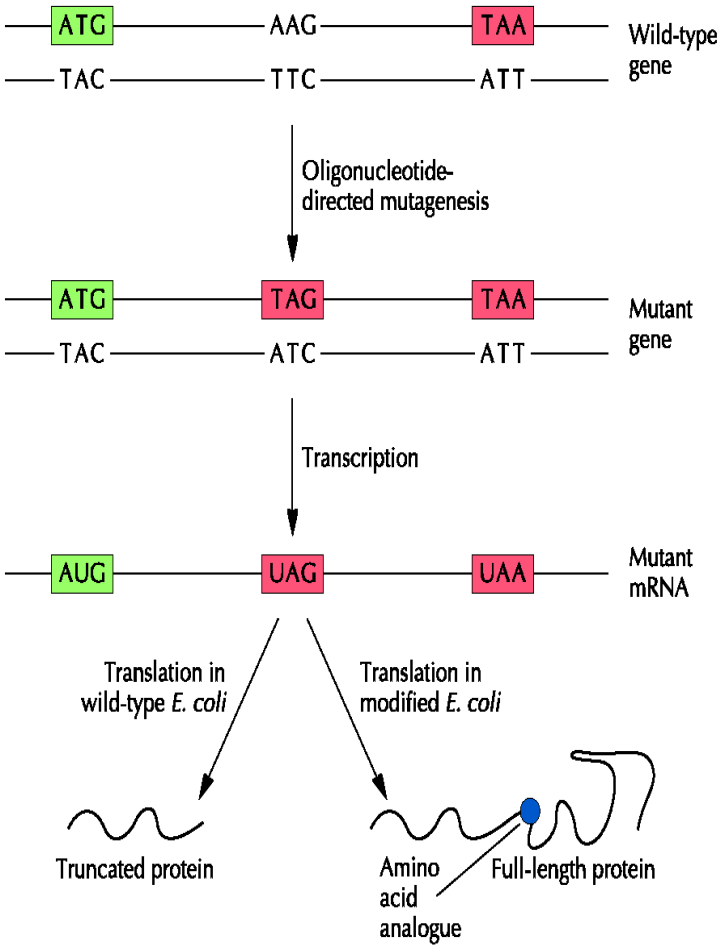
Insertion mutagenesis - transposon-based system to randomly insert a short sequence throughout a sequence of interest

Chemical mutagens - chemically modify DNA



Commercial kits (GeneMorph II from Stratagene…)

Insertion of novel amino acid analogues



Altered targeted codon to TAG (UAG)

Transformed into strain with modified tRNA and aminoacyl tRNA synthetase

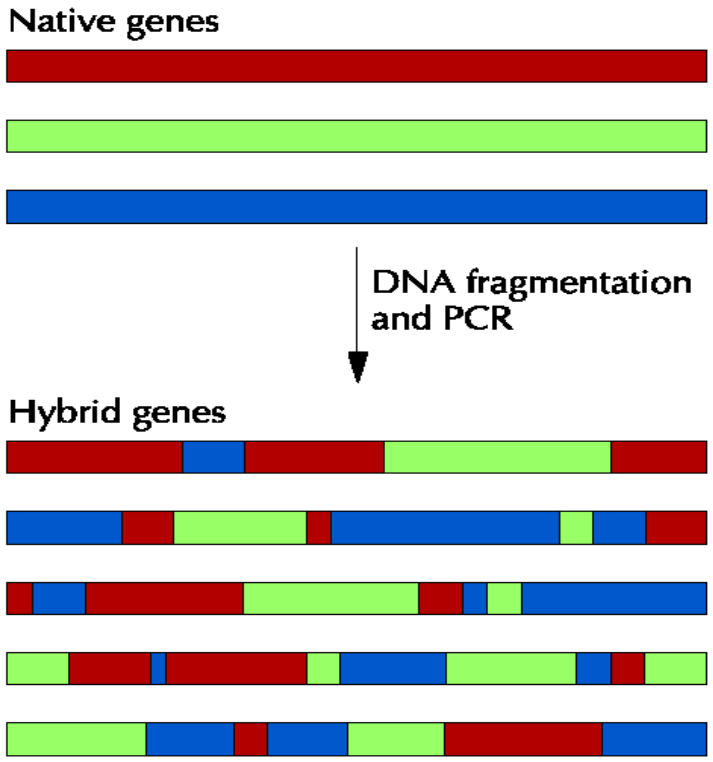
PCR-Based DNA Shuffling

In vitro homologous recombination of pools of selected mutant genes by random fragmentation and PCR reassembly

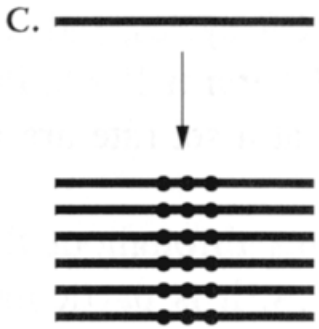
Mix sets of cut fragments (DNAse I)

During PCR fragments cross-prime each other

Final product obtained by including terminal primers ->full length fragments



Site-directed diversification

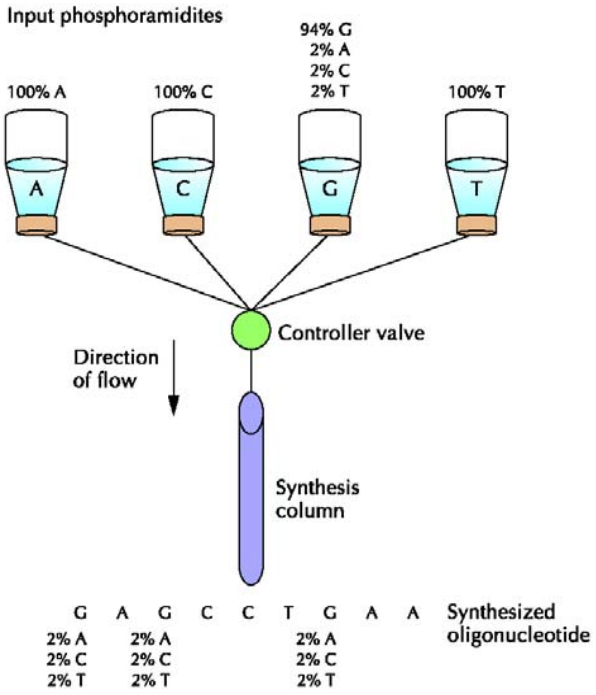


Directed to a specific position or set of positions in a gene

Based of using of **degenerated (random) PCR primers**:

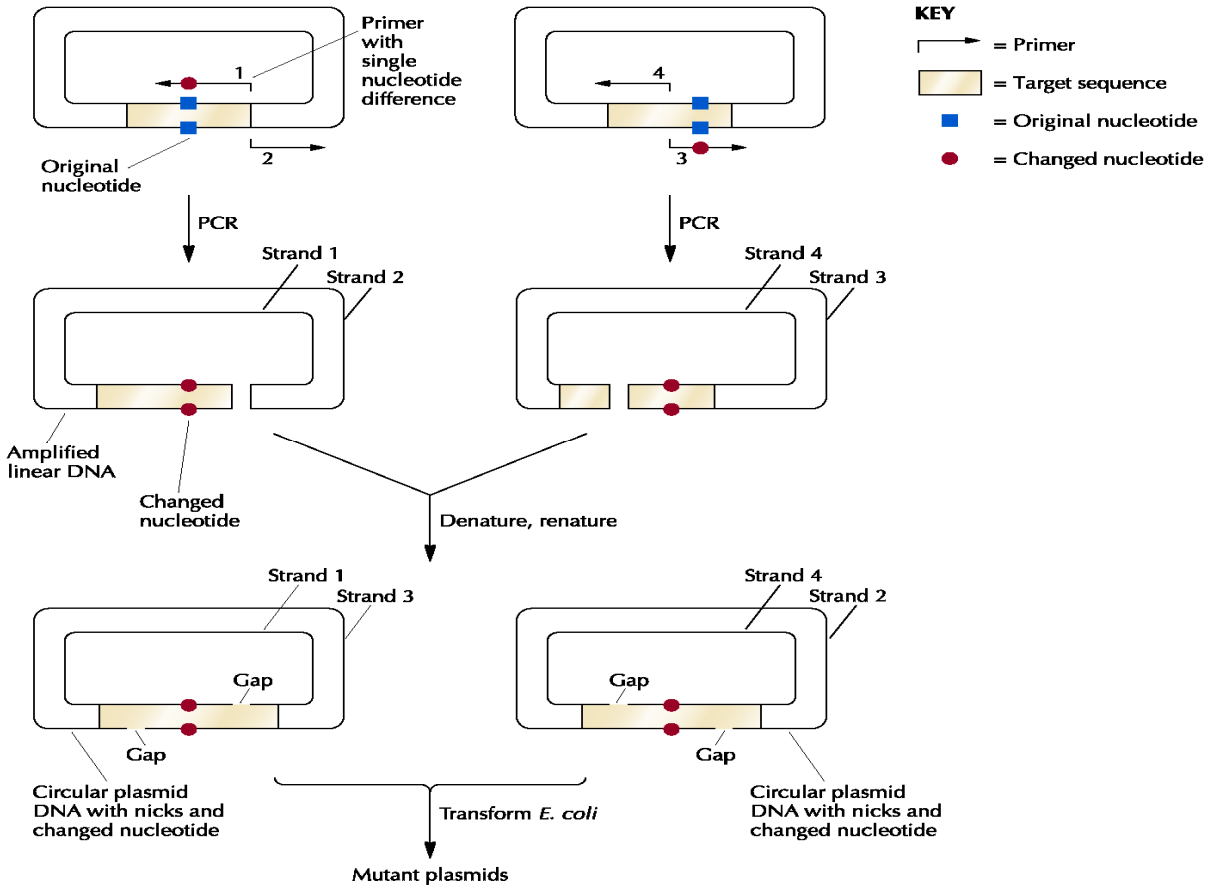
-> each codon of interest replaced by mixture of nucleotides described as NNN, NNS and NNK

N = A,C,T,G
S = G or C
K = G or T



Site-directed diversification

- Even a few number of sites mutated easily overwhelms the capacity of display/screening methodology (6 sites = 32^6 = ~ 10^9)

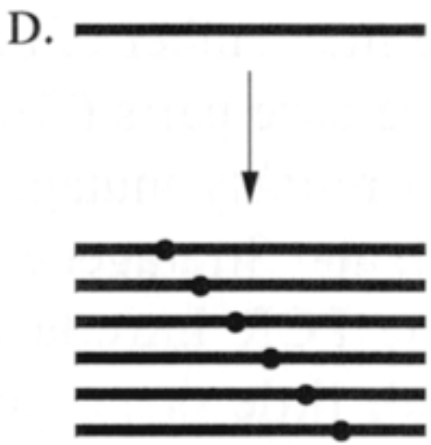


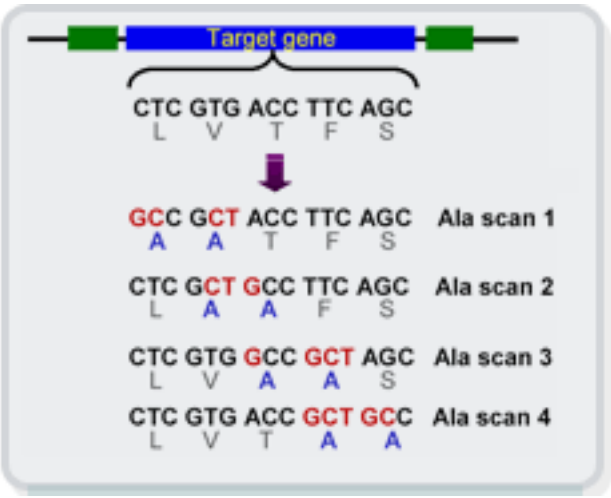
Quikchange system (Stratagene)

Scanning mutagenesis

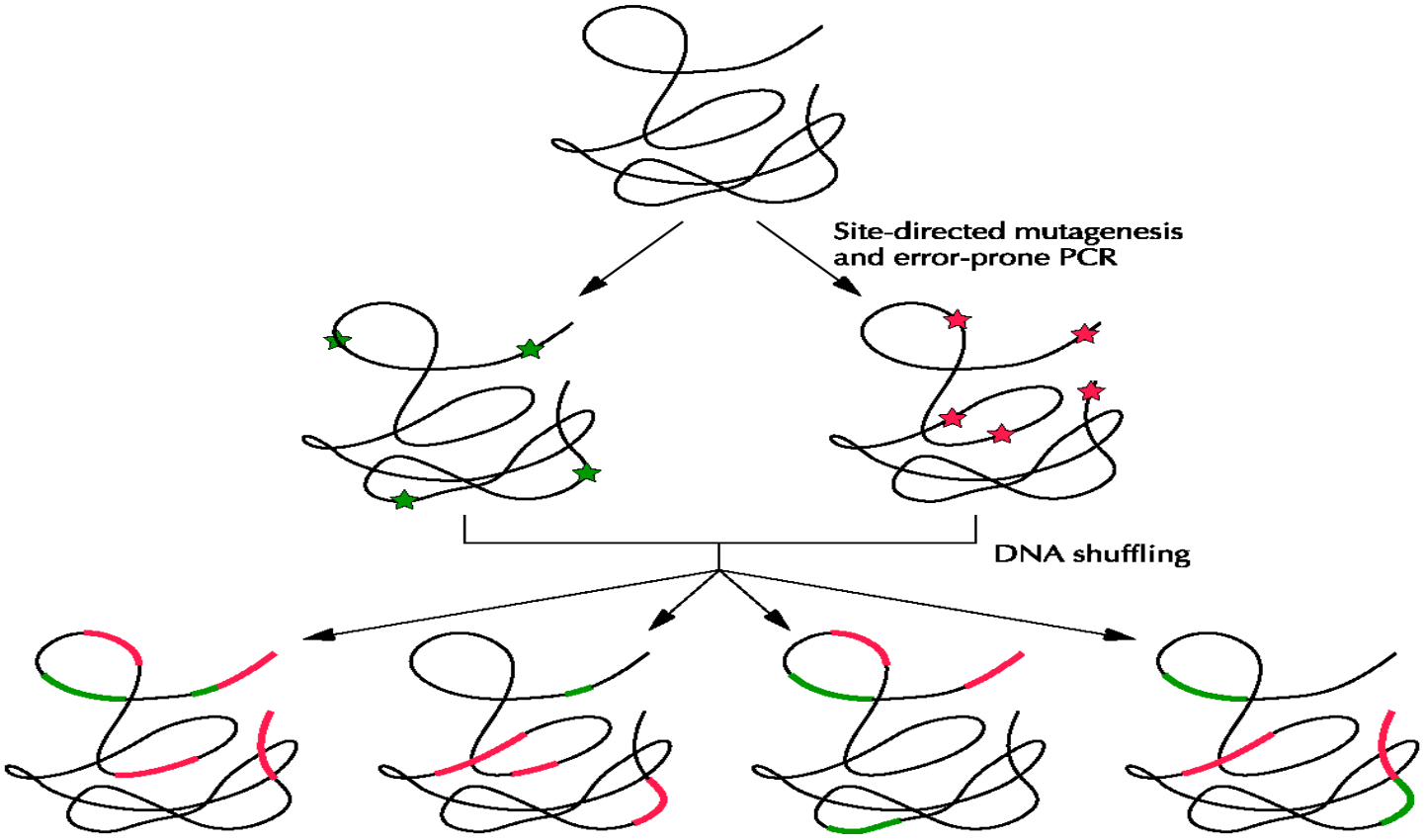
Alanin scanning mutants is frequently used to define functionally important residues or domains of a protein

Cystein, proline ...





Combination of more methods (error prone PCR + shuffling)



Protein display techniques

Protein expression in microtiter plates (~10^4 variants)

Phage display (up to 10^{11})

Cell surface display; bacteria, yeast, human cells (10^6 - 10^9)

Cell free (ribosome, mRNA) display (up to 10^{14})

Phage Display

An *in vitro* selection technique using a peptide or protein genetically fused to the coat protein of a bacteriophage

It is possible to create phage populations of different representativity, wherein each individual phage displays a random peptide on its surface

Polypeptides of interest can be screened using selection techniques

With the advance of DNA sequence recognition selected sequences can be easily identified

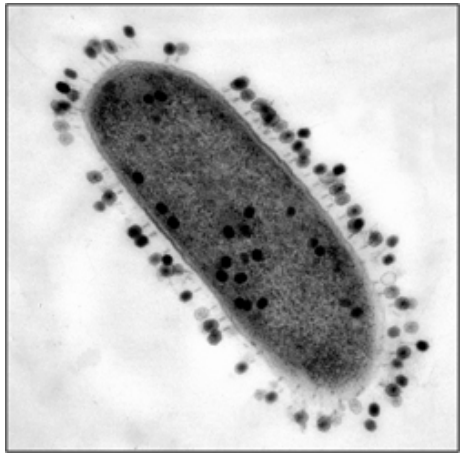
Bacteriophages

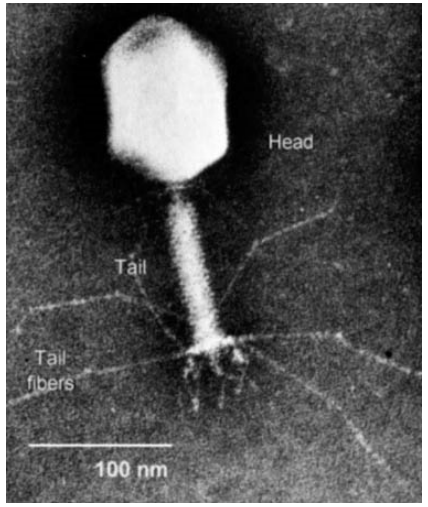
Bacteriophages are viruses that infect bacterial cells

Infected cells are used as hosts to replicate the virus

E. coli phages used due to ease of culture and quick regeneration

3 types of common phages used in phage display are the **M13**, F1 , FD





M13 bacteriophage

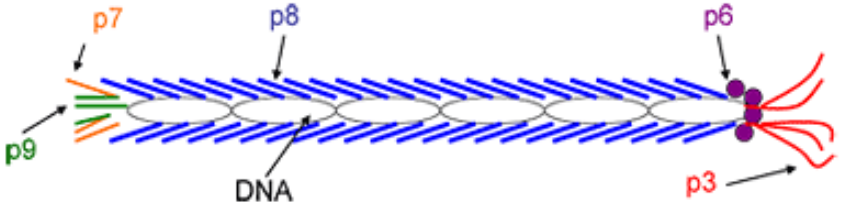
Filamentous bacteriophage composed of circular single stranded DNA, 6407 nucleotides

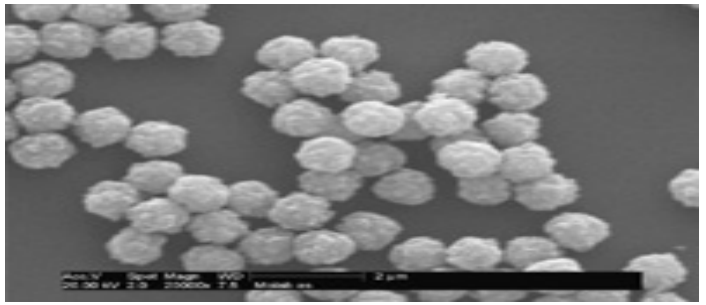
It is encapsulated in 2700 copies of the major coat protein P8, capped with 5 copies of two different minor coat proteins (P9, P6, and P3)

The minor P3 attaches to the receptor at the tip of F pilus of the host *E. coli*

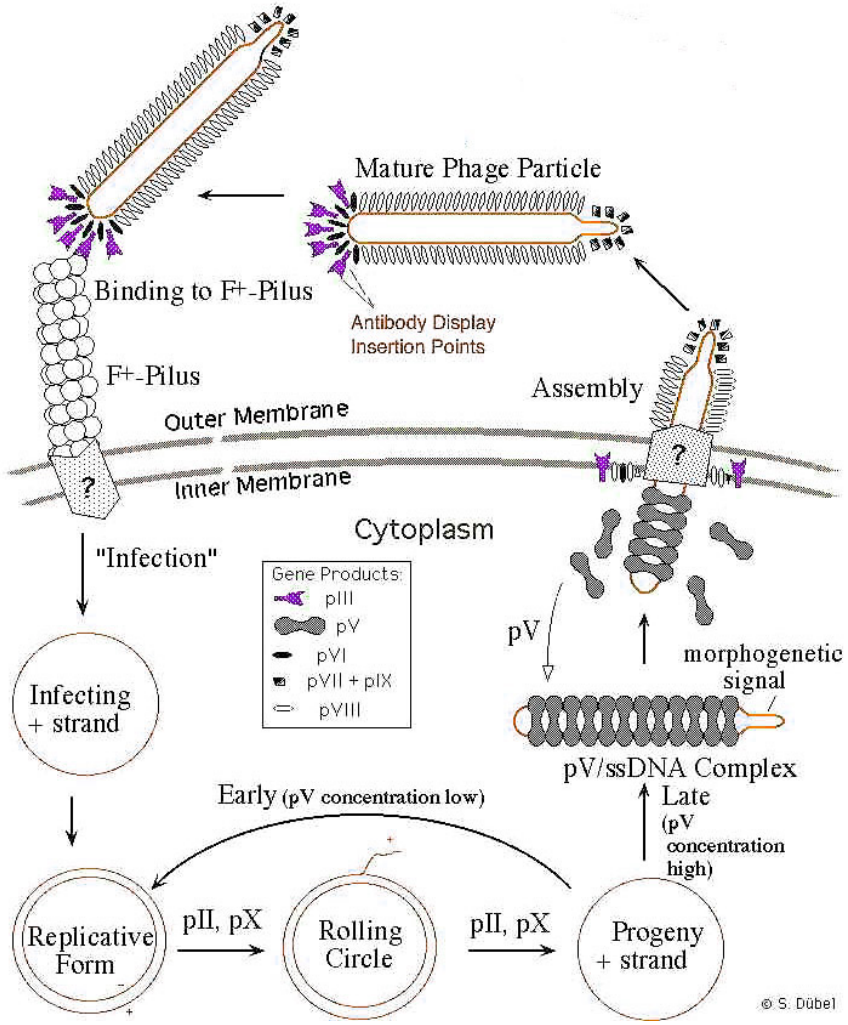
It is not lethal but causes plaques in the bacterial cell (areas of diminished cell growth), a non-lytic virus

The M13 phage is used for many recombinant DNA processes due to its extreme size and the virus has also been studied for its uses in nanostructures and nanotechnology

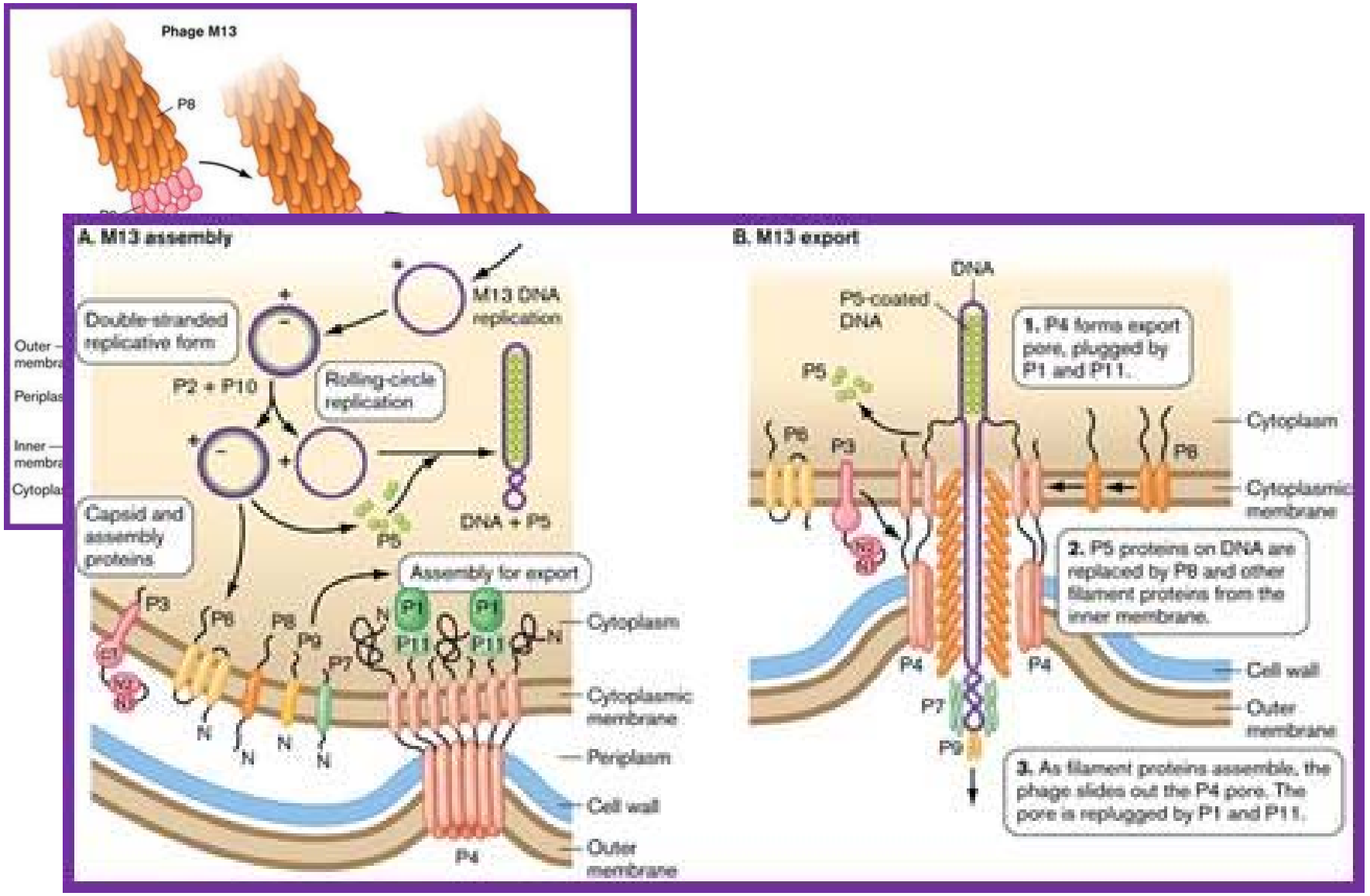




M13 bacteriophage life cycle

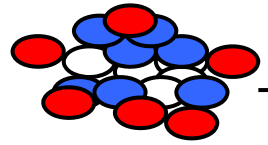


M13 bacteriophage life cycle

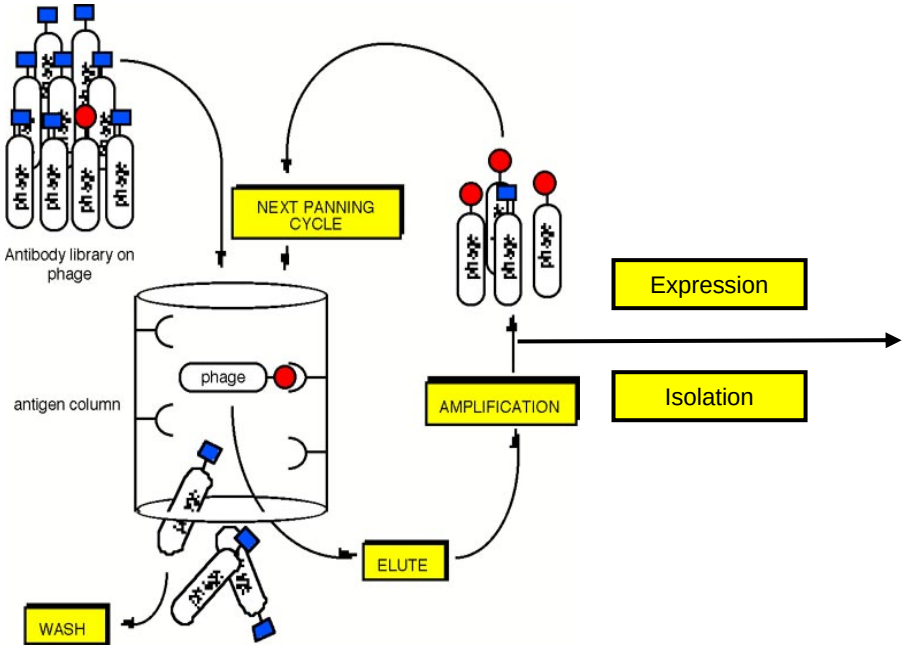


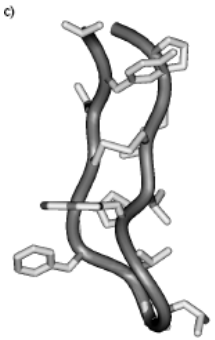
Steps involved in using phage display

Creation of vector
Binding/Selection
Wash
Elution
Amplification

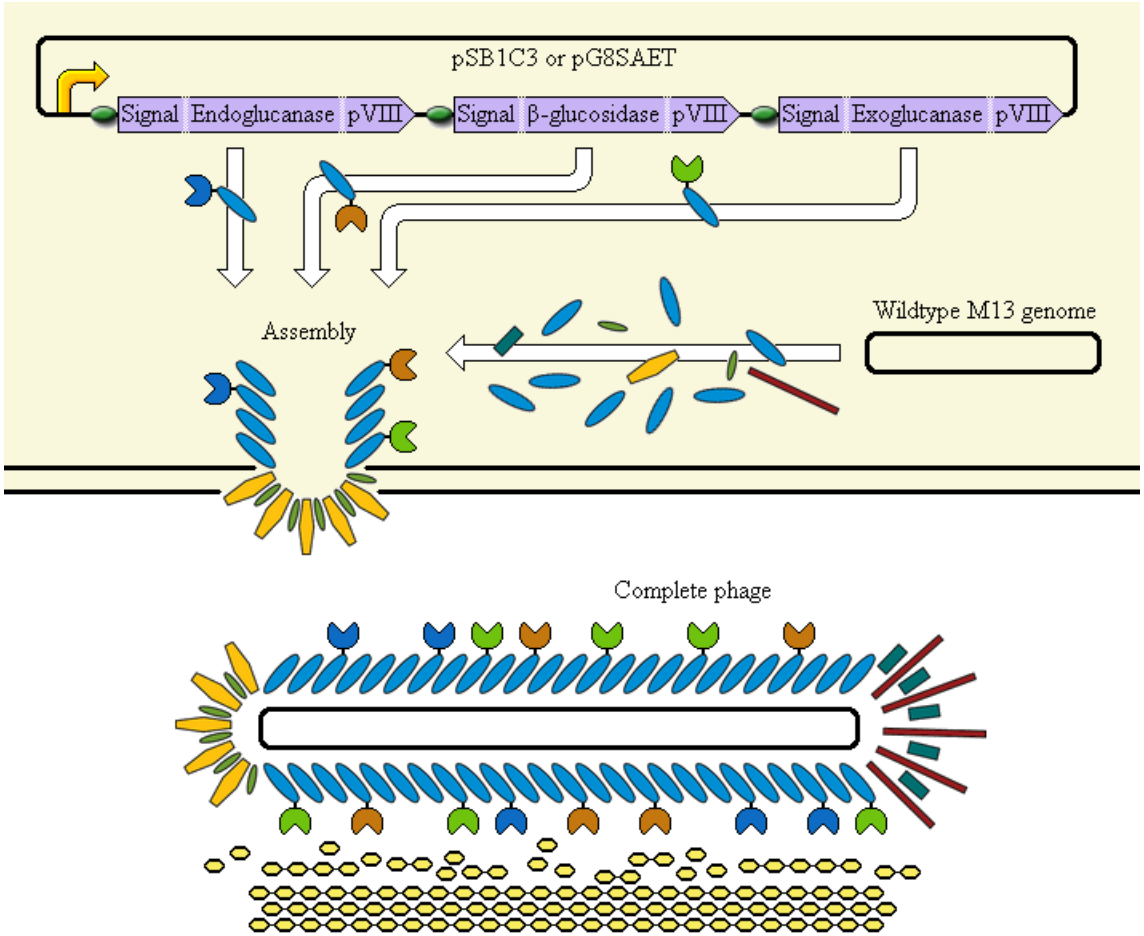


Library clones





Construction of viral vectors



Modified plasmid used to incorporate foreign DNA of interest into viral DNA.

Spliced into gene for a coat protein so the protein will be displayed on outside of phage particles

Construction of viral vectors

Incorporation of a gene (VH/K) protein into the phage p8 coat proteins

ompA leader sequence used to target fused protein into membrane

Outer membrane protein A (OmpA) -> a native *E. coli* protein

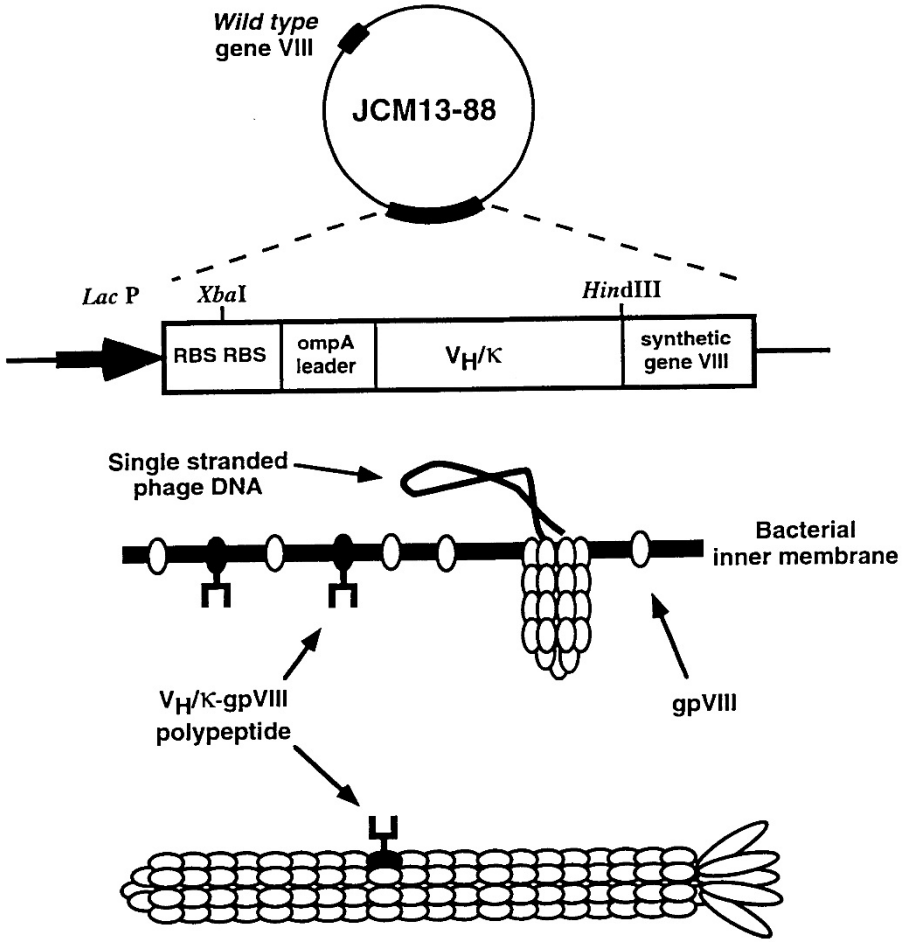
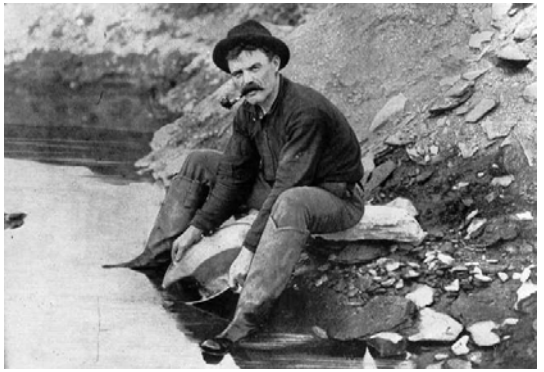


FIG. 1

Binding and selection - biopanning

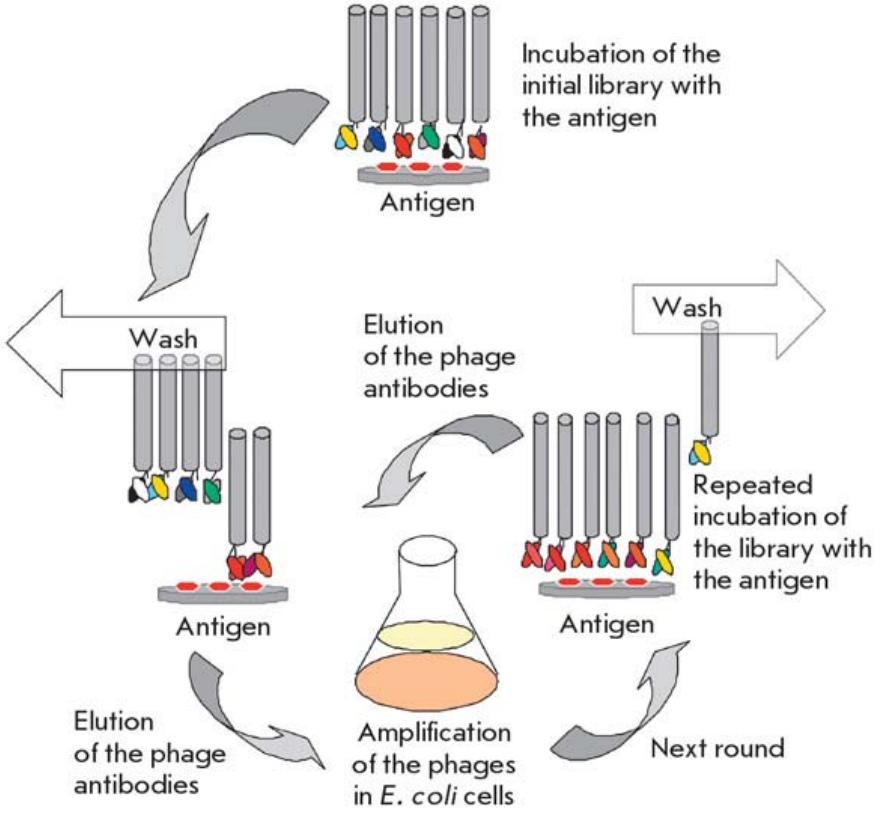


Can apply standard affinity techniques to capture phage by taking advantage of displayed proteins.

Pass solutions of amplified phages over a resin with antigens or receptors bound to it.

Phages with affinity bind to resin

Usually a few rounds of selection

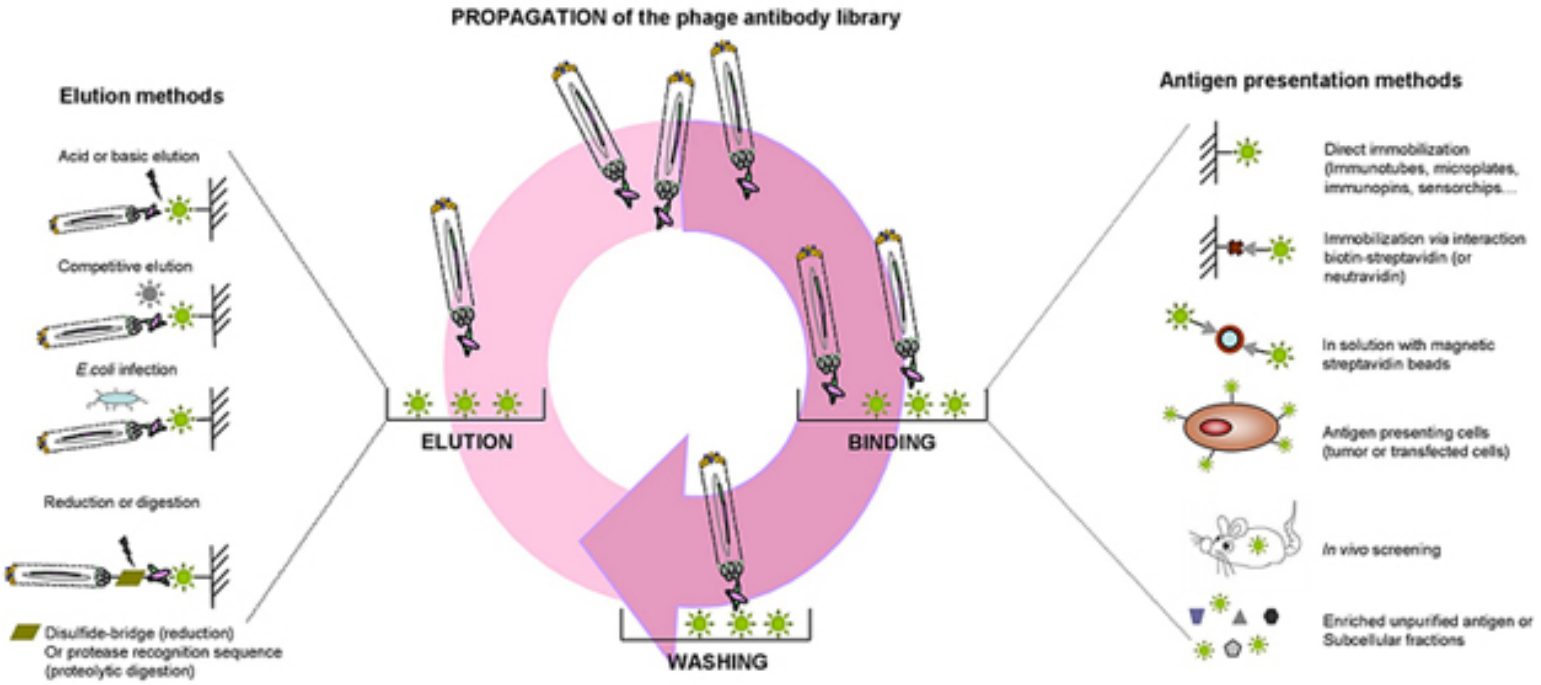


Elution and amplification

Eluted phages showing specificity are used to infect new host cells for amplification.

Cycle repeated 2-3 times for stepwise selection of best binding sequence

Final phages can be propagated then characterized with DNA sequencing



Application of phage display

Epitope mapping and mimicking (epitope = antigenic determinant)

Epitope discovery –> new vaccines

Identification of new receptors & ligands, **drug discovery**

Organ/tissue targeting

Creation of antibody libraries

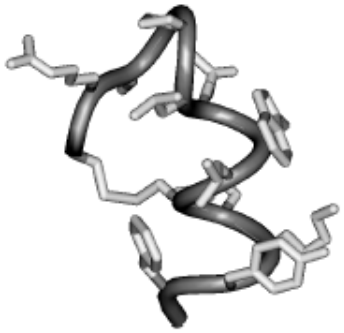
Material engineering

Drug discovery

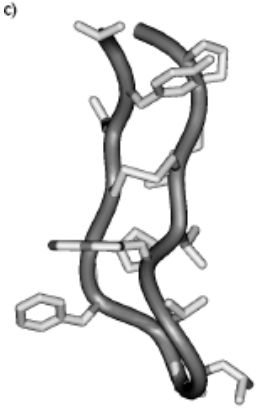
Library screened against immobilized targets (receptors …)

Peptides can act as antagonists, agonists, or modulators

Many candidates often required -> might not have good pharmacological properties



A turn-helix conformation is adopted by a peptide that binds to the insulin-like growth factor binding protein 1 (*Biochemisty,* 1998)

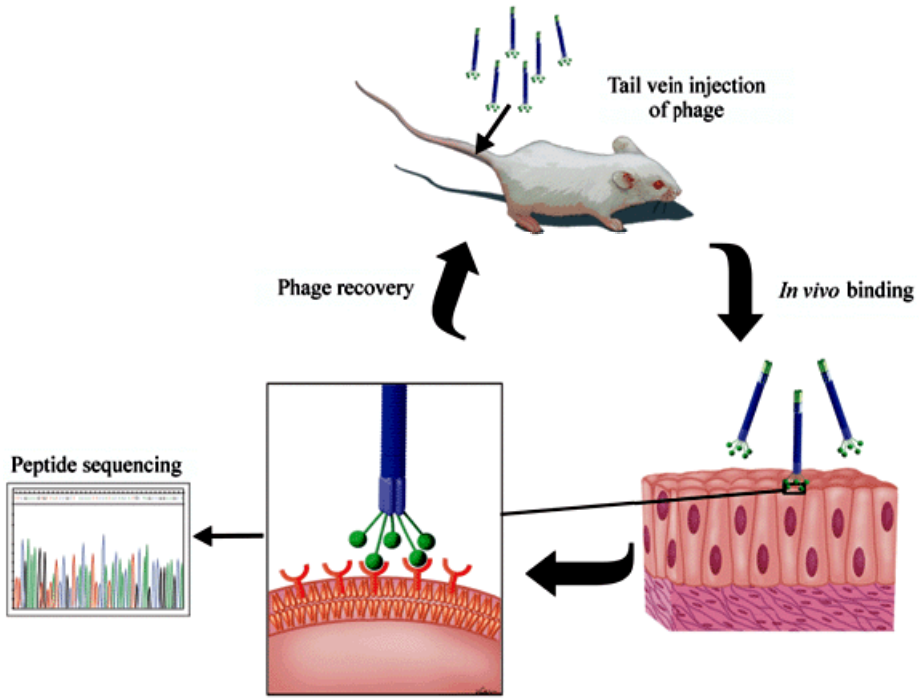


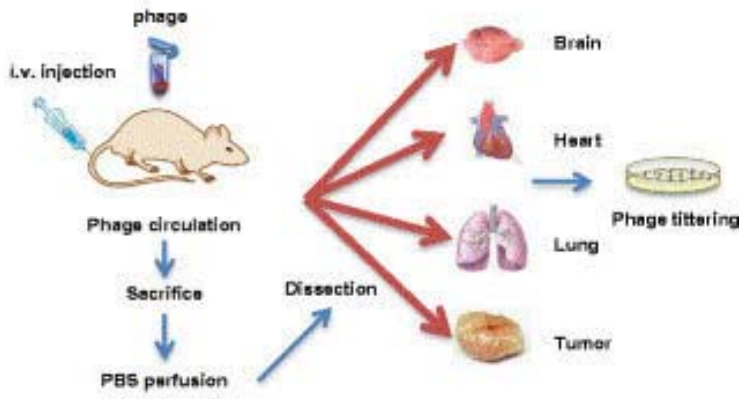
A β-hairpin peptide that binds to the Fc fragment of human IgG (*Science,* 2000)

Organ targeting

Identification of selective markers that target cells to help get drugs to selected tissue.

Inject phage into mouse then extract phages from different organs.
Identify common motifs possibly involved with localization





phage display – advantages

Easy to screen large # of clones $>10^{10}$

Easy to amplify selected phages in *E. coli*

Selection process easy and already in use in various forms

Phage library variation can be easily constructed by inducing mutations, using error prone PCR, etc.

– disadvantages

Might not have **long enough peptide** insert, so critical folding can be disrupted

Could lose phage variations if first bind/wash step is too stringent.

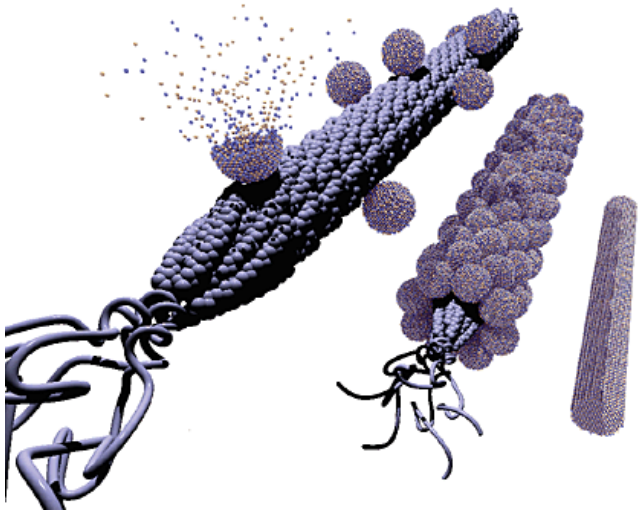
Affinities or binding that results during selection might not work *in vivo*

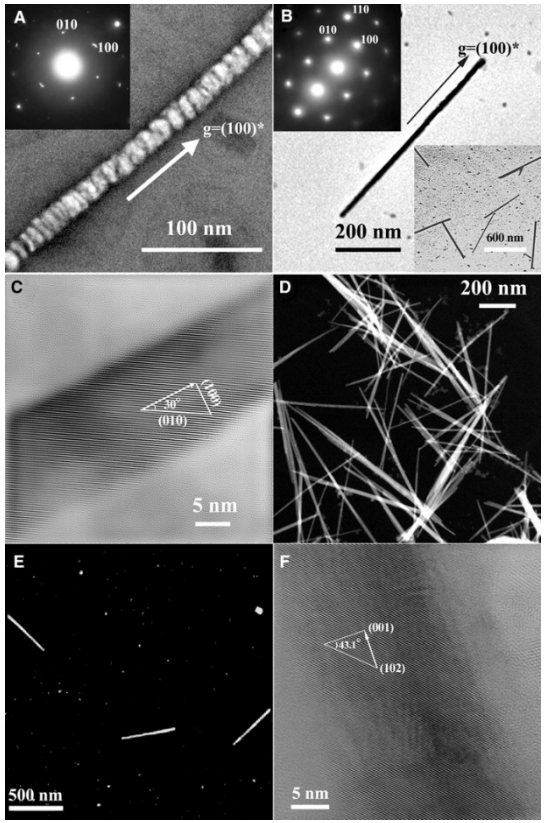
No eukaryotic post-translation modifications

Material engineering on phage surface

Modified M13 coat proteins allow precise assembly of nanowires using phage surface

Bacteriophage-based electrodes



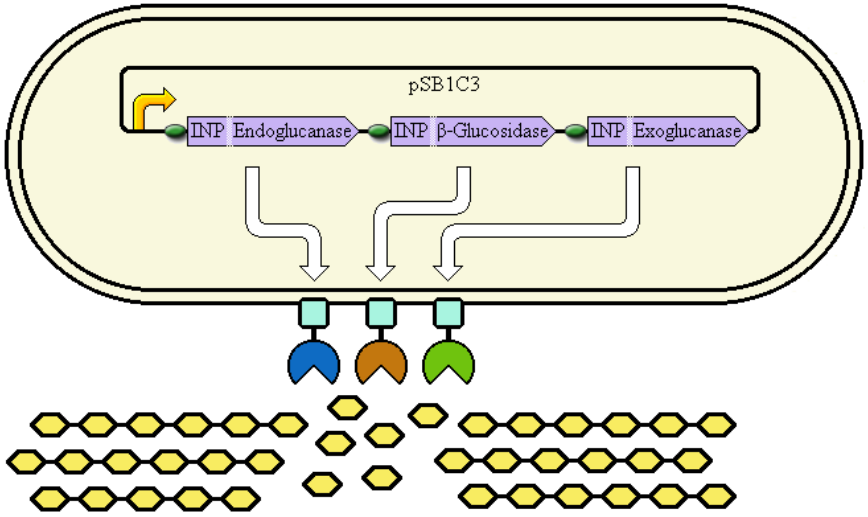


Cell surface display

Protein anchored on the surface of host (bacteria, yeast, human)

Express larger proteins with (eukaryotic) post-translation modifications

Screen the positive clones is possible by **Fluorescence Activated Cell Sorting** (FACS)



Bacteria cell surface display

Gram-negative bacteria is commonly used (*E. coli*)

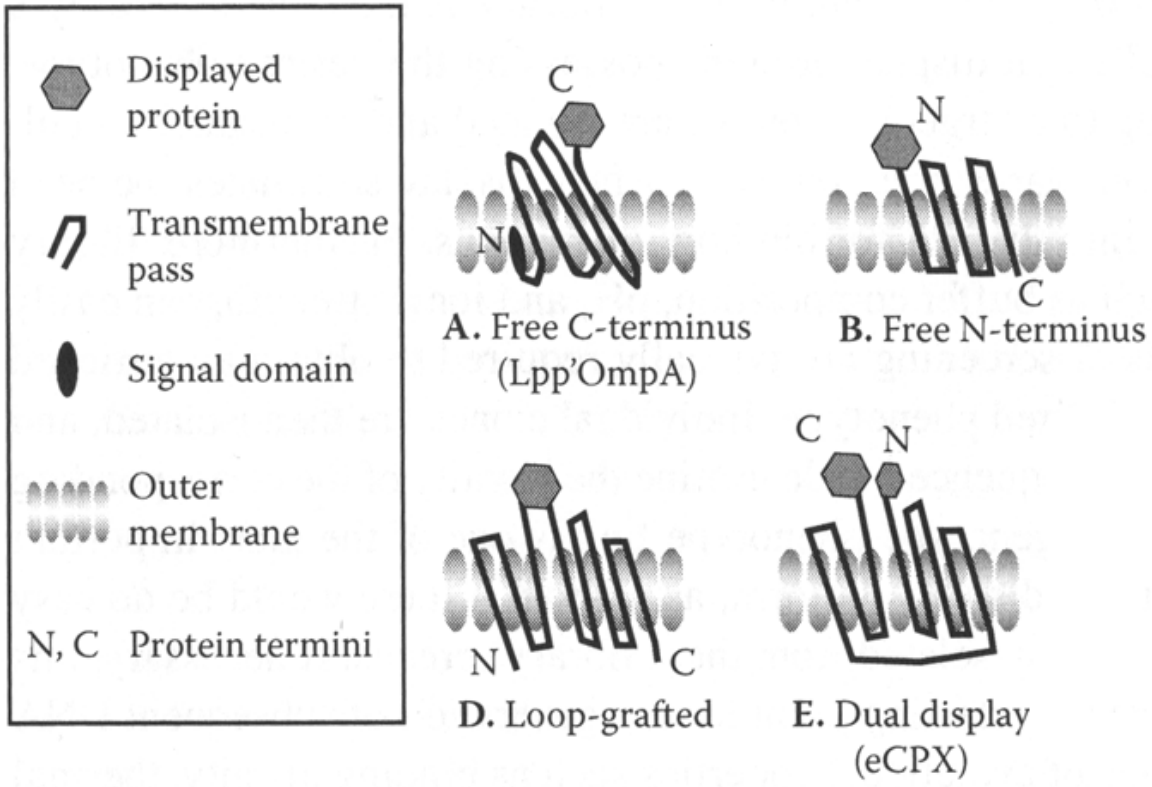
Genes from a library fused with a gene coding for outer membrane protein. Normally translated, transported to membrane, anchored on the outer membrane and face to extracellularly

First successful applications:

cellulases, esterase, organophosphate hydrolases, scFv enzymes (**scFv = antibody single-chain variable fragment**), β-lactamase, thioredoxin

Bacteria cell surface display

Membrane protein of *E. coli* used for surface display: LamB, OmpA, OmpC, OmpS, OmpT, FhuA, the lipoprotein TraT ...



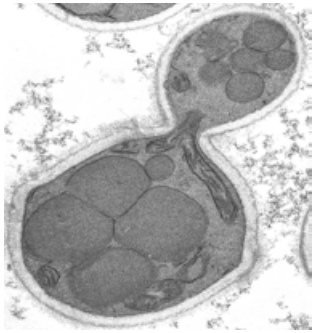
Yeast Surface Display

Similar protein secretion, post-translational modification with eukaryotes

Suitable for express antibody fragments, cytokines and receptor extra cellular domains

Saccharomyces cerevisiae
Pichia pastoris
Hansenula polymorpha (syn. Pichia angusta) -> excellent for the production of therapeutical and recombinant proteins on a industrial scale

Larger proteins can be displayed than in phage/bacterial systems

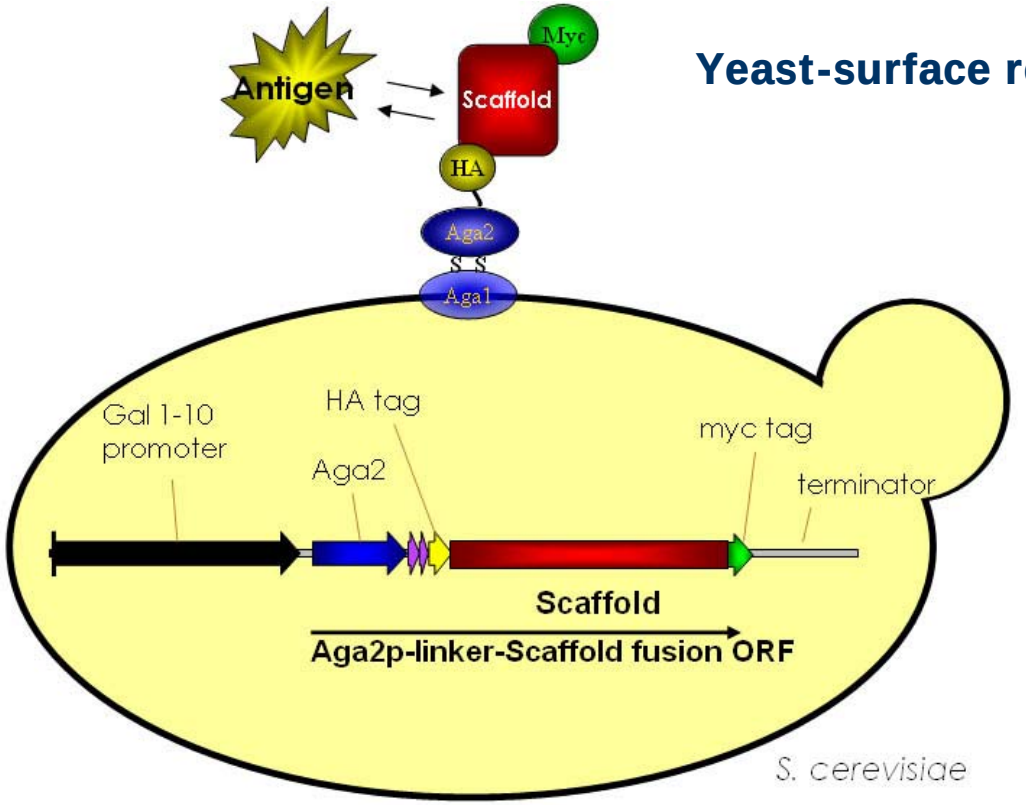


Eukaryotic post-translation modifications

Limited size of the library (typically 10^6-10^7 clones)

Yeast Surface Display

A protein of interest is displayed as a fusion to the Aga2p protein (a-agglutinin) on the surface of yeast. Agglutinin is naturally used by yeast to mediate cell–cell contacts during yeast cell mating.

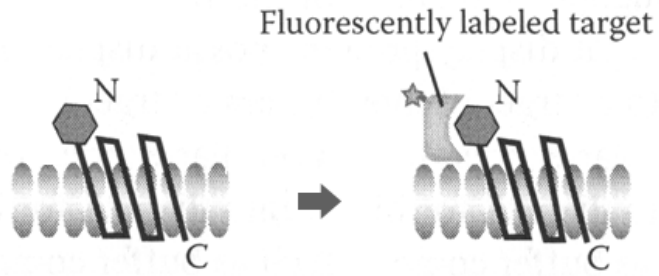


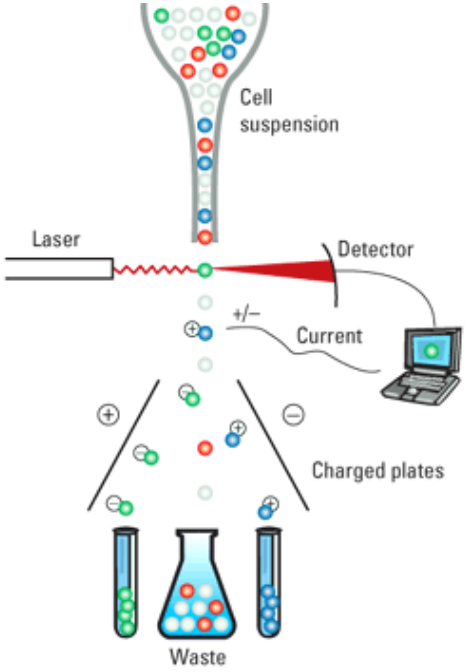
Yeast-surface receptor a-agglutinin

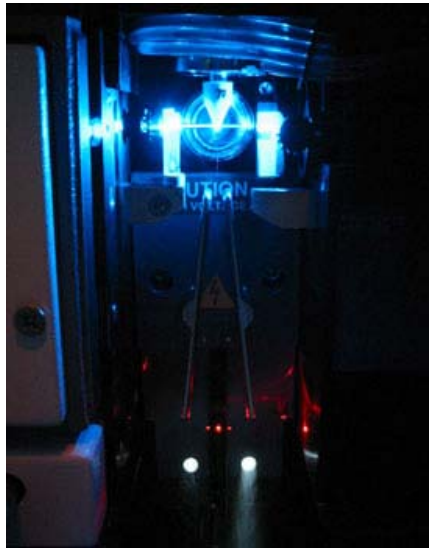
Fluorescence Activated Cell Sorting (FACS)

- Used to isolate distinct (‘fluorescent’) population from mixtures of cells
- Phages are to small to produce light scattering ..

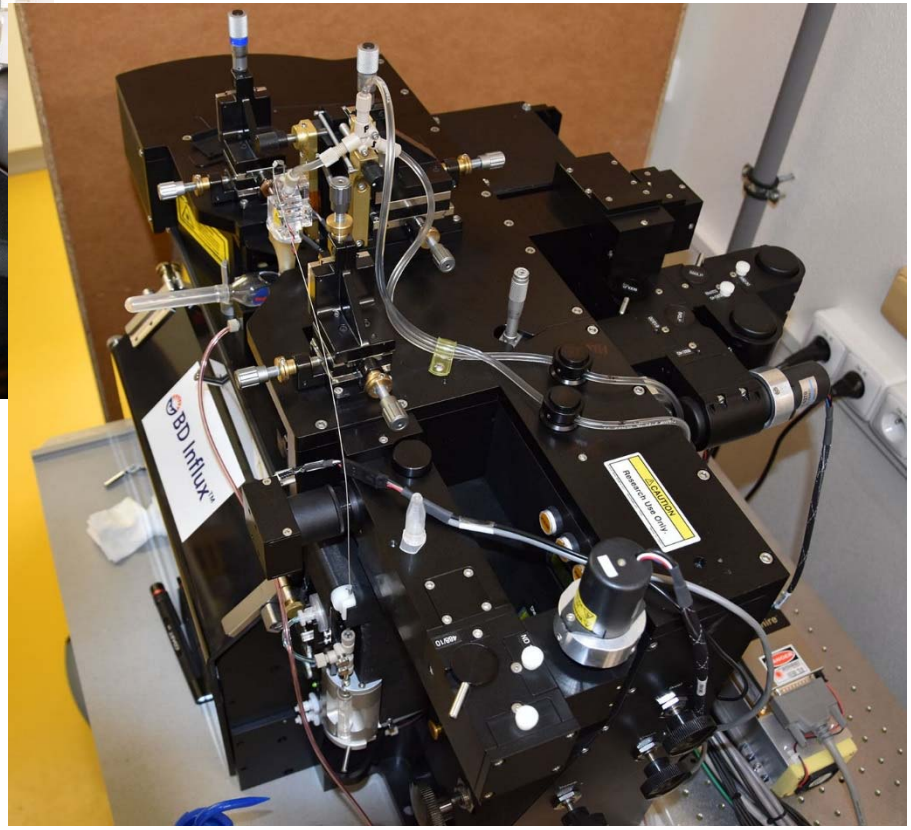
- For the cell surface display a target of library is fluorescently labeled ...



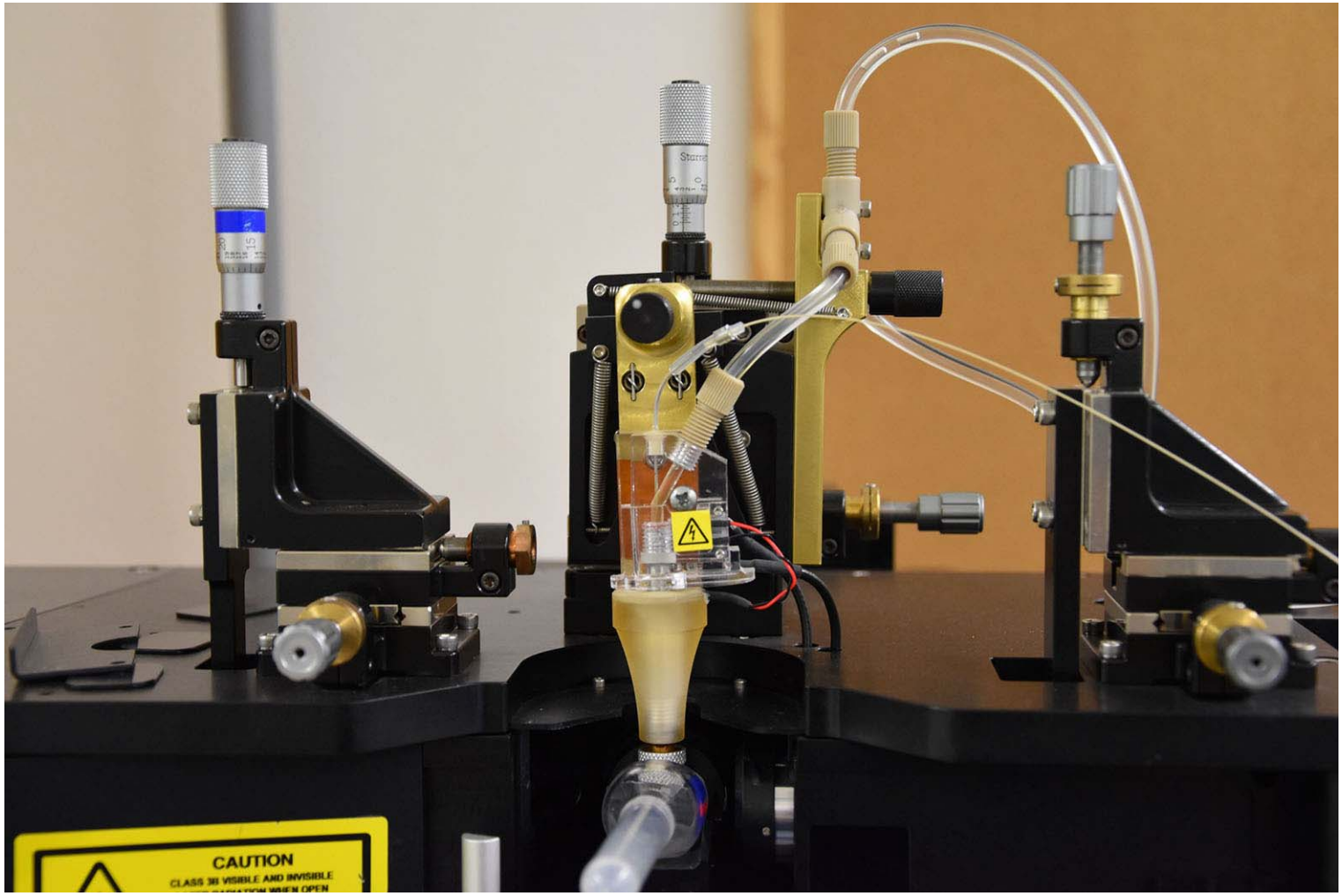




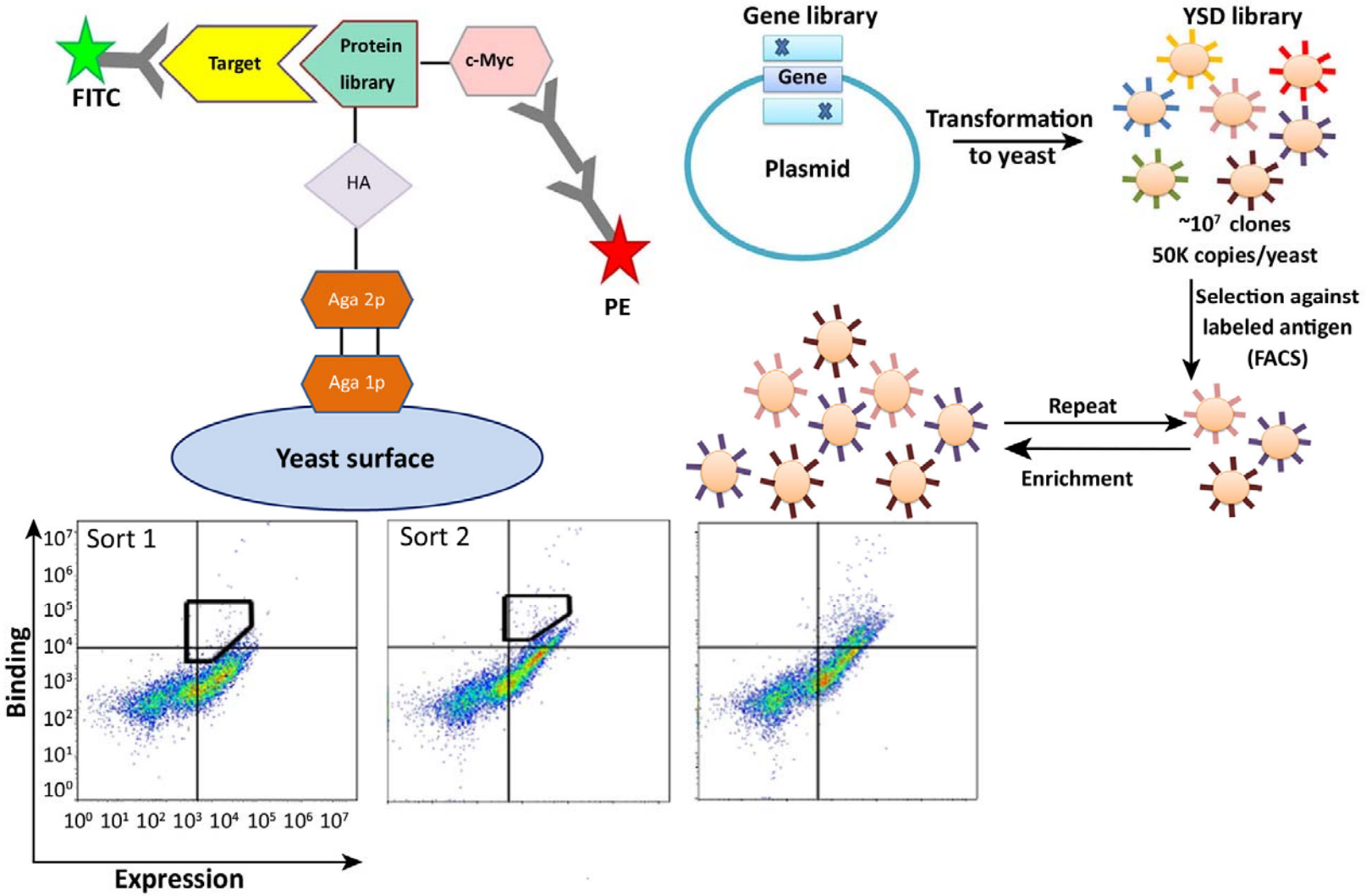
Fluorescence Activated Cell Sorting (FACS) – cell sorter



Fluorescence Activated Cell Sorting (FACS) – cell sorter



Fluorescence Activated Cell Sorting (FACS)



Cell-free display

Randomized library is translated using cell extract or purified ribosomes

Theoretical size of library depended only on number of ribosome in reaction (10^{14}/ml)

DNA first transcribed into mRNA library, which provides templates for the *in vitro* translation reaction later

Produced protein screened for binding to a substrate

- very big library (10^{10} *vs* 10^{14} – 10 000x more) -> increases the probability to select very rare sequences

- no unwanted selection pressure, such as poor protein expression, and rapid protein degradation

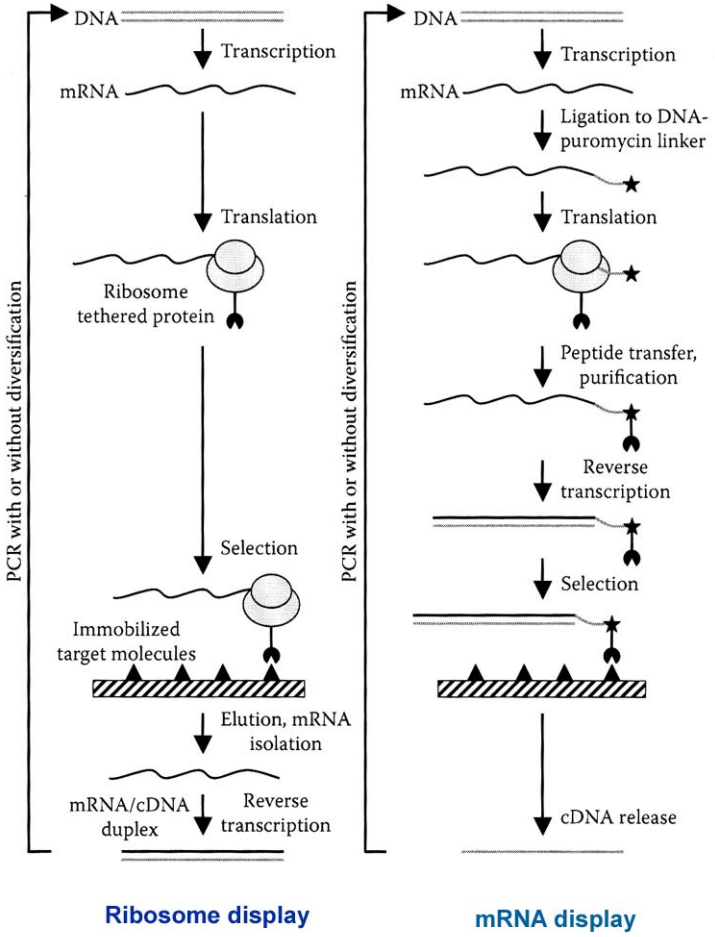
- technically more challenging
- potential problem with protein folding
- no post-translational modification

Cell-free display

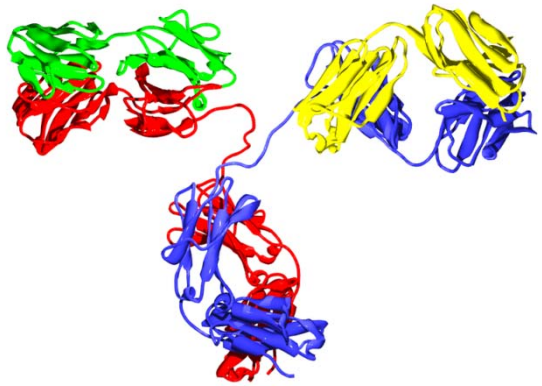
Ribosome display – generated protein remains together with ribosome and mRNA -> all this complex isolated via a screening process

mRNA display – generated protein covalently linked to mRNA using antibiotic puromycin

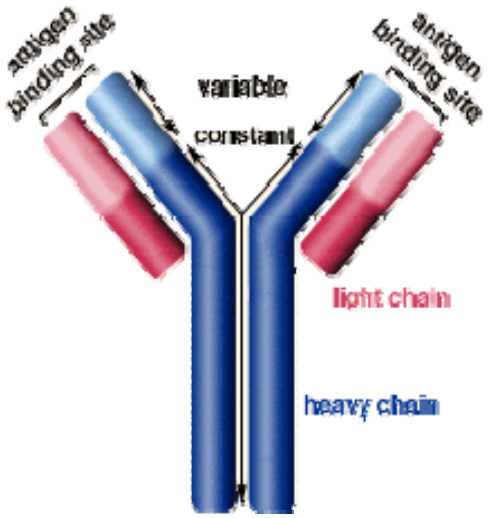
Stop codon has to be removed –> library genetically fused to a spacer sequence lacking a stop codon (**ribosome display**) or containing antibiotic puromycin at the 3' end (**mRNA display**)

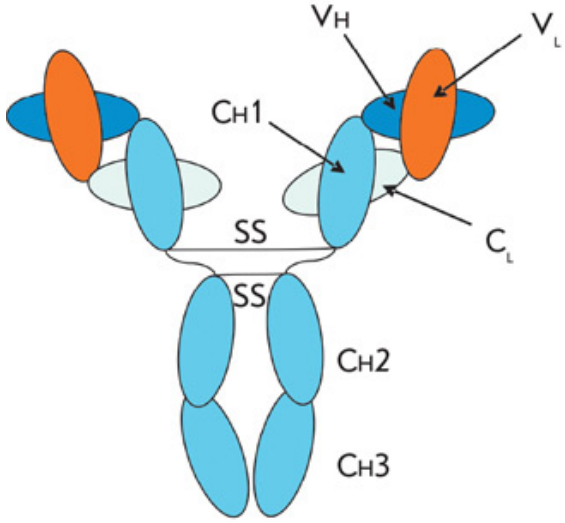


Engineered antibodies

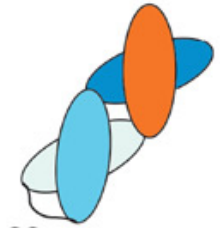


scFv = single-chain variable fragment
Fab = fragment antigen-binding

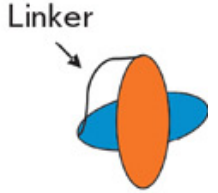




IgG



Fab

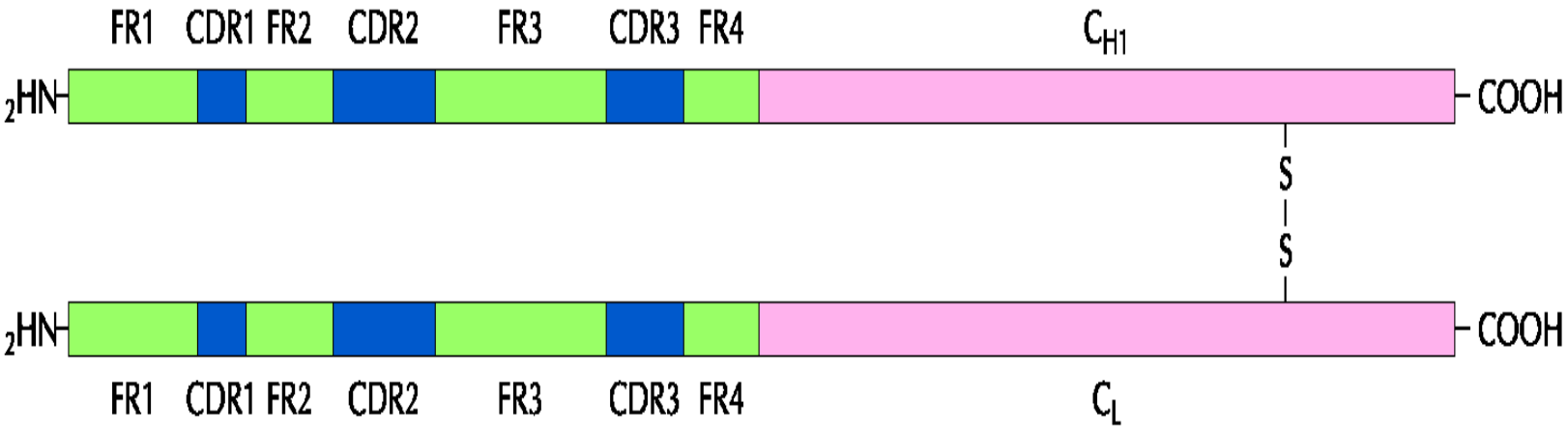


scFv

Engineered antibodies

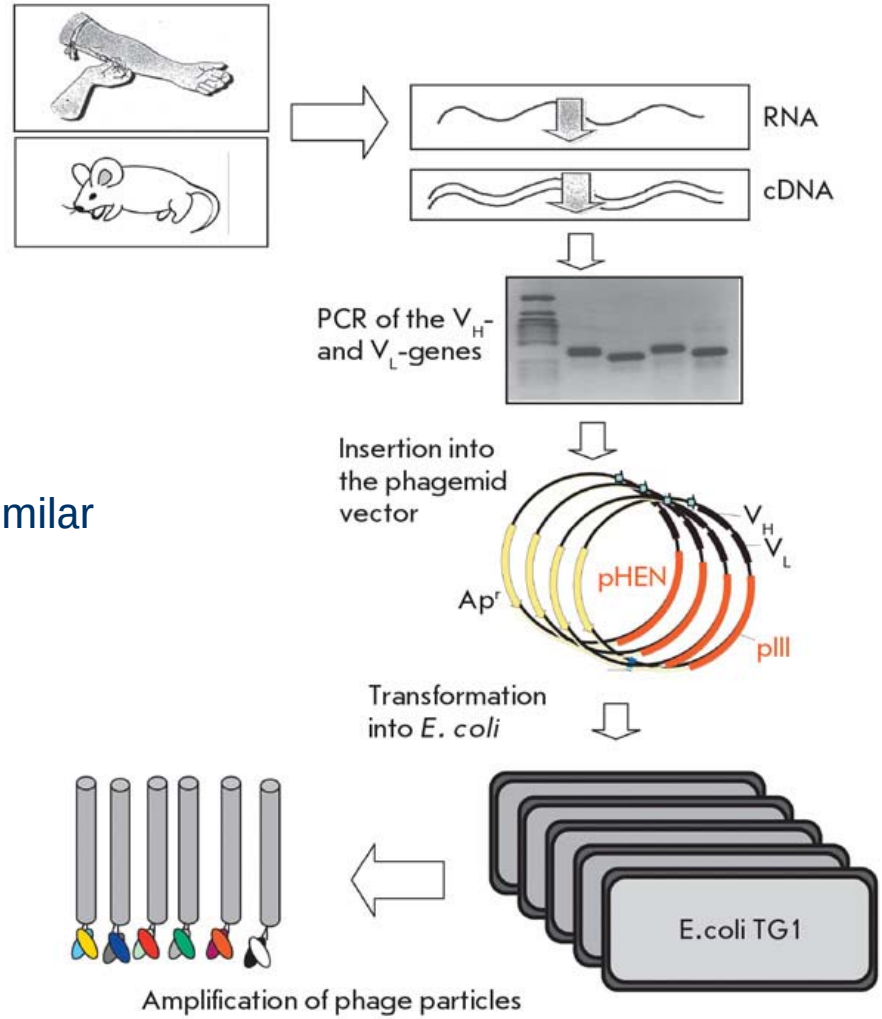
CDR - Complementarity-Determining Regions

Directed mutagenesis of CDRs gives antibodies with new specificities
Mutagenesis can be random (e.g. error-prone PCR or mutagenic primers)
or highly specific based upon predicted protein structure



Antibodies engineering by phage display

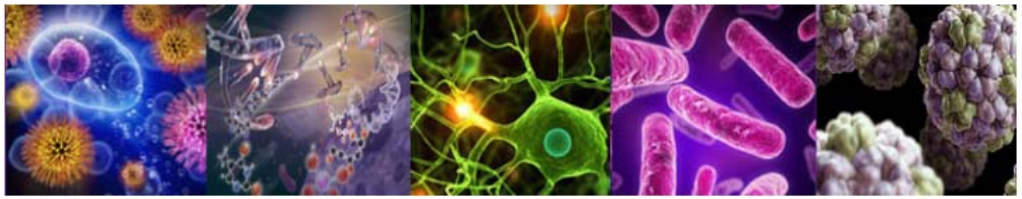
Once these phages are isolated and recovered they can be used to infect bacteria which will create a particle similar to a monoclonal antibody



Peptide Immunogen Phage Display Platform

A novel method to produce anti-peptide antibodies

CD CREATIVE DIAGNOSTICS

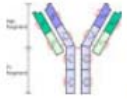


Dear Colleague,

Are you looking for antibodies to distinguish protein homologues? We, at Creative Diagnostics, have built up a novel platform, Peptide Immunogen Phage Display Platform, to efficiently raise short peptide specific antibodies with great selectivity, specificity and sensitivity.

This platform is based on our unparalleled expertise in phage display technology, in which the cDNA encoding a peptide is fused with the pVIII gene of a filamentous M13 phage. The peptide is expressed and displayed on the surface of filamentous phage at over 200 copies per phage particle. The displayed peptide together with the carrier phage virions is used for animal immunization; here the phage virion serves as one of the best immunization adjuvants. This platform supports antibody production in mice, rats, rabbits, chicken, sheep and llama.

Compared with conventional peptide-conjugation based methods, this platform supports



the formation of close-to-nature structure of the peptide that is bacterially expressed, processed and displayed on phage surface, thus increasing the chance to raise antibodies that can recognize the original antigen, from which the peptide immunogen is derived. Also, there are over 200 copies of the peptide on each phage virion, while a peptide conjugate has fewer copies of the peptide on each molecule. Importantly,

Table 1. Examples of antibody repertoires displayed on phage.

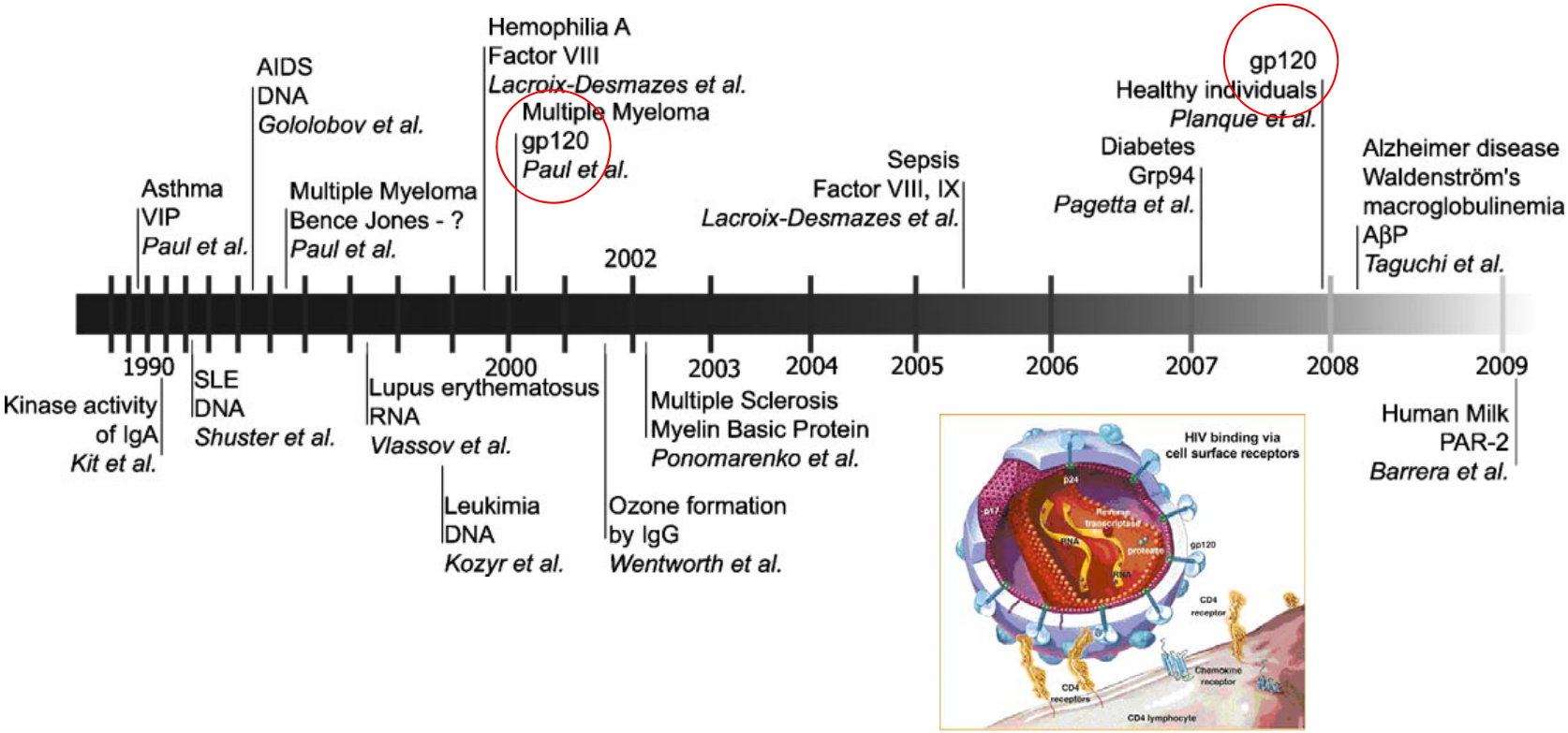
Donor	Source	Antibody format	Vector system (fusion gene)	Diversity	Reference
Immune					
Mouse	Spleen	scFv	Phage (gIIIp)	2×10^5	(Clarkson et al. 1991)
Mouse	Spleen	scFv	Phagemid (gIIIp)	6.5×10^6	(Kettleborough et al. 1994)
Mouse	Spleen	scFv	Phagemid (gIIIp)	2.4×10^6	(Lorimer et al. 1996)
Rabbit	Spleen, bone marrow	Fab	Phagemid (gIIIp)	2×10^7	(Lang et al. 1996)
Chicken	Spleen	scFv	Phagemid (gIIIp)	1.4×10^7	(Yamanaka et al. 1996)
Camel	PBL	V_H	Phagemid (gIIIp)	5×10^5	(Arbabi Ghahroudi et al. 1997)
Human	Bone marrow	Fab	Phagemid (gIIIp)	10^7	(Barbas et al. 1993)
Human	PBL	scFv	Phage (gIIIp)	5×10^8	(Cai and Garen 1995)
Human	Spleen (IgG) positive lymphocytes	scFv	Phagemid (gIIIp)	10^6	(de Wildt et al. 1996)
Human	Thymic lymphocytes	Fab	Phagemid (gIIIp)	4.8×10^6	(Graus et al. 1997)
Human	PBL	scFv	Phagemid (gIIIp)	2×10^8	(Mao et al. 1999)
Mouse	Pre-selected repertoires	Bispecific diabody	Phagemid (gIIIp)	7×10^4	(McGuinness et al. 1996)
Naïve					
Mouse	Bone marrow	Fab	Phagemid (gVIIIp)	5×10^6	(Gram et al. 1992)
Chicken	bursal lymphocytes	scFv	Phage (gIIIp)	2.7×10^7	(Davies et al. 1995)
Human	PBL (IgM)	scFv	Phagemid (gIIIp)	2.9×10^7	(Marks et al. 1991)
Human	PBL (IgG)	scFv	Phagemid (gIIIp)	1.6×10^8	(Marks et al. 1991)
Human	PBL, tonsil, bone marrow	scFv	Phagemid (gIIIp)	1.4×10^{10}	(Vaughan et al. 1996)
Synthetic					
Mouse	VH: HCDR3: 9[a]	V_H	Phagemid (gIIIp)	4×10^8	(Reiter et al. 1999)
Human	HCDR3: 16[a]	Fab	Phagemid (gIIIp)	5.0×10^7	(Barbas et al. 1992)
Human	LCDR1: 4, LCDR3: 6, HCDR2: 2, HCDR3: 5[a]	Fab	Phagemid (gIIIp)	3×10^8	(Garrard and Henner 1993)
Human	HCDR3: 4, LCDR3: 4[a]	scFv	Phagemid (gIIIp)	3×10^8	(Pini et al. 1998)
Human	HCDR3: 8[a]	scFv	Phagemid (gIIIp)	8×10^8	(Braunagel and Little 1997)
Human	49 VH, HCDR3: 5[a]	scFv	Phagemid (gIIIp)	10^7	(Hoogenboom and Winter 1992)
Human	50 VH, HCDR3: 4–12[a]	scFv	Phagemid (gIIIp)	10^8	(Nissim et al. 1994)
Human	49 VH, HCDR3: 4–12, 26 V_κ, LCDR3: 0–2, 21 V_λ, LCDR3: 0–3[a]	scFv	Phagemid (gIIIp)	6.5×10^{10}	(Griffiths et al. 1994)
Human	49 VH, HCDR3: 6–15[a]	scFv	Phagemid (gIIIp)	3.6×10^8	(de Kruif et al. 1995)
Human	VH: HCDR3: 5–12[a]	V_H	Phagemid (gIIIp)	2×10^8	(Davies and Riechmann 1995)

[a]The number of randomised residues in the CDR regions, and the number of applied V gene segments (only one if nothing is stated).
PBL – peripheral blood lymphocytes.

Catalytic antibodies?

Catalytic antibodies (abzymes) bind to a substrate, with the very high affinity that's typical of many antibodies, and facilitating the chemical reaction.

Old concept (Linus Pauling), described first in the mid-1980s



Anti-cocaine catalytic antibody?

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

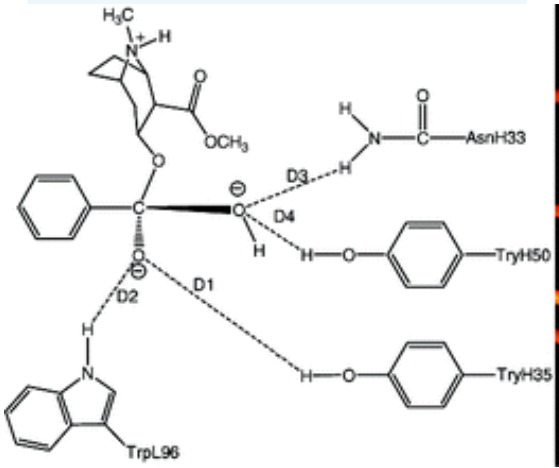
Prev.

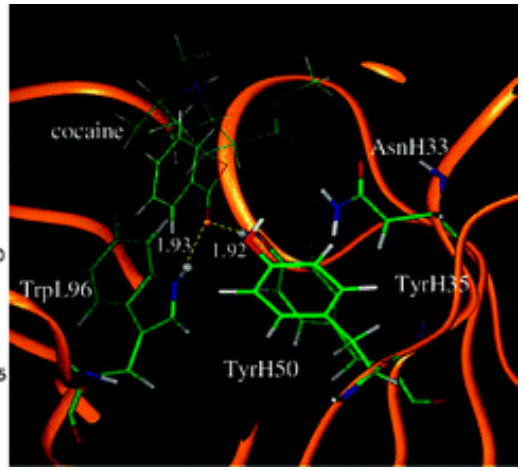
Modeling the Catalysis of Anti-Cocaine Catalytic Antibody: Competing Reaction Pathways and Free Energy Barriers

Yongmei Pan , Daquan Gao and Chang-Guo Zhan
Department of Pharmaceutical Sciences, College of Pharmacy, University of Kentucky, 725 Rose Street, Lexington, Kentucky 40536

J. Am. Chem. Soc., **2008**, *130* (15), pp 5140-5149
DOI: 10.1021/ja077972s
Publication Date (Web): March 15, 2008

ACS Section: Enzymes





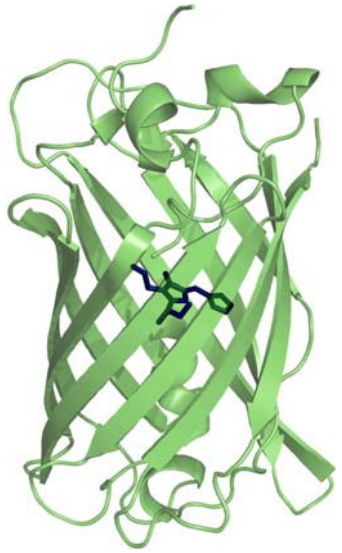
GFP is a frequent target of protein engineering

Cloned from the jellyfish *Aequorea victoria* in 1992

30 Da protein that exhibits bright green fluorescence when exposed to blue light

Original GFP had several drawbacks:

- low fluorescence yield
- two excitation maxima
- slow maturation
- oligomerization
- chlorine sensitivity
- green fluorescence only



GFP is a frequent target of protein engineering

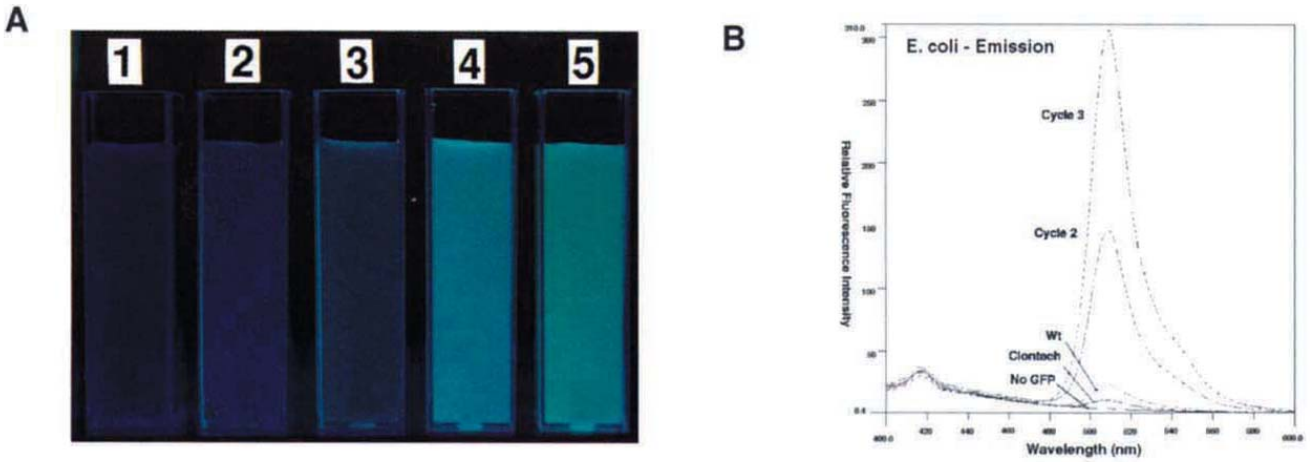
(*Nature Biotechnology*, 1996)

Started with a (almost) original GFP gene from the jellyfish *Aequorea victoria*

Performed recursive cycles of DNA shuffling.

Screened for the brightest *E.coli* colonies (using UV light)

After 3 cycles of DNA shuffling, a mutant was obtained with 45-fold greater fluorescence



GFP is a frequent target of protein engineering

Till now improved:

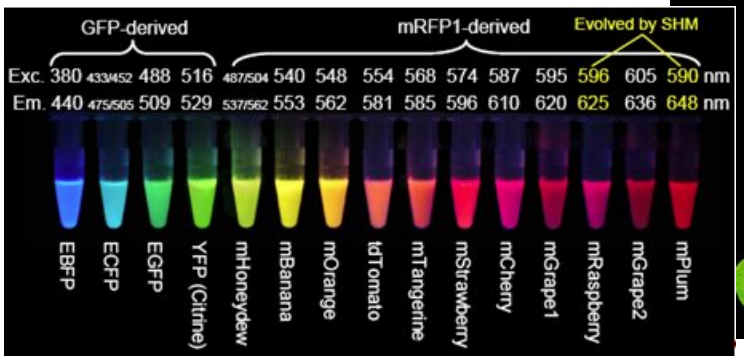
Excitation/emission wavelengths - Red, Yellow, Cyan, Blue FP

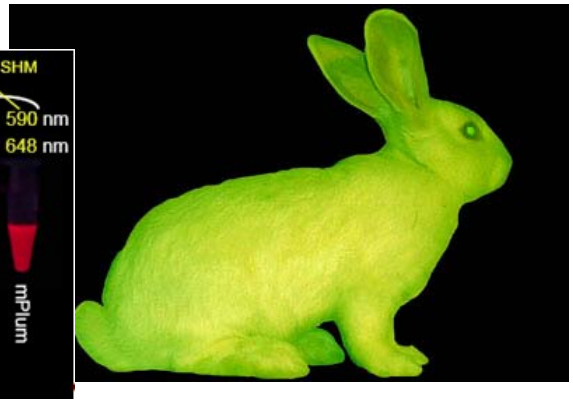
Fluorescence quantum yield

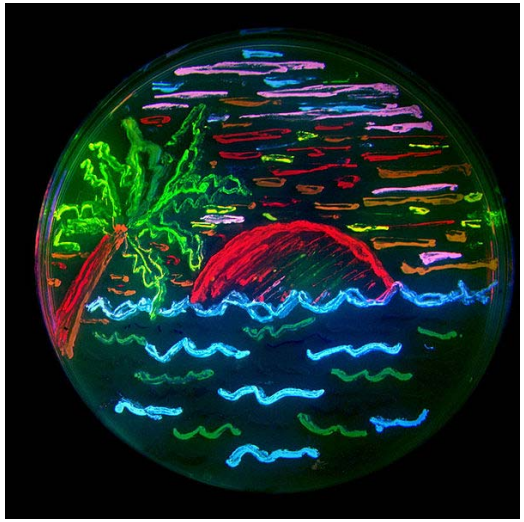
pH stability, chloride sensitivity,

Photostability and folding at 37°C (superfolded GFP – 2006)

Redox, Ca^{2+}, pH sensitive versions of GFP







"Fluorescent timer": Protein that changes colour with time

A fluorescent protein that changes color with time was generated from the red fluorescent protein (RFP) by error-prone PCR

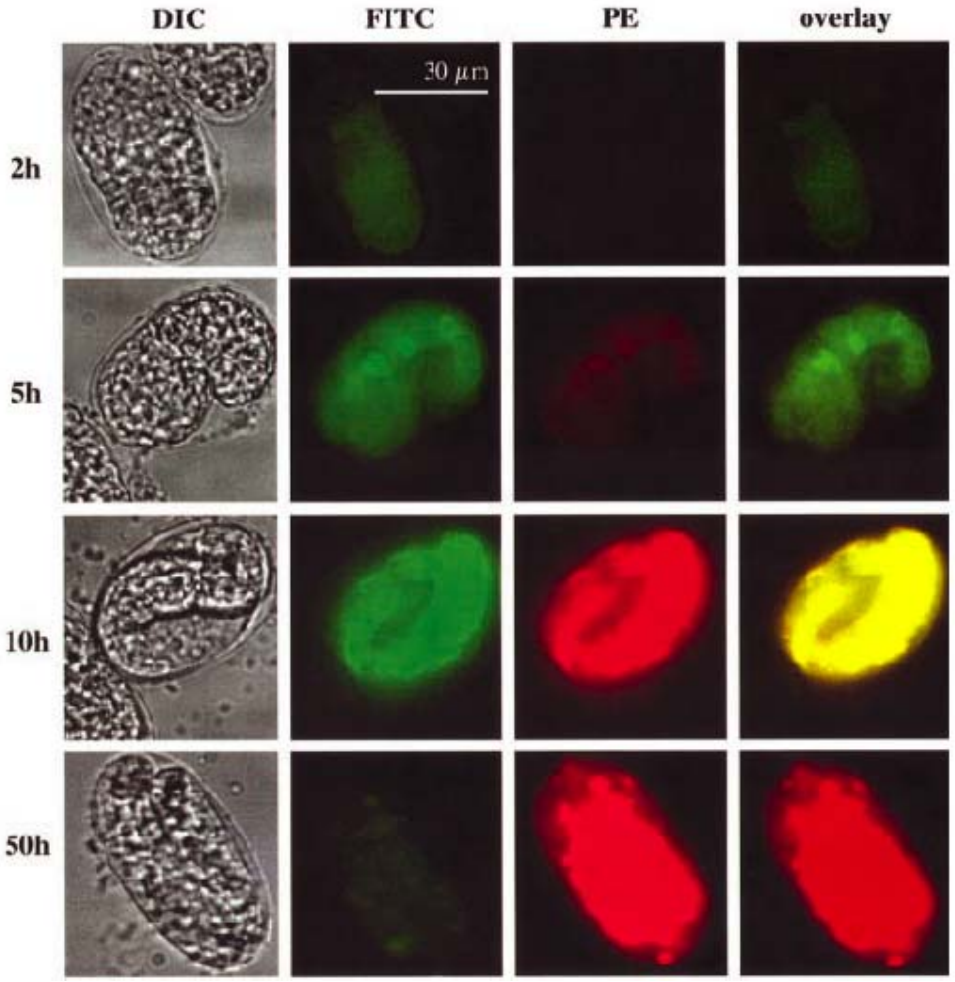
Mutants with various properties, such as faster maturation, double emission (green and red), or exclusive green fluorescence were isolated.

One mutant protein (E5) changes color over time: initially bright green, then to yellow, orange, and finally red.

E5 has two replacements: V105A and S197T

Science, 2000, 290:1585-1599.

"Fluorescent timer": Protein that changes colour with time



E5 used as a fluorescent clock:

heat shock-regulated expression of the E5 mutant RFP in *C.elegans*.

Science, 2000, 290:1585-1599.

‘Fluorescent timer’: Protein that changes color with time

E5 fluorescent protein be used as a fluorescent clock (timer) to examine promoter activity

Green: recent activation

Yellow -to- orange: continuous promoter activity;

Red: promoter activity has ceased after an extended ‘on’ period

Photoactivatalbe / switchable fluorescent proteins

Reversible...

	Ex1	Em1	Ex2	Em2	Photoconversion	
Dronpa	503	**518**	OFF	OFF	490 nm	reversible, 390 nm

Ireversible...

	Em1	Ex1	Em2	Ex2	Photoconversion	
Eos protein	506	**516**	571	**581**	Ultraviolet	ireversible