

Introduction:

Dr. Jonathan Nelson
OHSU Dept. of Nephrology and Hypertension

Employment opportunity

Basic gene cloning: plasmids and transformation

- Plasmid biology
- How to add DNA to a plasmid
- Transformation of *E. coli*

The key role of *E. coli* and an RNA expression plasmid in mRNA vaccine production (Pfizer/BioNTech, specifically)

<https://www.nytimes.com/interactive/2021/health/pfizer-coronavirus-vaccine.html?action=click&module=Spotlight&pgtype=Homepage>

-

Readings:

- 1) *22 MC4 Plasmids*. History of plasmid development, plasmid replicons, care of *E. coli* and plasmids, specialized plasmids, transformation, alpha complementation.
- 2) *25 MC4 Antibiotics*. Summary of several antibiotics used for plasmid maintenance, and how they work
- 3) *24 MC4 Cloning in Plasmids*. Practical aspects of cloning and transformation, plasmid dephosphorylation. Discussion of recombinational and ligation independent cloning.
- 4) Gibson assembly 2009. Report demonstrating easy DNA assembly using overlapping sequences.
- 5) IVA (in vitro assembly) 2016. Cloning without ligation.
- 6) Link: www.addgene.org, a nonprofit plasmid repository

Cloning: isolation and propagation of a specific piece of DNA

Cloning vector: a movable DNA element that is modified to contain another piece of DNA

General features:

- Easy to isolate and purify
- Can be manipulated in the test tube
- Can be moved into a new cell
- Selectable: easy to propagate and test for presence or absence

Types of cloning vectors:

- Plasmid
- Virus
- Artificial chromosome
- transposon

Bacterial plasmid: Non-chromosomal DNA molecule

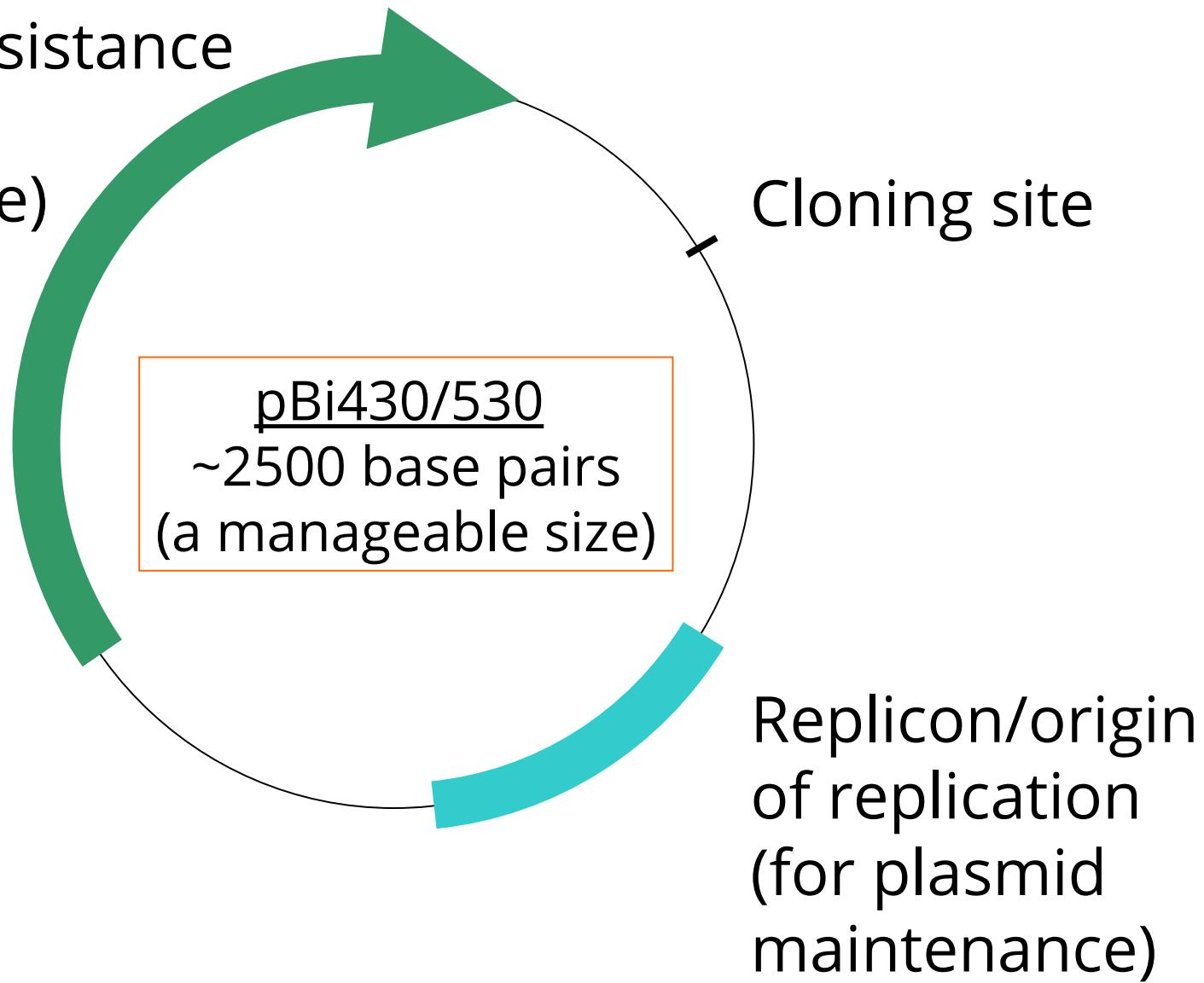
- Small, circular, supercoiled
- Replicates independently of the bacterial chromosome
- Copy number in cell is maintained through an origin of replication (replicon)
- Selectable through a gene that is beneficial for the host bacterium, eg. antibiotic resistance

A useful plasmid

1. Is relatively small
 - Gives higher copy #, stability, and transforming efficiency
 - Can accept larger pieces of DNA
 - Easier to handle (less susceptible to breakage)
2. Has a selectable marker (e.g. antibiotic resistance)
3. Has single sites for many restriction enzymes

Plasmid: basic parts list

gene that encodes
antibiotic resistance
(for plasmid
maintenance)



Replicon: how the plasmid replicates

- A DNA sequence and associated factors
 - origin of replication, ori: a site on the DNA
 - protein and RNA factors that manage replication
- The replicon helps define the “ copy number ” , the number of plasmid copies per cell

A few common plasmid replicons

<u>PLASMID</u>	<u>REPLICON</u>	<u>COPY #</u>
pBR322	pMB1	15-20
pUC19	Modified form of pMB1 (RNA II mutation)	500-700
pACYC	p15A	18-22
pSC101	pSC101	~5

Plasmid copy number is an important consideration

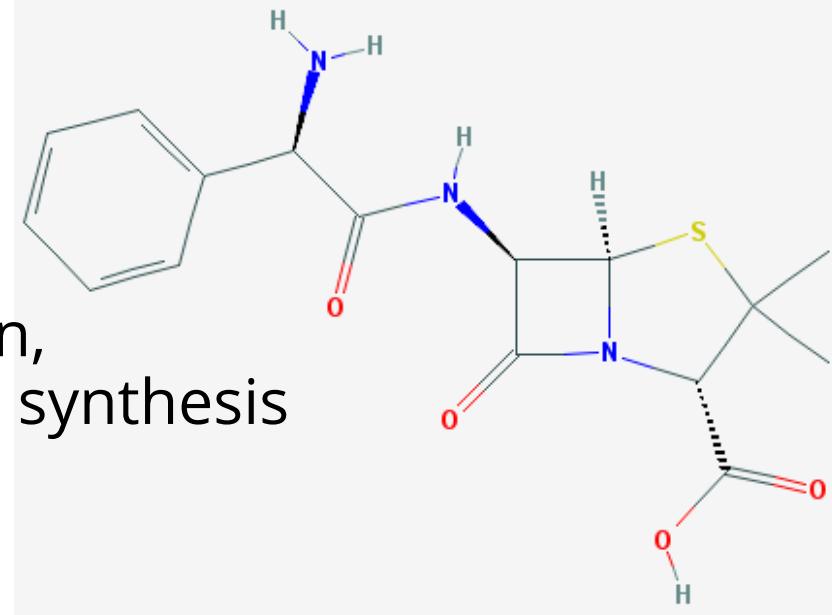
- High copy number plasmids
 - More copies/cell = easy to make a lot of DNA in a plasmid prep
 - Used for routine manipulation of small (<15 kb) recombinant DNAs
- Low copy number plasmids
 - For genes that are lethal or unstable in high copy number plasmids
 - Bacterial Artificial Chromosomes (BACs) that can propagate large (>100 kb) recombinant DNAs. Lots of DNA is a drag on cell physiology

How to make sure a plasmid is maintained

- Plasmids contain **selectable markers**: genes carried by the plasmid that confer functions required for host survival
- Selection: only those cells with the plasmid will survive
 - Allows transformation (a rare event) to be easily detected
 - Cells cannot lose the plasmid, even if it causes a selective disadvantage (eg. slow growth or toxic gene product)

Antibiotic resistance genes

- Antibiotic
 - **ampicillin** (related to penicillin, carbenicillin) inhibits cell wall synthesis
- Resistance gene
 - beta lactamase (bla) breaks down ampicillin, so cells carrying this gene are called **amp^r**

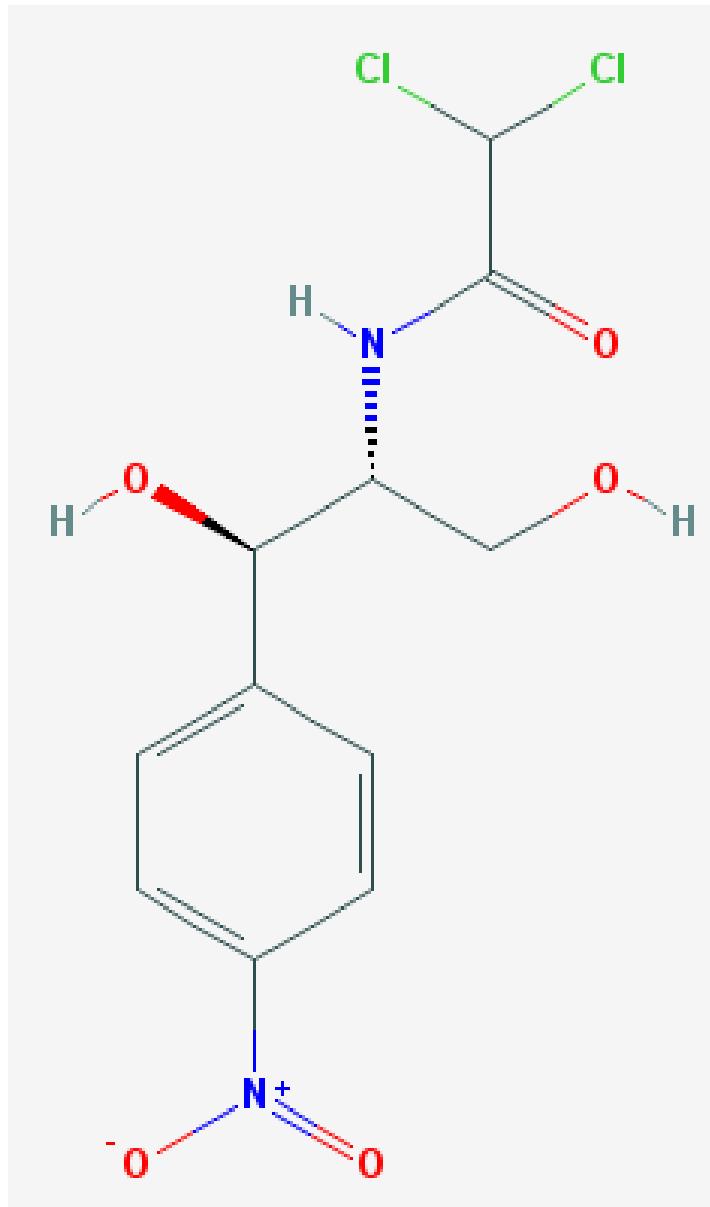


Keep in mind

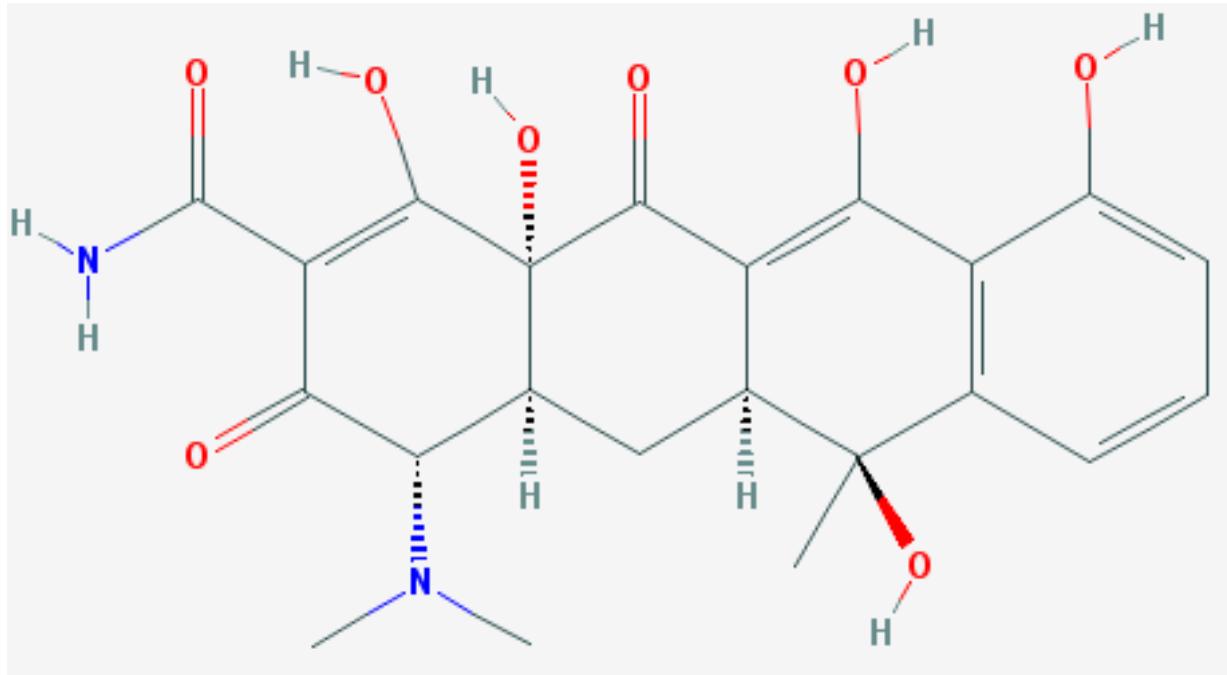
- beta-lactamase is secreted into the medium where it breaks down the antibiotic and depletes it
- ampicillin sensitive cells can grow, defeating the selection

Antibiotic resistance genes

- Antibiotic
 - chloramphenicol (cm) inhibits translation (peptidyl transferase activity of the ribosome)
- Resistance gene
 - chloramphenicol acetyl transferase (CAT) inactivates chloramphenicol



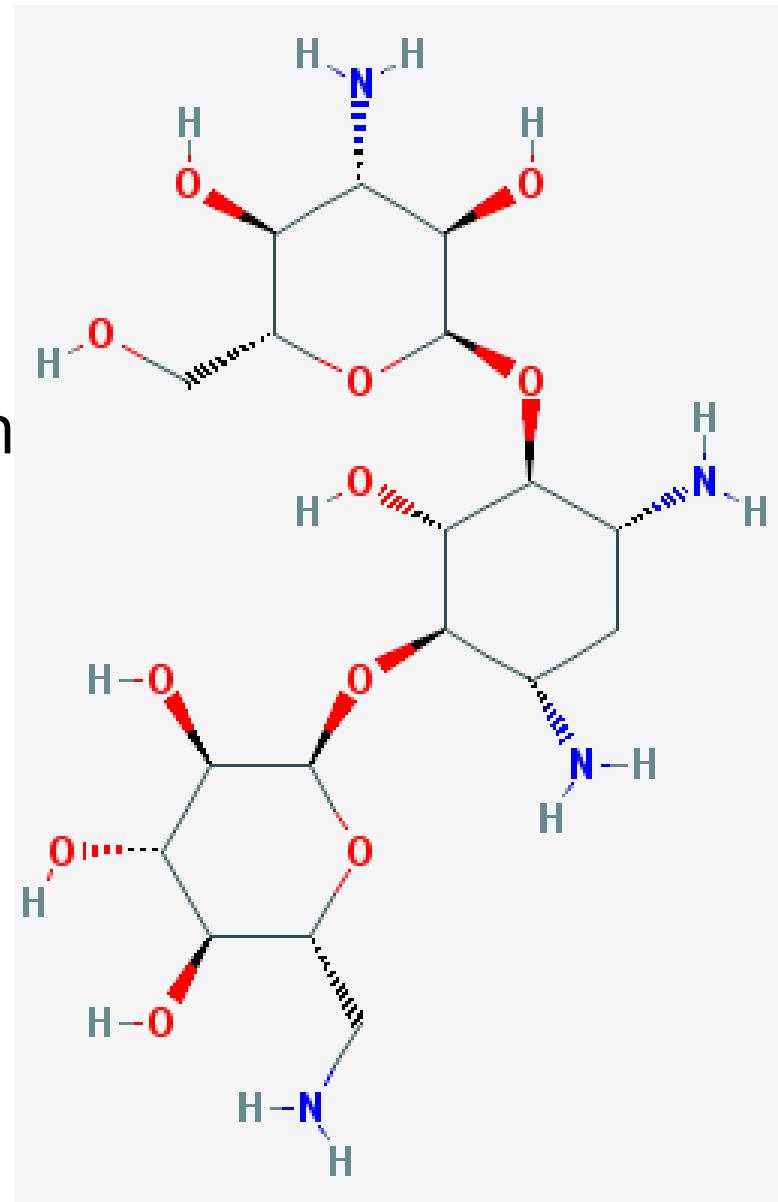
Antibiotic resistance genes



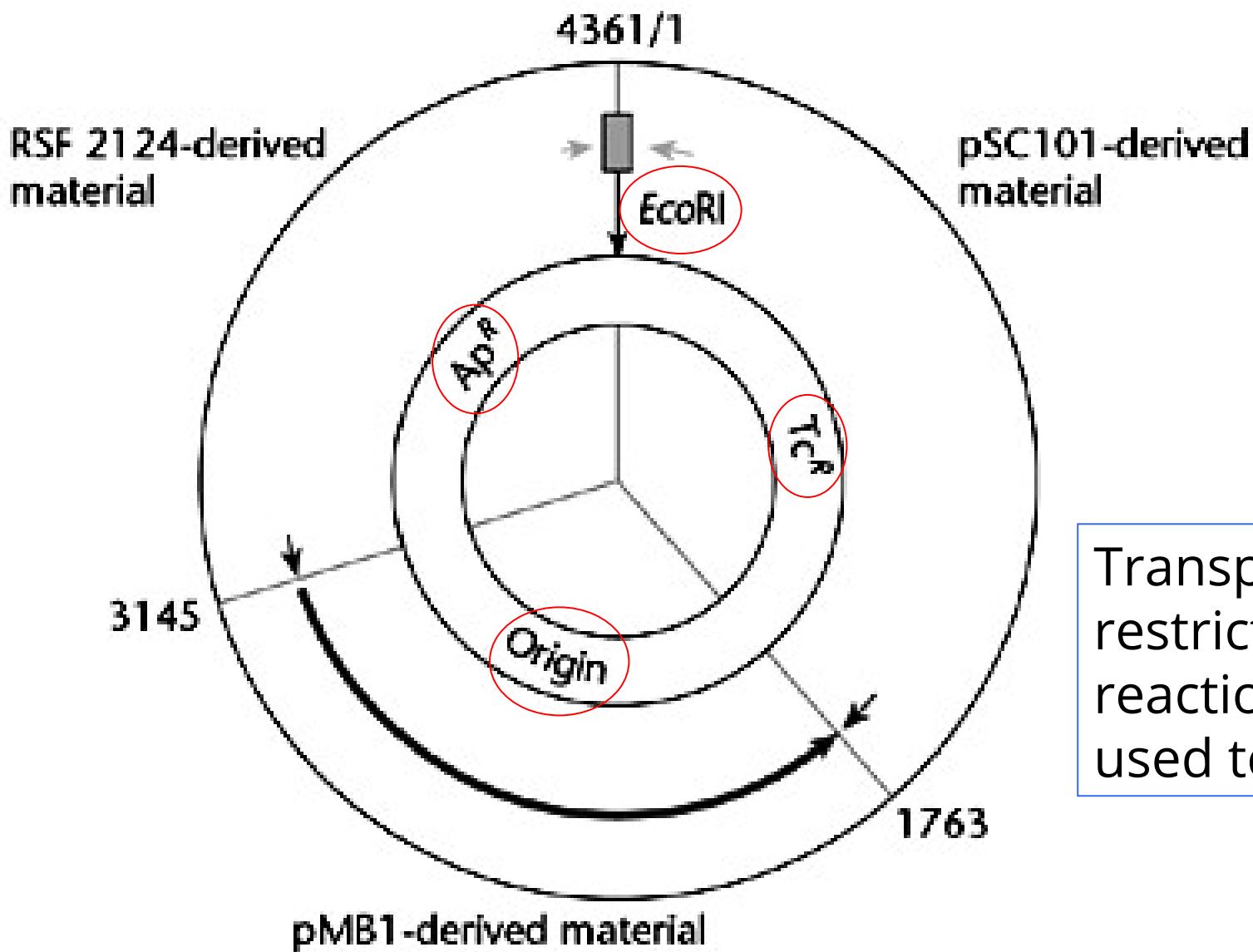
- Antibiotic
 - **tetracycline** inhibits translation (interacts with ribosome 30S subunit)
- Resistance gene
 - Tet A (C) protein confers resistance to by moving this antibiotic out of the cell

Antibiotic resistance genes

- Antibiotic
 - kanamycin: aminoglycoside antibiotic, inhibits translation (interacts with ribosome 30S subunit)
 - others of this class include G418, neomycin
- Resistance gene
 - bacterial aminophosphotransferase inactivates kanamycin by phosphorylation

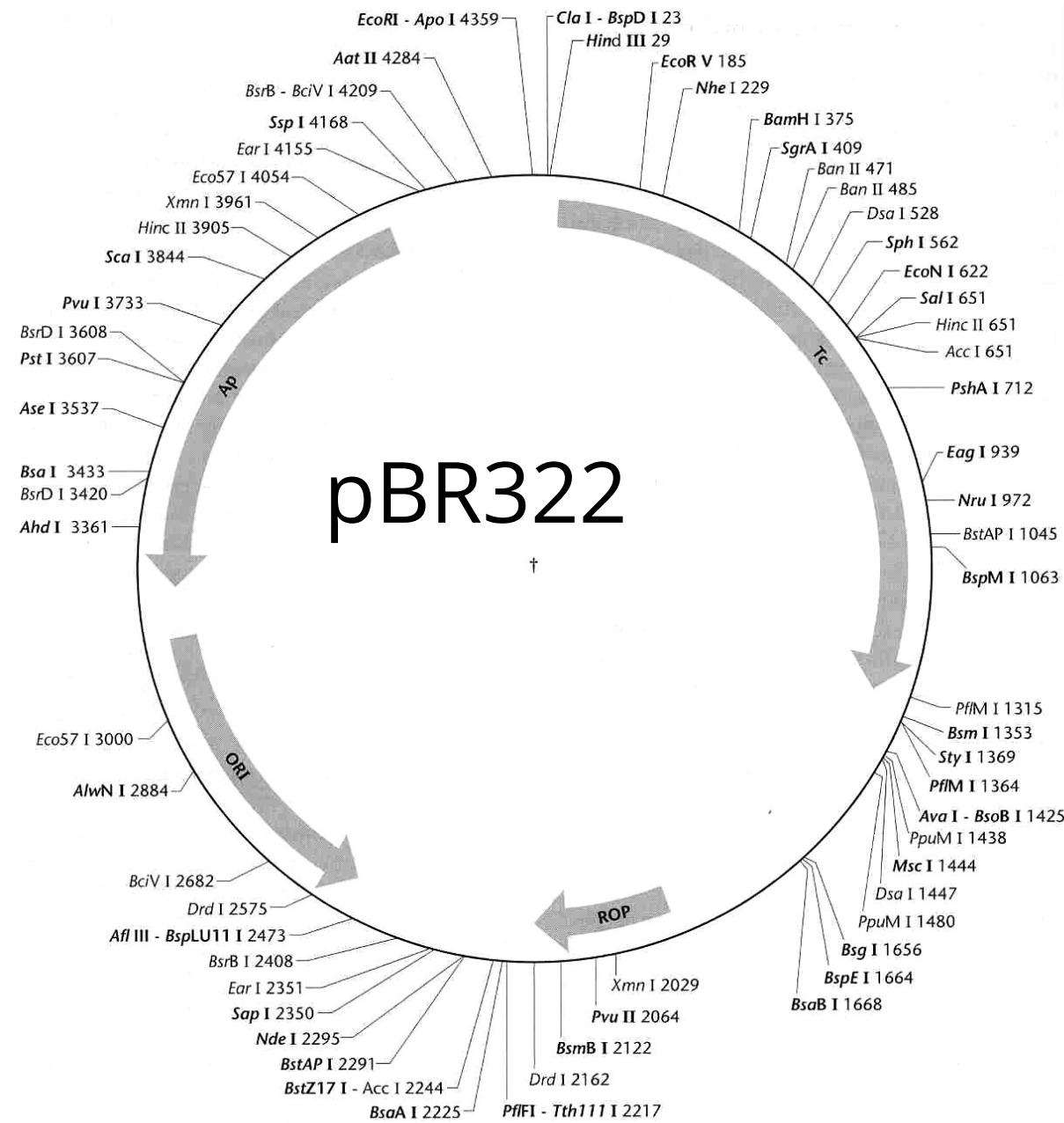


pBR322: the first widely used plasmid



Transposition and
restriction/ligation
reactions were
used to make it

Using pBR322



Clone into sites in the Tc^r gene,

Plate on ampicillin,
then 'replica plate' on
tetracycline.

Recombinants are amp
resistant, tet sensitive

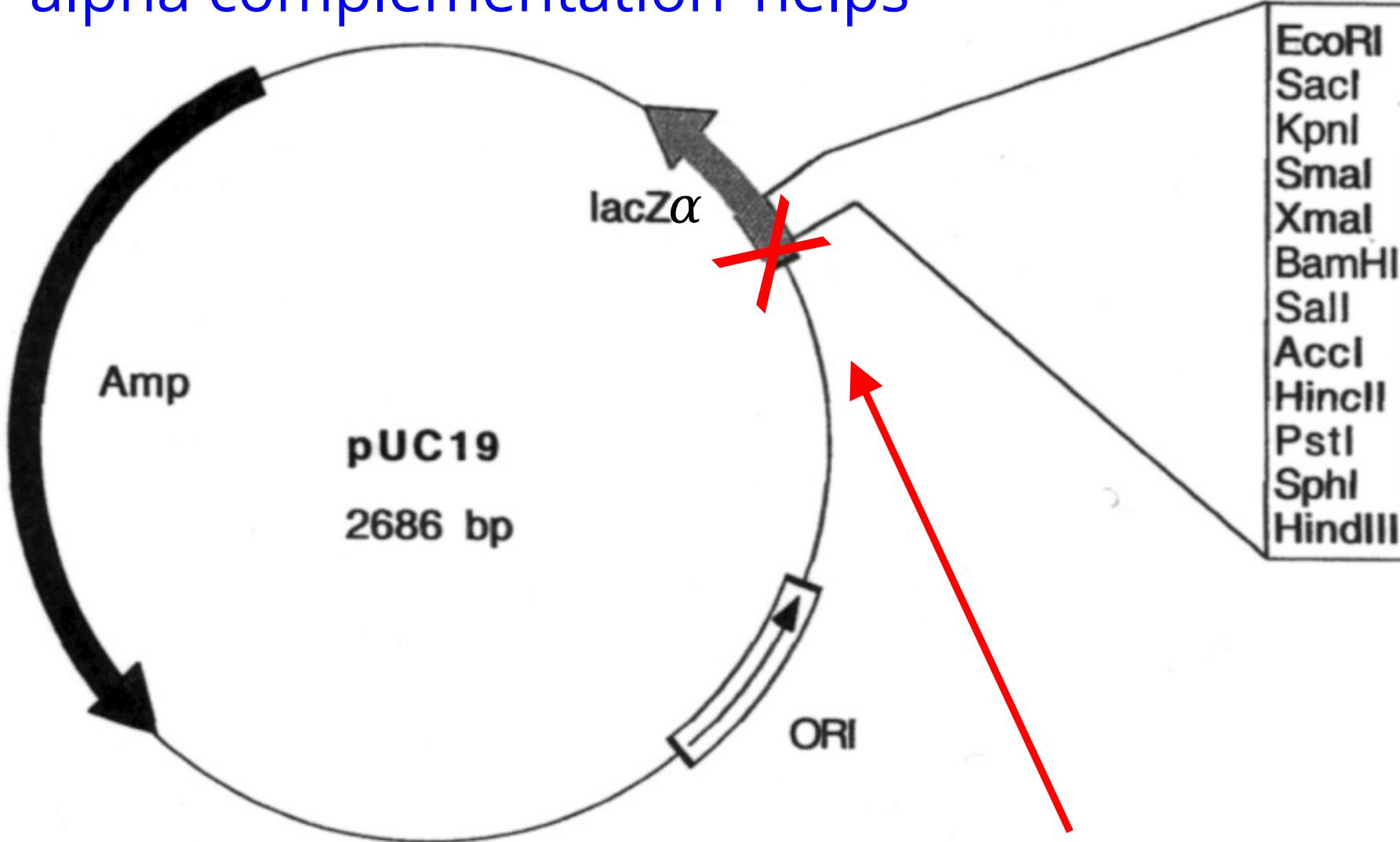
But: pBR322 has **low copy number, large size**

pUC plasmids: improved cloning vectors

- Reduced size (about 2000 bp), easier for cells to handle
- Multiple cloning site (MCS, also called “ poly-linker”): unique sites for lots of different restriction enzymes
- Very high copy number (mutation in RNA II)
- A “ blue-white” screening tool for recombinants (“ alpha complementation” is disrupted by foreign DNA in the MCS)

Identifying recombinant plasmid: 'alpha complementation' helps

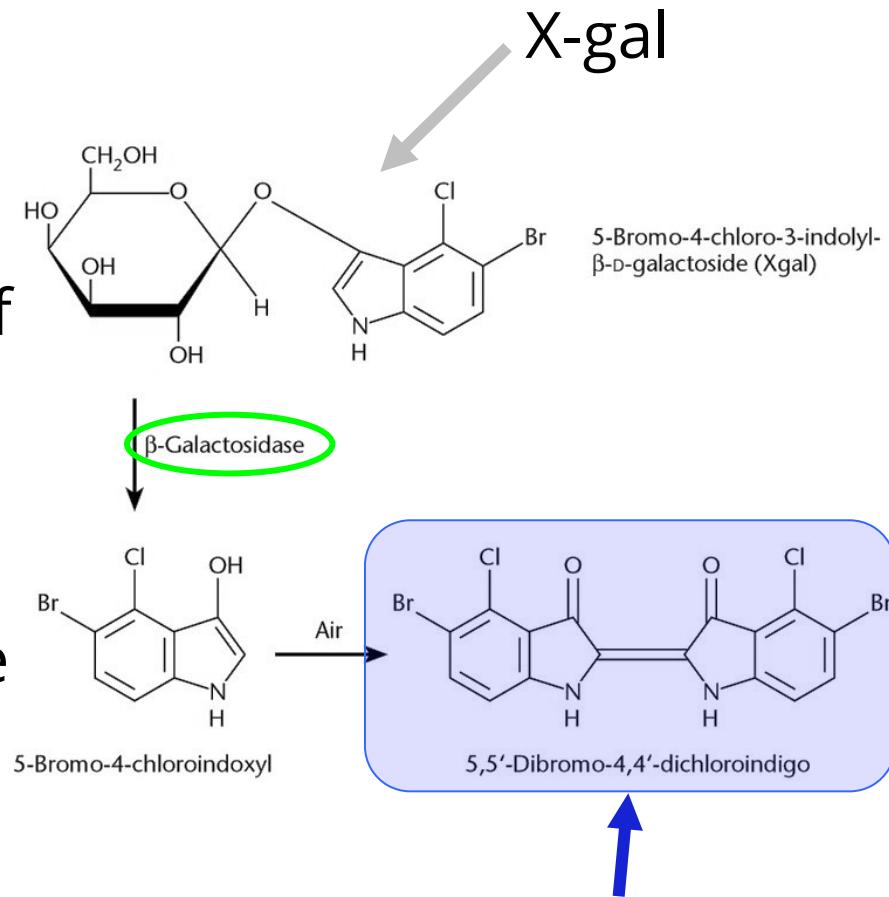
Polylinker
(MCS)



DNA cloned in the MCS interrupts the lacZ gene alpha fragment

Alpha complementation

- Plasmid has the N-terminus of beta galactosidase (alpha fragment)
- Host *E. coli* has the C-terminus of beta galactosidase (omega fragment)
- Both the N- and C-terminal fragments are needed for enzyme activity
- if beta galactosidase is present, X-gal is cleaved, producing a bright blue product

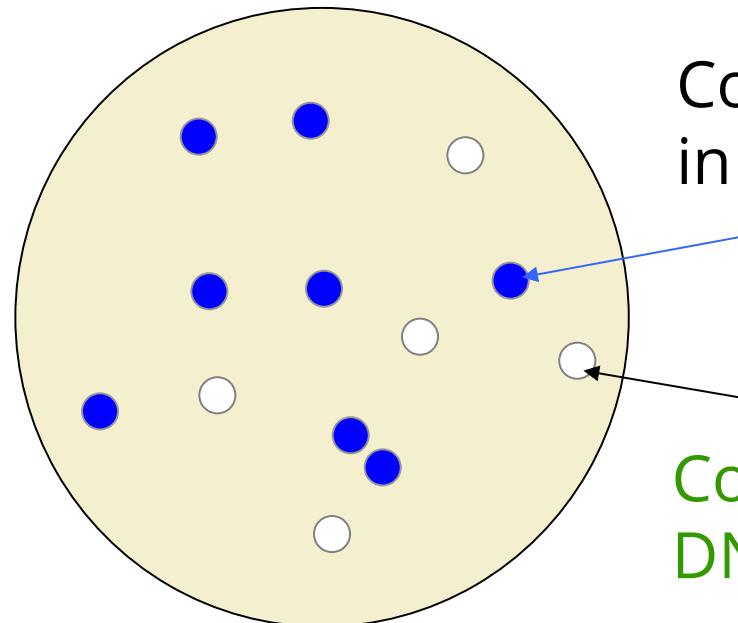


Bright blue

Alpha complementation on a petri plate

- Plasmid has alpha fragment of beta galactosidase with an MCS, bacterial host has omega fragment
- Clone DNA into the MCS, no alpha fragment
- No alpha fragment, no functional B-galactosidase
- No B-gal, no blue color
- white colonies

pUC19
transformation
plate



Colony with no DNA
in MCS. Don't want.

Colony with cloned
DNA in MCS. Want.

Many specialized plasmids are available

- For expression of specific genes
 - Plasmids are engineered to include “tags” on expressed protein, to assist in purification or tracking
- For cloning of unstable or toxic genes
 - Low copy number
- For cloning gigantic pieces of DNA
 - Low or single copy number

Want a new plasmid? Addgene.org, a plasmid repository

- <https://www.addgene.org/>

Keeping track of your plasmid sequence and structure

- A plasmid editor (ApE)
 - Save plasmid sequence
 - Include annotations to the plasmid
 - <http://biologylabs.utah.edu/jorgensen/wayned/ape/>

How to create a recombinant plasmid (general protocol)

1) Prepare the plasmid vector

- a) Make the plasmid linear (PCR, or restriction enzyme)
- b) Purify, quantify the DNA

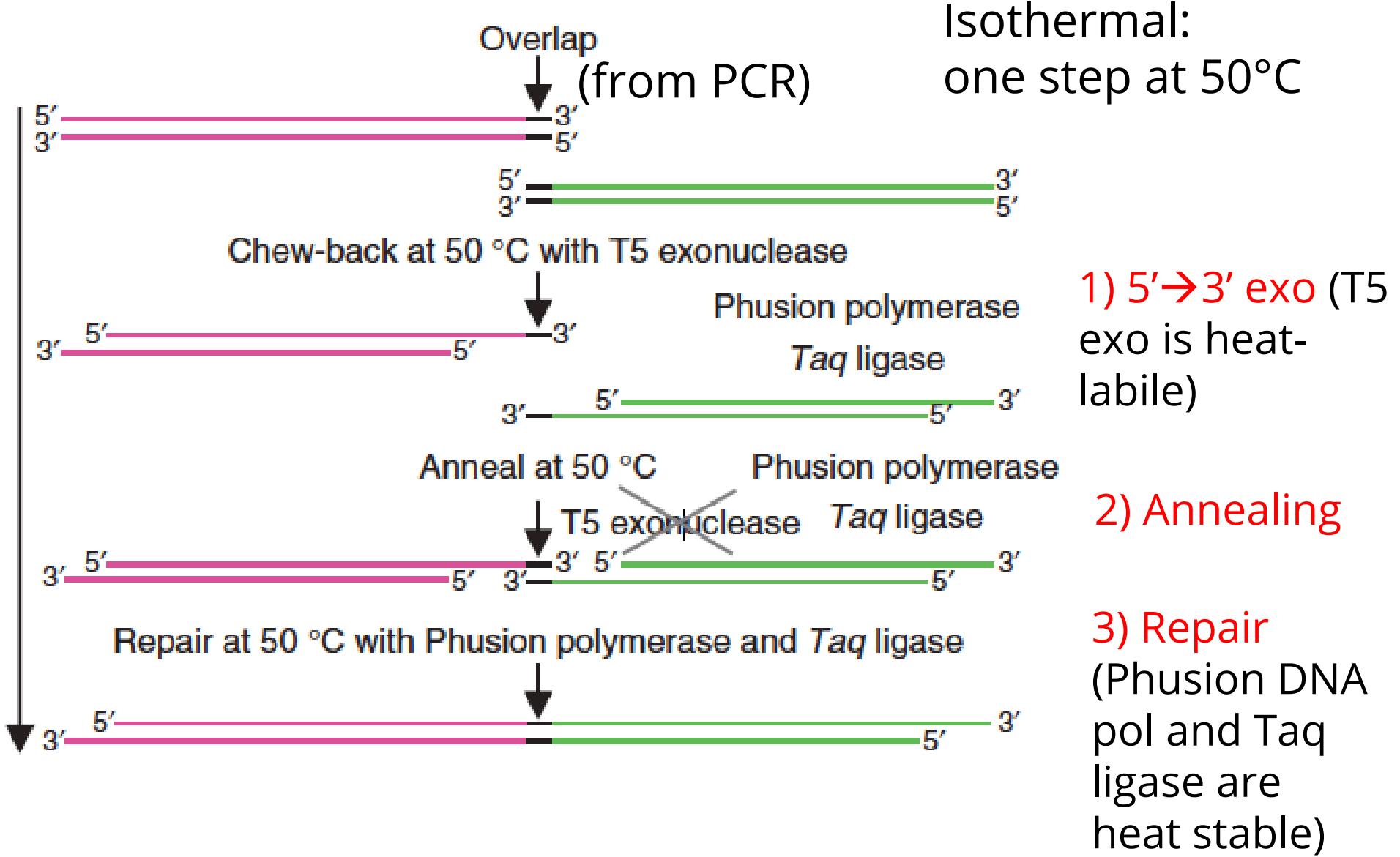
2) Prepare the DNA to be cloned

- a) Obtain the DNA
 - PCR
 - Small pieces of genomic DNA
- b) Treat the ends of the DNA to make them clonable
 - Primers with modified 5' ends
 - Add sequence to 3' ends with terminal transferase
 - Ligate adaptor sequences to ends

3) Combine the plasmid and clone DNA

- a) Standard cloning: ligase
- b) Gibson Assembly: Exonuclease, DNA pol, ligase
- c) In vivo assembly (IVA): the cell assembles pieces with overlapping ends

Gibson DNA assembly: make synthetic genes, pathways, or *entire genomes*.



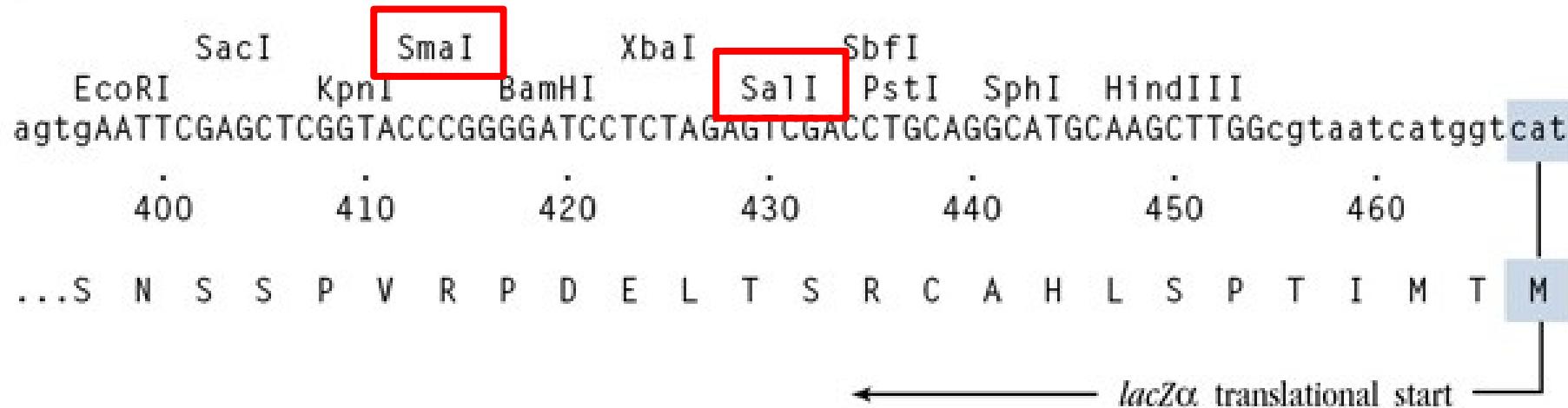
Gibson assembly example

Cloning vector: plasmid pUC19

Cloned DNA: Pf1831 (*Pyrococcus furiosus* histone protein)

```
ATGGGAGAATTGCCAATTGCCCGAGTTGACAGGCTTATAAGAAAGGCAGG  
TGCTGAAAGAGTTAGCGAGCAAGCAGCAAAGTCCTAGCAGAGTACCTCG  
AAGAGTACGCTATTGAGGTGCAAAGAACGGCAGTAGAGTTCGCAAGGCAC  
GCAGGTAGAAAGACTGTTAAGGTTGAAGACATTAAGCTCGCAATTAGAG  
CTGA
```

pUC19 MCS



Cloning strategy: <http://nebuilder.neb.com/>

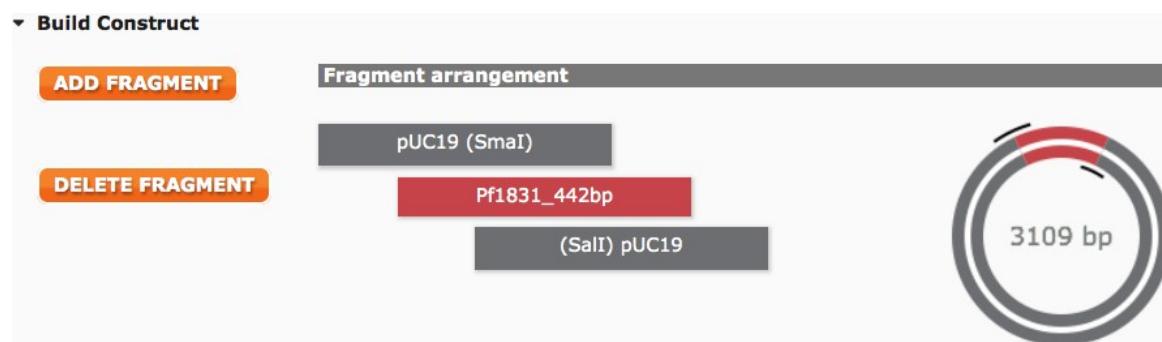
Parameters for cloning by Gibson assembly:

The final product has to be circular

Need at least a 15 bp overlap at the cloning junctions (this is accomplished by extra sequences at the 5' end of the PCR primers)

Picking primers: can be done manually, or with an automated online tool

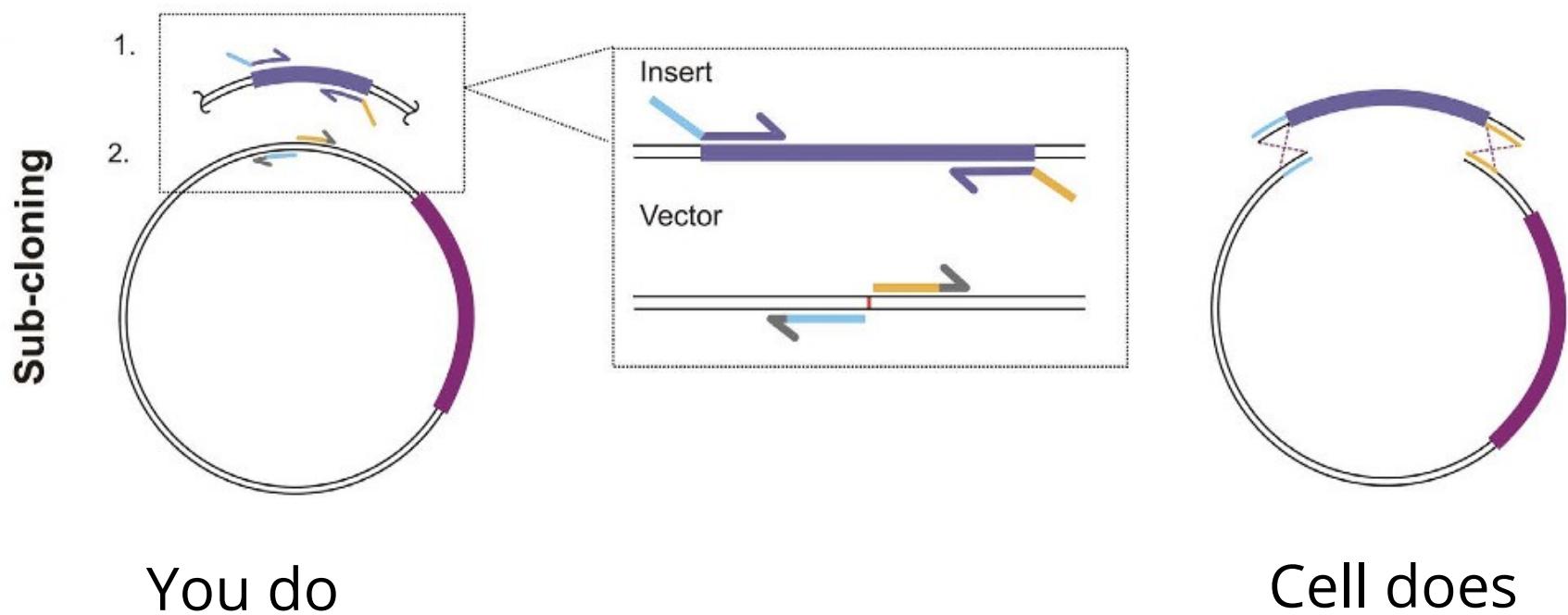
Overlaps	Oligo (Uppercase = gene-specific primer)	Anneals	F/R	3' Tm	3' Ta *
pUC19	tcgagctcggtaccc GGCAATCTATTGGAAATC	Pf1831	Fwd	56.0°C	56.0°C
pUC19	ttgcatgcctgcagg GATATACTTTAATTCGGGAGG	Pf1831	Rev	57.2°C	56.0°C



"IVA" cloning: In Vivo Assembly

- *E. coli* has a recombination machinery that only requires short regions of homology (15-20 base pairs)
- Transform cells with DNA with overlaps, the recombination machinery will put them together automatically, and with reasonably high efficiency

b



Moving DNA: transformation of *E.coli* with plasmid DNA

Cells acquire DNA by

- Chemical competence
- Electroporation
- Natural competence (in some lucky cases)

E.coli strain: must be antibiotic sensitive, and should not have restriction enzyme systems

Following transformation, the cells are cultured on a petri plate

Making chemically competent *E. coli*

- Grow cells to “ mid-log” phase (A_{600} of 0.4) spin to get cell pellet
- Suspend cells in CaCl_2 (100 mM), pellet again
- Suspend in small volume of CaCl_2 /glycerol
- Freeze cells (-80°C) or go straight to transformation protocol

Transformation of chemically competent cells

DNA binds to cells



- Mix DNA and competent cells, on ice for 30 min.

DNA uptake by cells



- Heat shock (42°C) for 1.5 minutes
- Add growth media, 37°C for 1 hour

Cells recover



- Plate on growth medium plus selection (antibiotic) for the plasmid

Selection occurs



Efficiency: $\sim 10^5 - 10^8$ antibiotic resistant colonies/microgram plasmid DNA

Transformation by electroporation

- Grow cells to A_{600} of 0.4
- Centrifuge and resuspend in water + 10% glycerol (do this 4 times to reduce conductivity)
- Place cells with DNA in electrode-containing cuvette, deliver electrical pulse
- If there is arcing (sparks) transformation efficiency will be poor (uneven transfer of charge). To avoid this make sure the ion concentration is very low (less than 10 mM salt)
- $> 10^9$ transformants/microgram DNA (ideally)

Setting up a transformation – treatments and controls

1. *No plasmid* (negative control, nothing should grow on this plate)
2. *Supercoiled plasmid* of a known concentration (to determine efficiency of competent cells)
3. *Plasmid DNA* backbone: without clone DNA (background transformants)
4. *Plasmid DNA with clone DNA* (desired products)

Ideal results from a transformation

- 1) No DNA--No colonies
- 2) 2 nanograms (2×10^{-3} micrograms) supercoiled plasmid DNA--500 colonies (efficiency of cells: 2.5×10^5 transformants per microgram DNA)
- 3) Plasmid alone--small number of colonies
- 4) Plasmid + insert—lots more colonies than #3

Identify recombinant plasmids

- Screen colonies/plasmids for cloned DNA by PCR
- Alpha complementation: white colonies represent presence of insert DNA blocking functional beta galactosidase

Confirm clones by sequencing

When cloning a piece of DNA consider:

- 1) Choice of plasmid: cloning sites? antibiotic? replicon?
- 2) Adding DNA to plasmid: how will the reaction be set up to give the desired product?
- 3) Moving DNA by transformation: what strain of *E. coli* will you transform into? Which method for transformation?
- 4) Screening for successful ligation products (recombinant plasmid DNA): how will the recombinant plasmids be identified?

Basic gene cloning: plasmids and transformation, cutting and pasting

- Plasmid biology
- Adding DNA to a plasmid
- Transformation of *E. coli*

Vectors for *E.coli* part II

- I. Bacteriophage (bacterial viruses): lambda and M13
- II. Moving and storing large DNA molecules: PACs, and BACs

Readings:

- 1) 28 *MC4 Lambda and M13*. Short introduction to these historically important molecular biology tools, also two M13 protocols.
- 2) 33 *MC4 High capacity vectors*. Summary of vector types, with an emphasis on bacterial artificial chromosomes (BACs)
- 3) 29 *MC4 Cre/Lox*. Discussion of the cre-lox system for forcing site-specific recombination

Bacteriophages: viral vectors for molecular cloning

I. M13: a filamentous phage

- Life cycle
- genome structure

II. Lambda: a “head and tail” phage

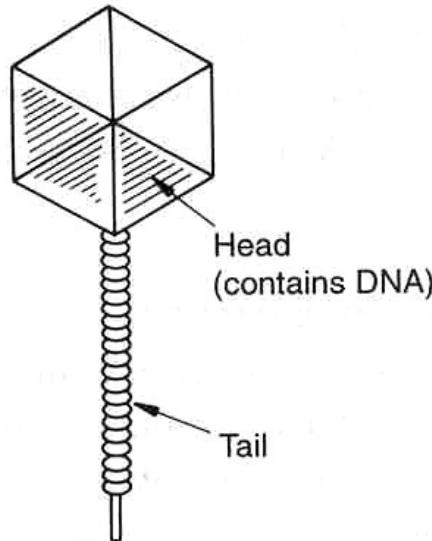
- Life cycle
- Basic cloning in lambda

Bacteriophages

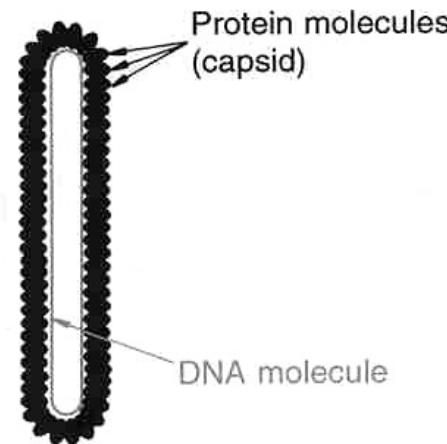
Viruses that infect bacteria

Morphologies:

- a) Head and tail
- b) Filamentous



(a) Head-and-tail



(b) Filamentous

- Nucleic acid molecule
 - Carrying genes for infection and replication
 - Surrounded by a protective protein coat (capsid)
- Infection (instead of transformation):
 - Phage attaches to outside of bacterium, injects DNA
 - Phage DNA is replicated, capsid proteins are made
 - Phage assembled and released

Use of bacteriophages in cloning:

M13 -- single-stranded DNA genome

- Then:
 - Sequencing
 - Site-directed mutagenesis
- Now:
 - Phage display of foreign peptides

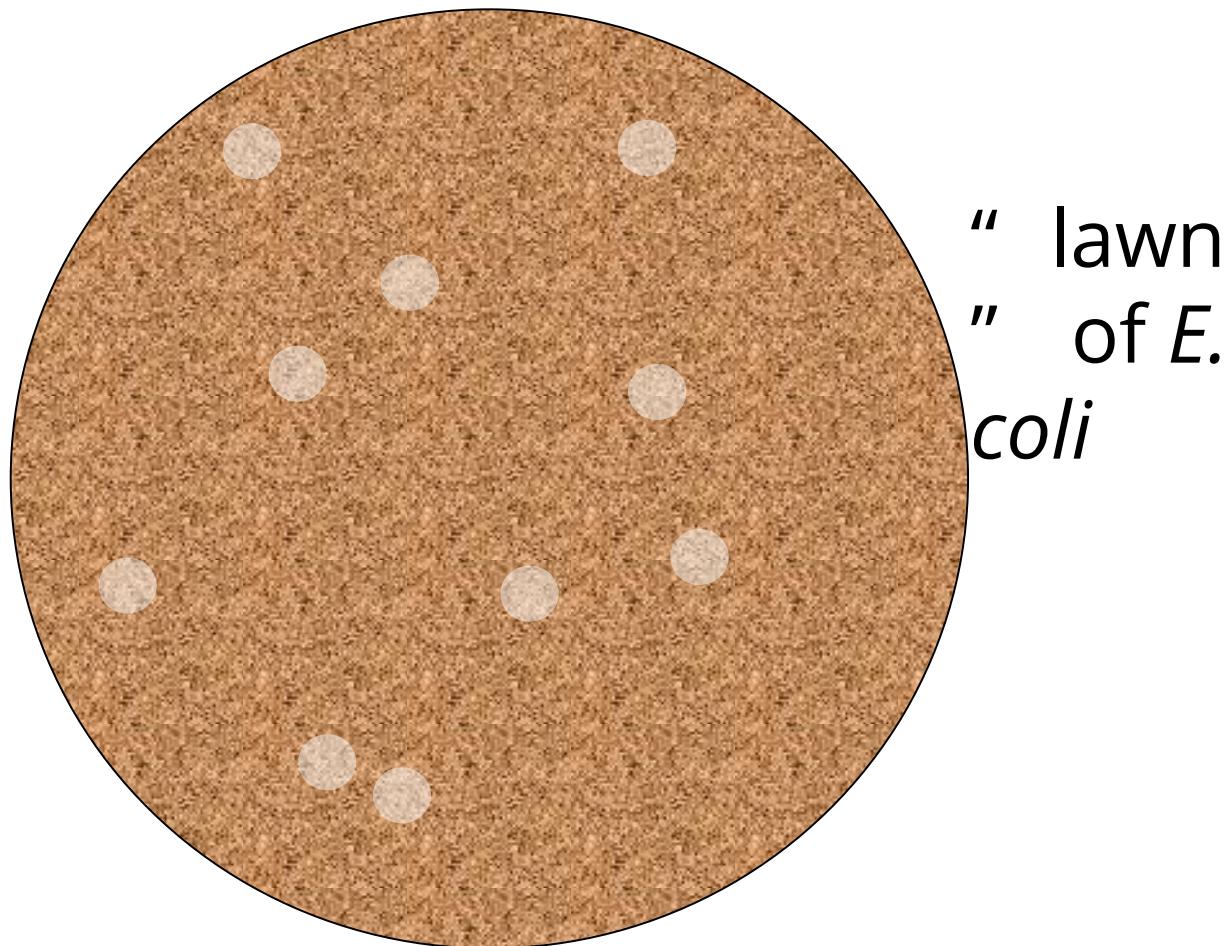
Lambda -- large-ish DNA fragments (25 kilobases)

- For gene cloning (large eukaryotic genes)
- Good selection schemes for recombinants
- Clone lots of precisely-sized DNA fragments for library construction

M13: a filamentous bacteriophage

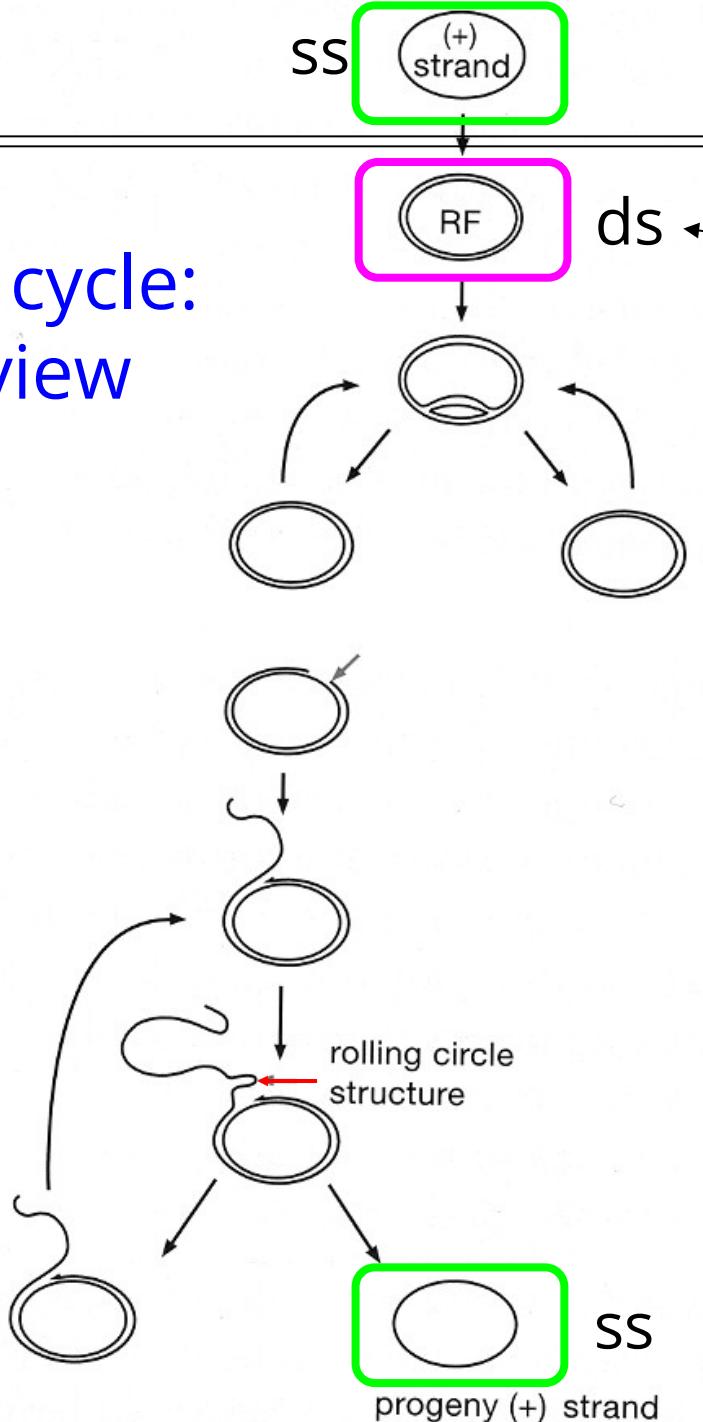
- Single-stranded, circular genome, 6.4 kb
- Can carry foreign DNA up to 6X the M13 genome size (36 kb) -- **the larger the DNA, the less stable the clone**
- Drawback: **foreign DNA can be unstable** (slow host cell growth – deletions confer a selective advantage)

M13 doesn't lyse cells, but it slows them down



M13 infections form 'turbid' plaques

M13 life cycle: an overview



Infecting single-stranded circular viral DNA is converted to double-stranded replicating form (RF) DNA by host-encoded enzymes.

Isolate for cloning

Several rounds of replication occur through θ structures.

The (-) strand of the RF DNA is transcribed into viral mRNAs.

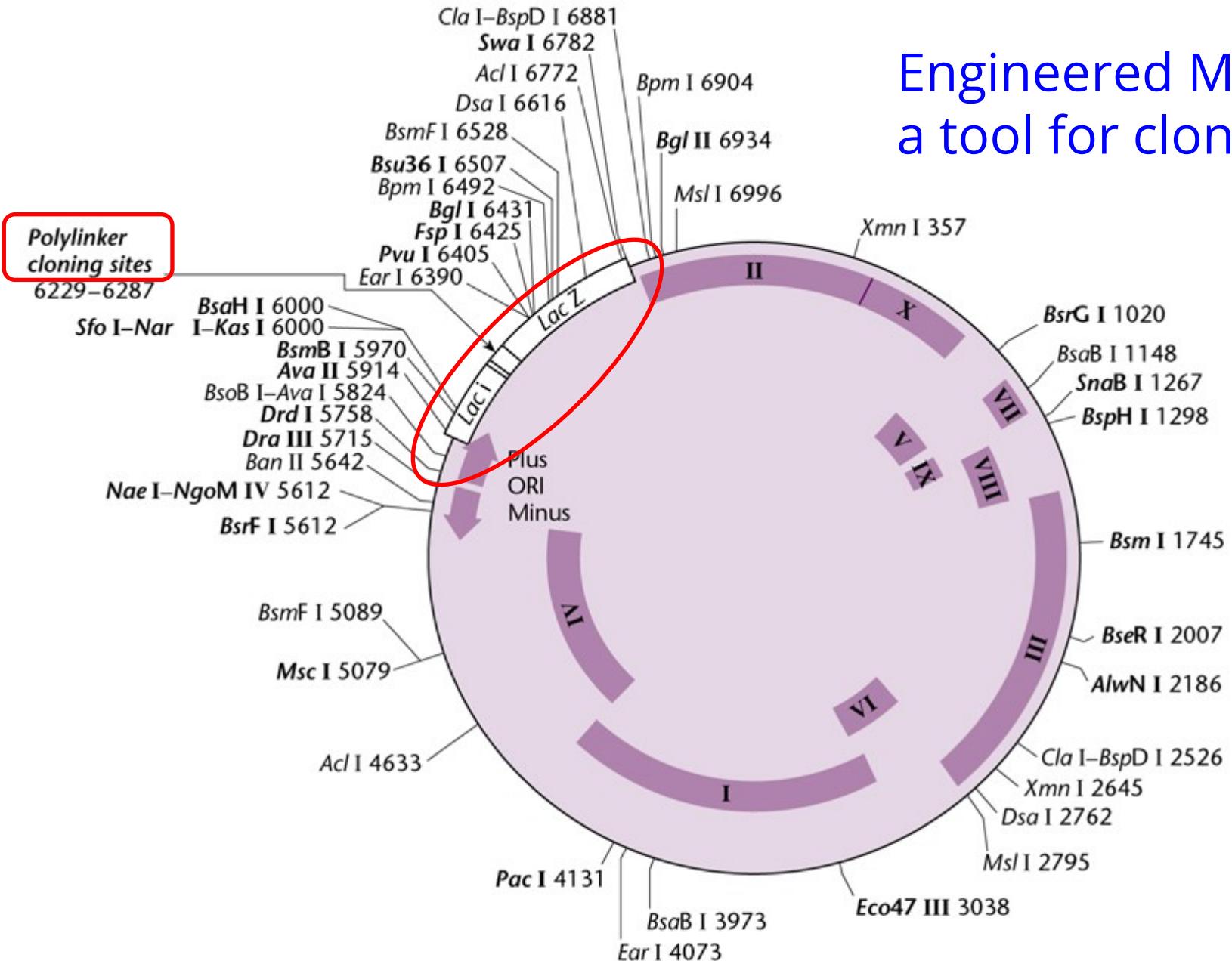
The viral gene II product introduces a nick at a specific site (red arrow) in the (+) strand of the RF DNA.

Progeny (+) strand is synthesized continuously by movement of the replication machinery around the (-) strand template (rolling circle replication).

The completed progeny (+) strand is cleaved from the rolling circular structure by the viral gene II product (red arrow). The progeny strand then circularizes.

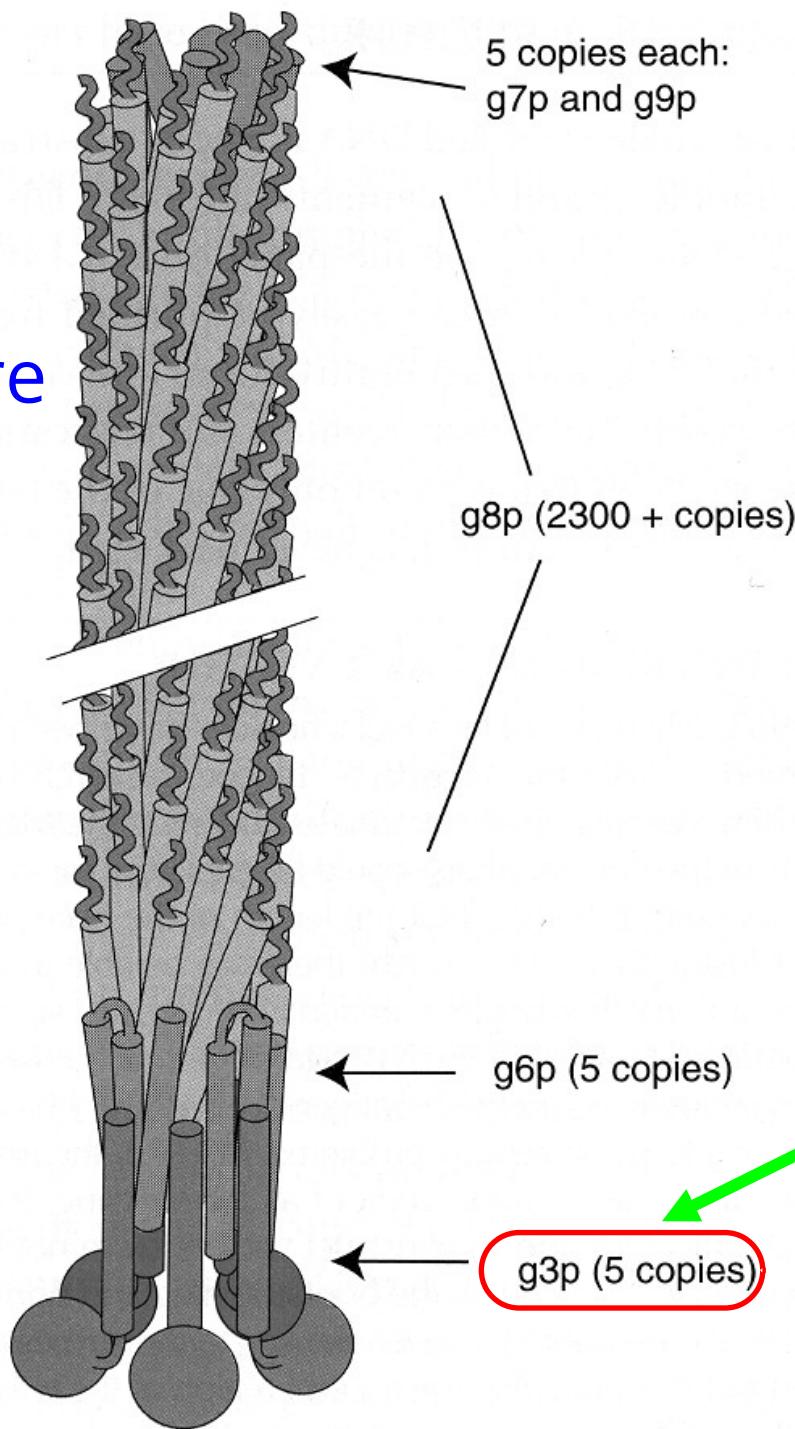
Synthesis of progeny (+) strand continues.

Engineered M13: a tool for cloning



alpha complementation (like pUC plasmids)

M13 phage particle structure



Variable length,
determined by
DNA size

Used in
'phage
display'
techniques

Bacteriophage lambda

- “ head and tail” phage, well-studied model
- Large, linear genome--48.5 kb
 - Central region of genome (“ stuffer”) is dispensable for infectious growth--it can be engineered out
- Two infectious phases:
 - Lytic: replicating phase (killing/lysing cell)
 - Lysogenic: latent phase (integrated, waiting for opportunity)
- Can hold 5-25 kb DNA fragments

Lambda as a cloning vector

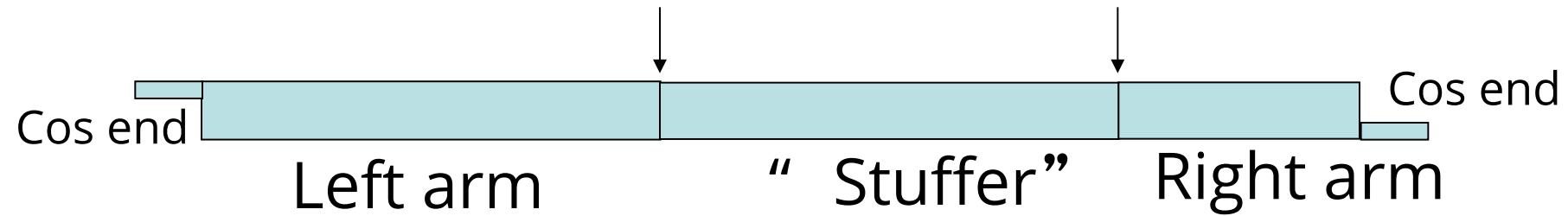
Phage capsid:

- Size of head is invariant
- Capsid must be filled
 - Too little DNA and it's not infectious
 - Too much DNA and the genome won't fit

a) Insertional vectors: can only increase genome size by 5% (about 3 to 5 kb)

b) Replacement vectors: remove “ stuffer” , can clone larger pieces of DNA, 8 to 24 kb (sufficient for many eukaryotic genes)

Cloning in lambda phage – the genome view



- 1) Cut out stuffer, save right and left arms
- 2) Ligate with foreign DNA
- 3) “ Package” ligation mixture into phage heads
- 4) Plate mixture on *E. coli*, individual plaques represent recombinant clones

Selection for *recombinant* lambda

There is a minimal size of DNA that can be packaged in lambda phage heads

- If the stuffer is removed the ligated “ arms” cannot be packaged (too small)
- Presence of recombinant DNA makes the phage large enough, so the recombinants are selected

Cloning large DNA fragments

- 1) Make genomic *libraries*: the larger the DNA fragment, the fewer you need to make a complete *library* (representing the entire genome in fragments)
- 2) Clone DNA large enough to contain an entire eukaryotic gene

To get a functional gene, you need the whole gene and its regulatory regions

- Average human gene (exons and introns): 27 kilobases
- Regulatory regions may add another 10 to 100 kilobases upstream/downstream of gene

Vectors for large recombinant DNA fragments

- Bacteriophage P1 plasmid: 70-100 kb
- YAC: 250-400 kb (or higher)
- PAC: 130-150 kb
- BAC: 120-300 kb (up to 700 kb is possible)

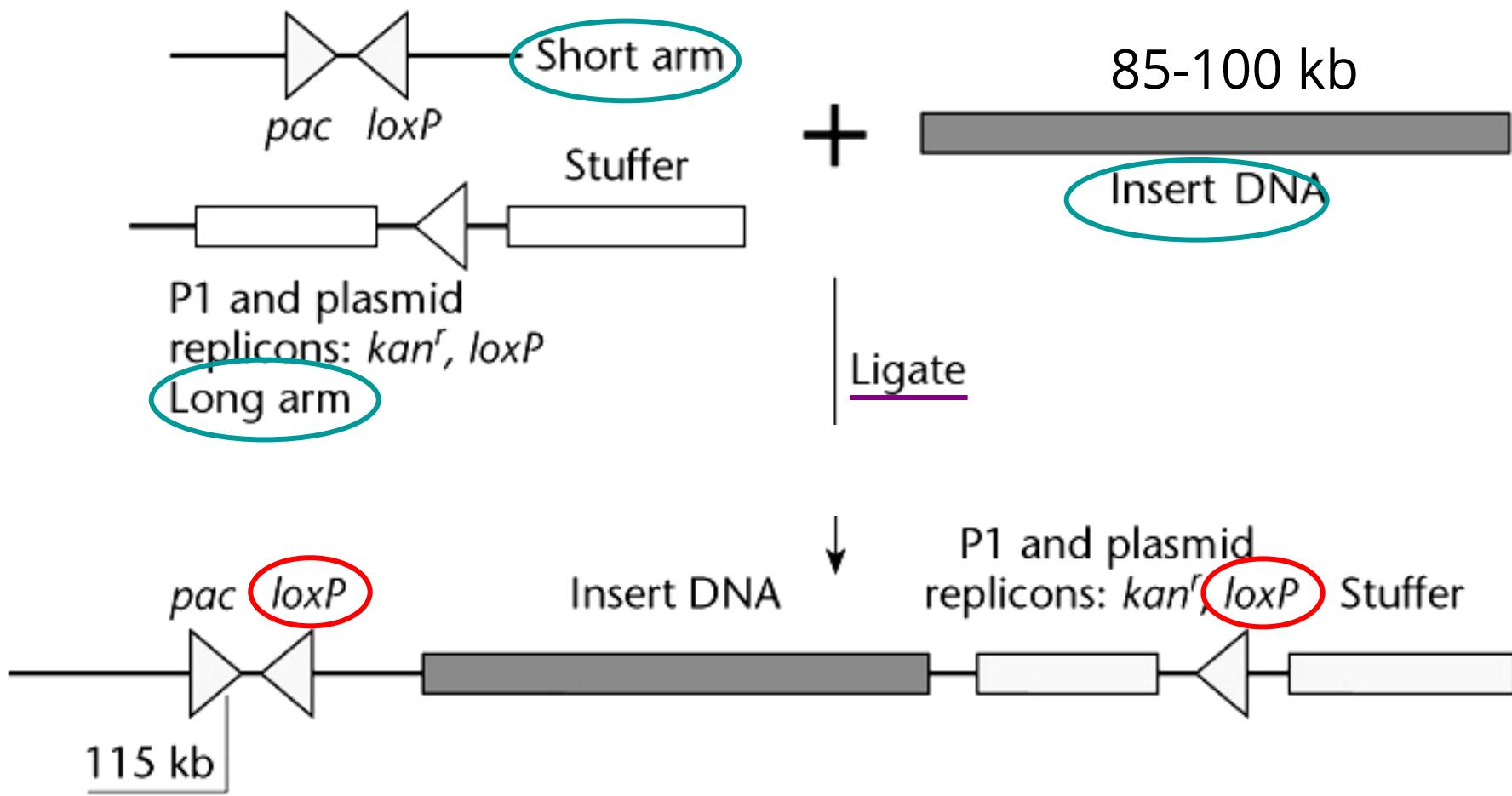
TABLE 1. High-capacity vectors for genomic cloning

Vector	Capacity (kb)	Replicon	Host	Copy number	Recovery of cloned DNA
P1	70–100	P1	<i>Escherichia coli</i>	1 (amplifiable)	Alkaline extraction
YAC	250–400	ARS	Yeast	1	Pulse-field gels
PAC	130–150	P1	<i>E. coli</i>	1	Alkaline extraction
BAC	120–300	F	<i>E. coli</i>	1	Alkaline extraction

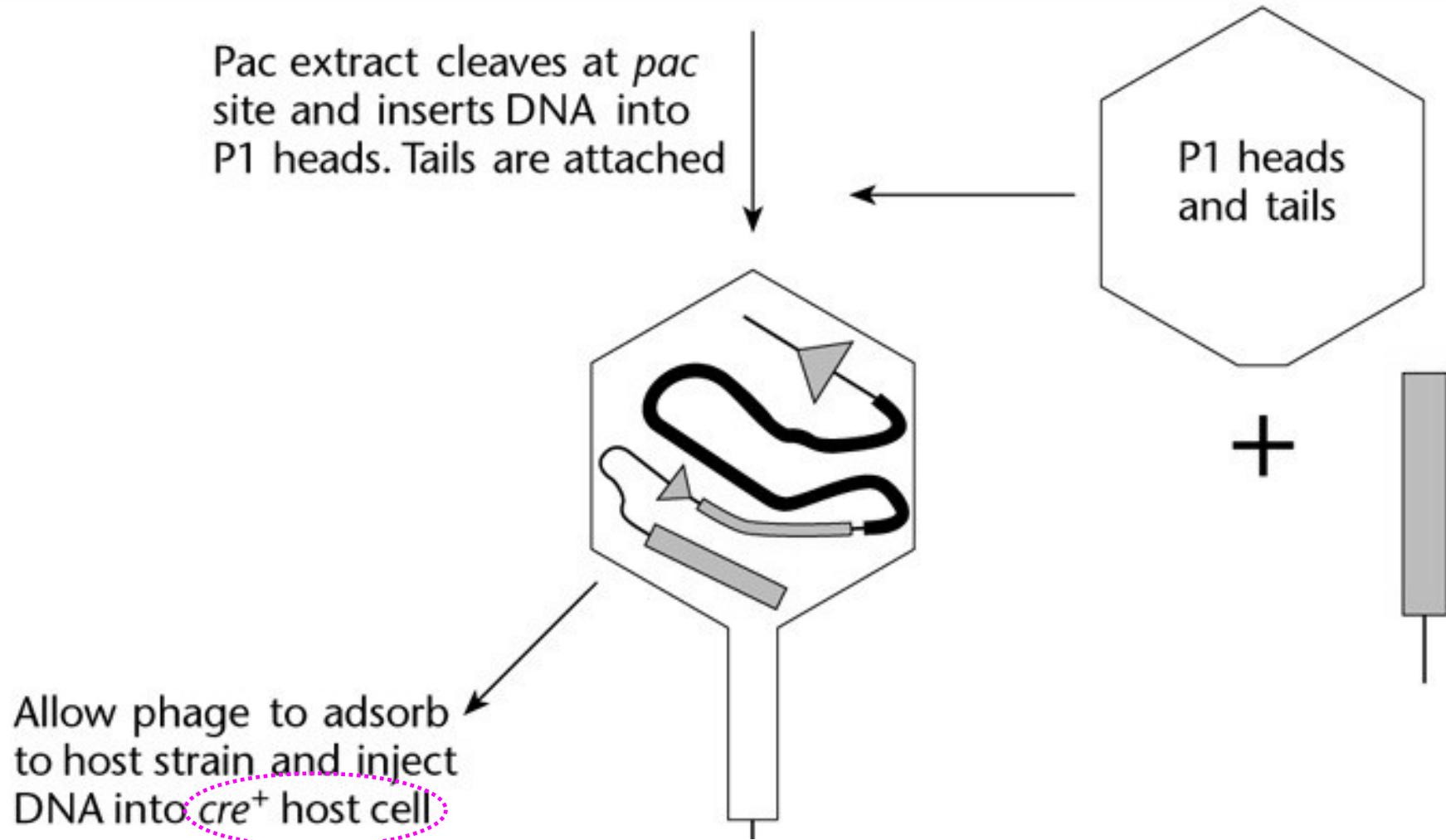
BACs are most commonly used

Phage P1 vectors:

clone large DNA fragments (85-100 kb)

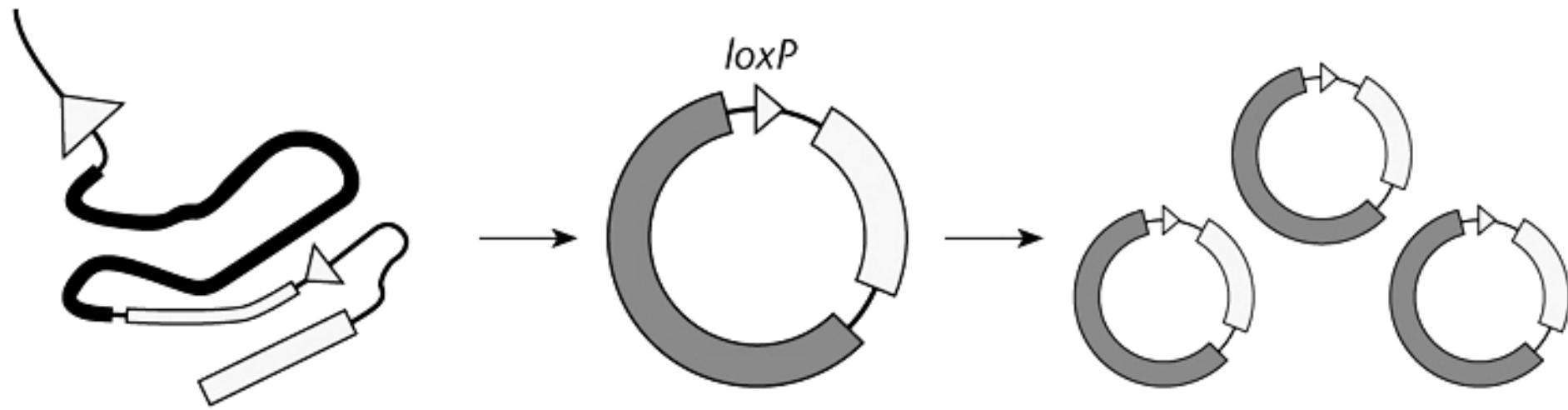


Phage P1 vectors



Efficiency of packaging is typically low: thus it is not good for making large genomic libraries

Phage P1 vectors

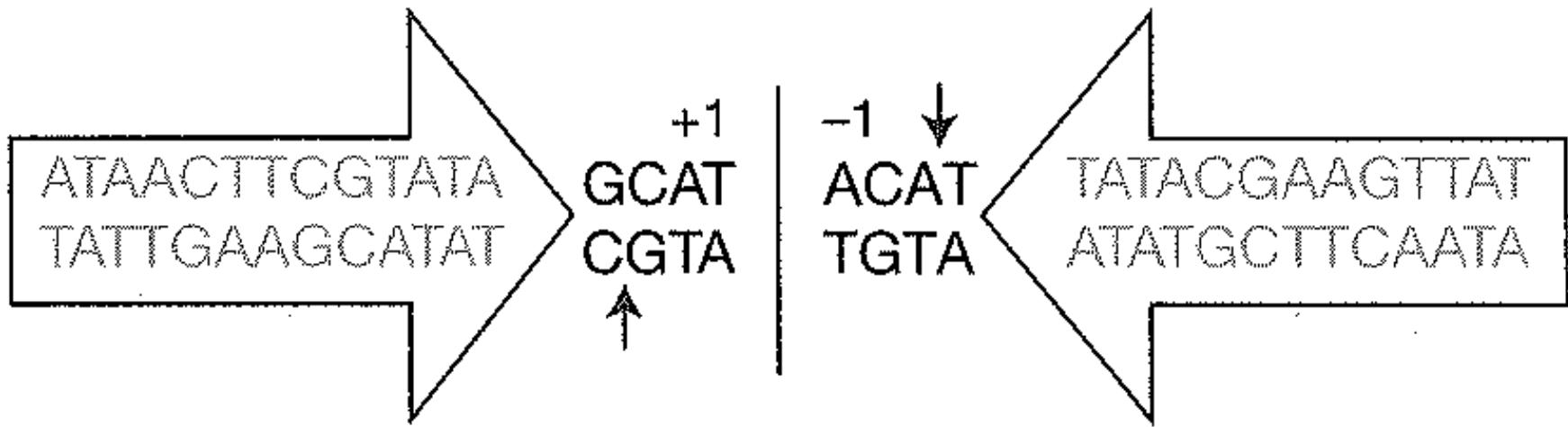


Cre recombinase protein circularizes injected DNA at the *loxP* sites. DNA replicates using plasmid replicon. Plasmid copy number is increased by induction of P1 lytic replicon.

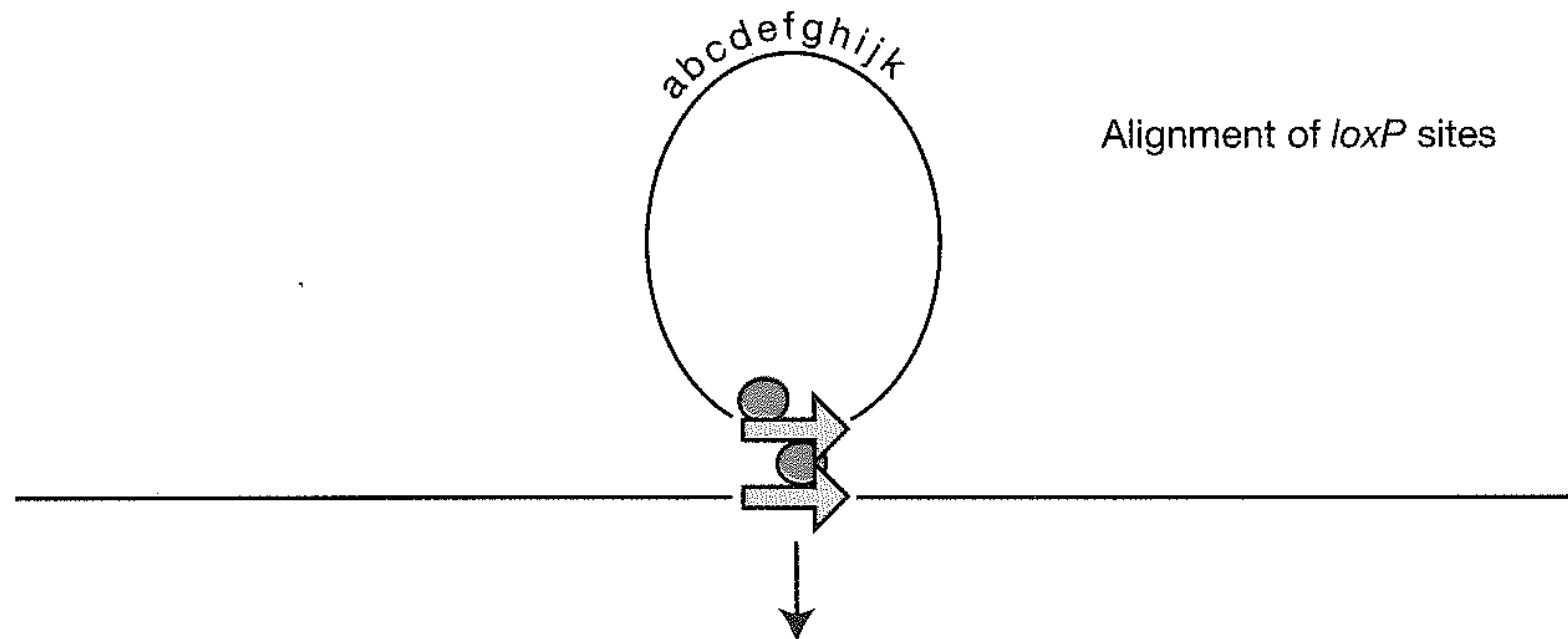
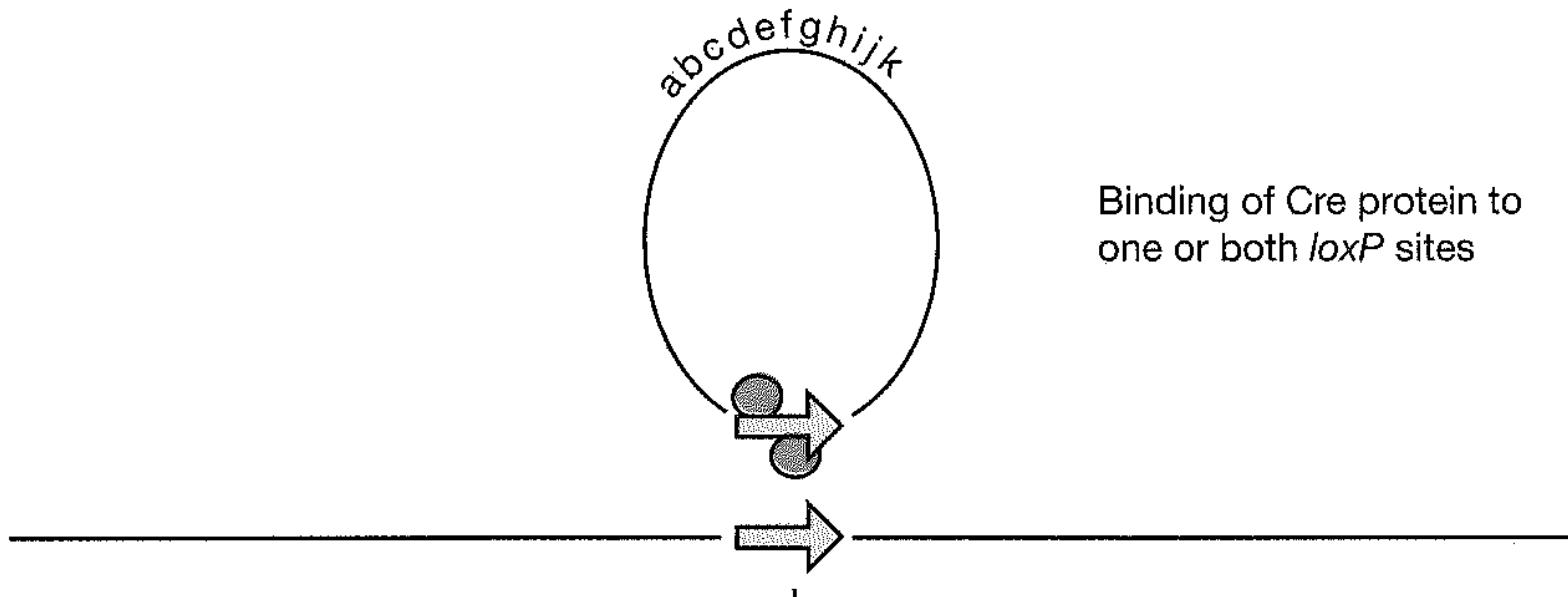
PACs: like P1 vectors but the DNA is not packaged (transfer by electroporation)

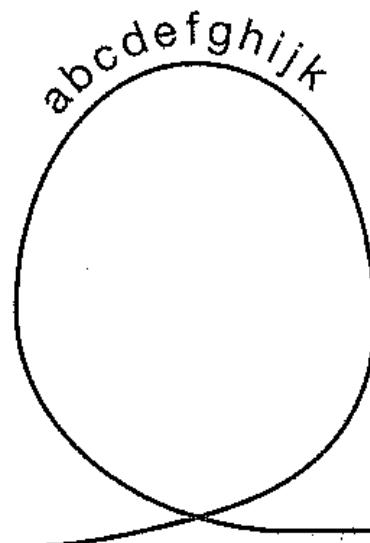
Cre and Lox: a site-specific recombination system

The loxP site



Two 13 bp inverted repeats separated by an asymmetric linker



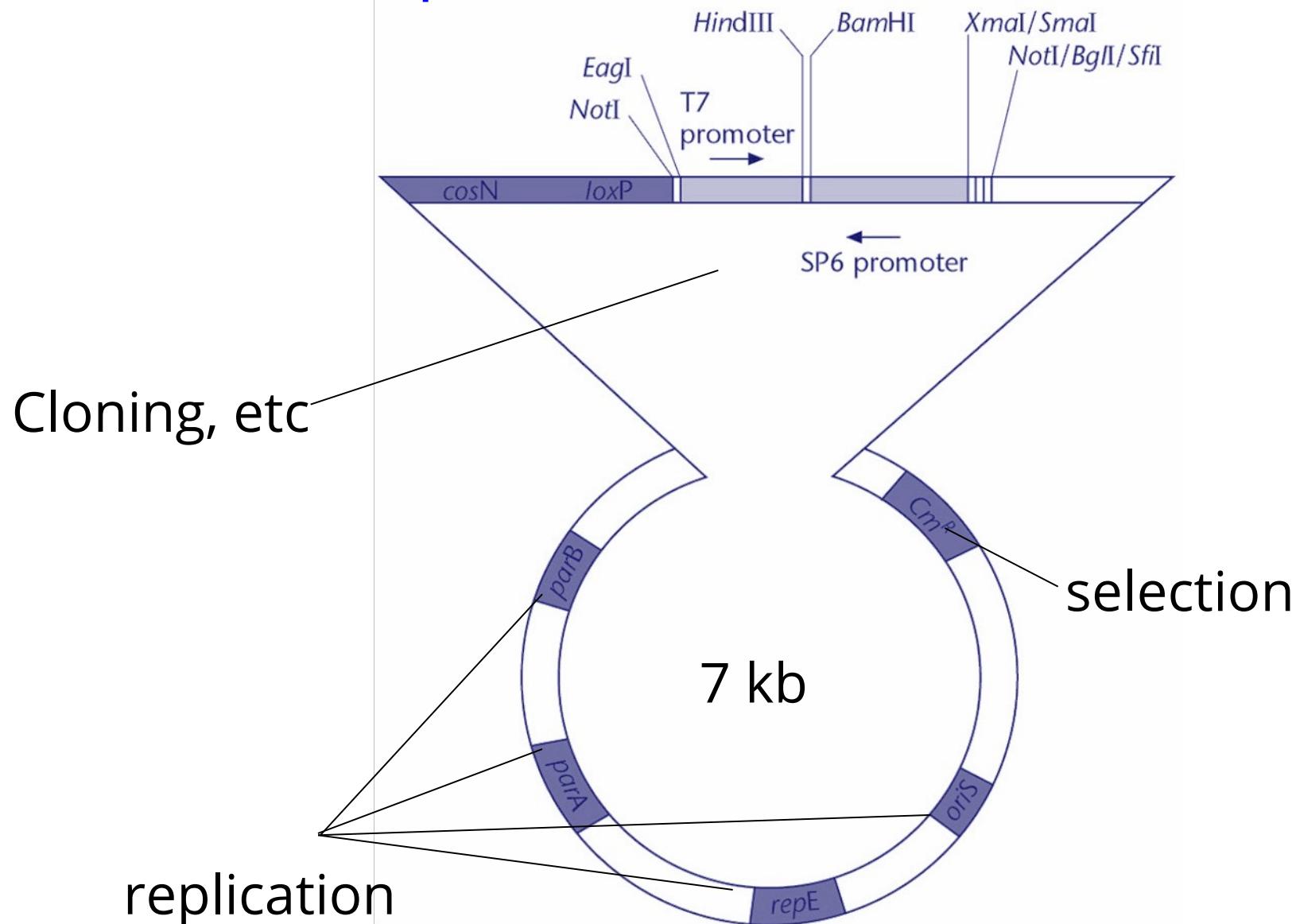


Asymmetric cleavage
of *loxP* sites followed by
strand invasion and synapse

BAC: Bacterial Artificial Chromosome

- F factor of E. coli:
 - 100 kb plasmid, propagates through conjugation
 - low copy number (1-2 copies per cell)
 - 2 genes (parA and parB): accurate partitioning during cell division
- BACs: engineered to have par genes, replication ori, cloning sites, selectable marker
- Holds very large pieces of DNA: **up to 300 kb**
- Fairly easy to manipulate: move into cells by transformation (electroporation)

BAC vector map



Vectors for *E.coli* part II

- I. Bacteriophage lambda and M13
- II. Moving and storing large DNA molecules: PACs and BACs

Gene cloning: Expression of protein and RNA

Controlled gene expression

I. proteins

- o Various tags and their functions
- o A protocol for purification of 6xHis tagged proteins expressed in *E. coli*

II. RNA

- o For in vitro uses
- o For RNA interference (RNAi) in vivo

Guide to readings:

- 1) 30 *MC4 Protein expression*. Intro to protein expression in bacterial and eukaryotic systems, vectors, optimization.
- 2) 32 *MC4 Epitope tags*. A list and description of tags/handles that can be added to recombinant proteins.
- 3) *Protein expression tags Malhotra 2009*. More good info on tags.
- 4) 31 *MC4 His tag protein purification*. A detailed protocol for production of 6 x His tagged proteins.
- 5) *RNAi delivery methods 2015*. Trends in RNAi –based therapies
- 6) *SARS 1 vaccine 2009*. Using spike protein for immune response
- 7) *Hotez and Bottazzi 2020*. Production of affordable SARS CoV2 protein antigen vaccine.

Controlled protein or RNA production is important

1) Production of proteins & RNA:

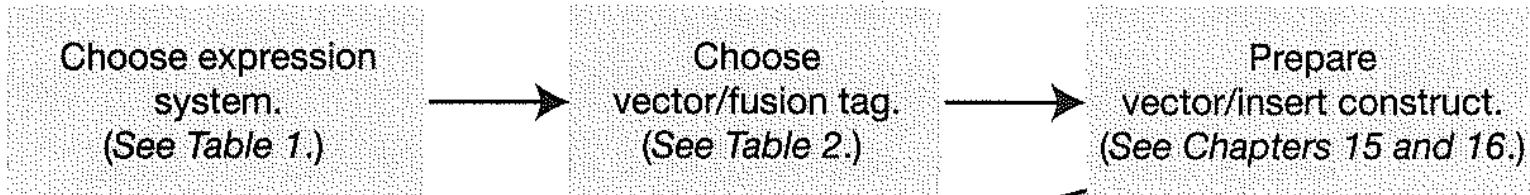
- Biochemical studies (enzyme function)
- Structural studies (crystallography, NMR)
- Commercially useful proteins or RNAs
- Protein pharmaceuticals
- Induction of RNA interference (RNAi)

2) Living systems

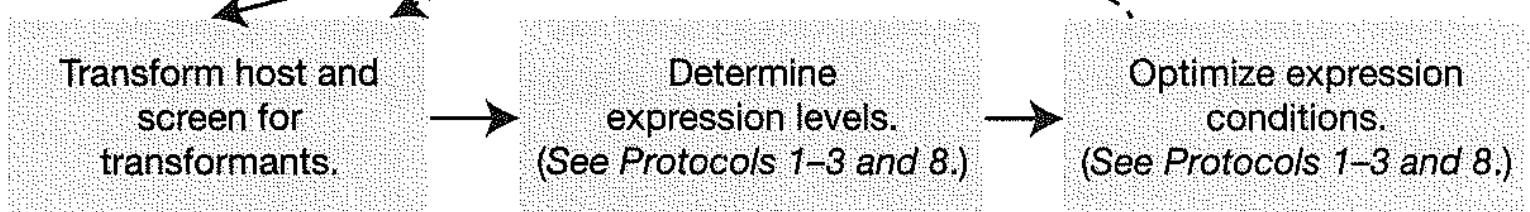
- Cell biology -- protein localization
- Transgenic plants and animals
- Human gene therapy

Strategy for protein “overexpression”

(1) Cloning



(2) Expression



(3) Purification

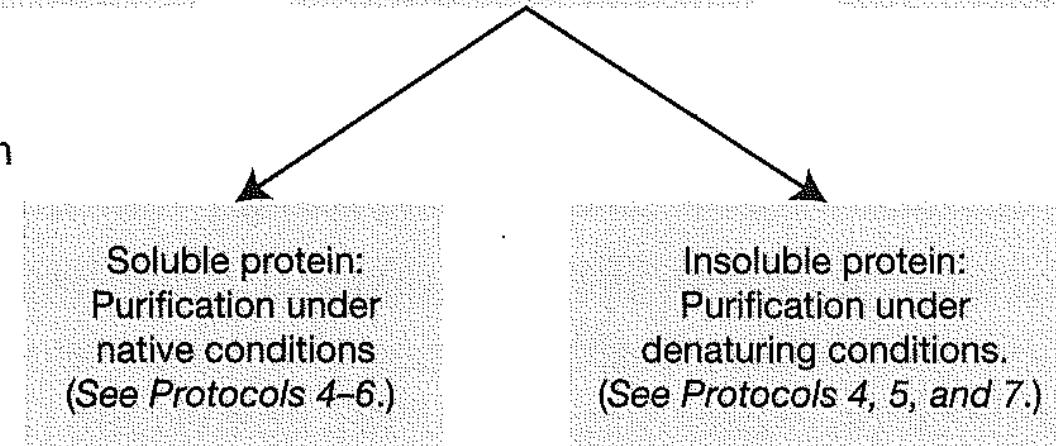


FIGURE 1. Overall strategy for recombinant protein expression and purification.

Expression vectors for protein/RNA purification

- Make lots of protein or RNA
 - Control timing of synthesis
 - Control amount of synthesis
-
- Optimized for transcription, translation
 - Facilitate post-translational handling by adding “tags”

The most common expression platform is bacterial, specifically *E. coli*

- Lots of vectors available
- Lots of host *E. coli* strains available
- Large variety of tags for proteins
- Straightforward, easily adaptable protocols
 - Add gene to expression plasmid
 - Transform host strain with expression plasmid
 - Grow cells to mid-log (exponential phase)
 - Induce protein expression
 - Purify the protein

Alternative expression hosts sometimes needed

Eukaryotic protein expression in bacteria not always successful:

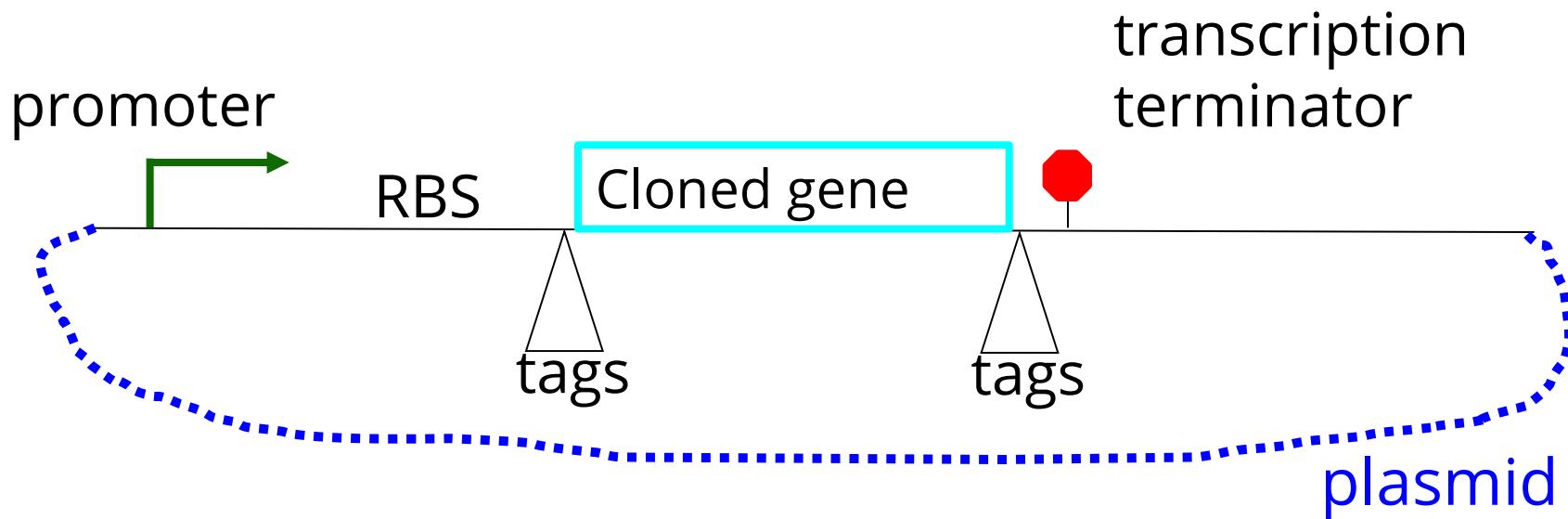
- Some proteins do not fold properly
- May lack post-translational modifications required for activity, especially glycosylation
- Protein might be toxic to bacteria

Some alternatives to bacterial expression:

- 1) Baculovirus-infected insect cell culture
- 2) *Pichia pastoris*, an easily cultured yeast
- 3) Mammalian cell culture
- 4) *Cell-free* translation system

See Table 1 in *30MC4 Protein Expression* for comparison of these methods

A bacterial protein expression vector



- 1) Promoter for transcription
- 2) Transcription terminator
- 3) Ribosome binding site (RBS) for translation
- 4) Epitope tags for purification or detection of the protein

To maximize synthesis of gene product:

- 1) Bring cells to mid-log phase (all cells in the process of growth and division) without expressing the gene: cell physiology is optimized
- 2) When conditions are right, “induce” the gene, by turning on transcription/mRNA synthesis
 - lots of ribosomes available
 - protein synthetic machinery is at a maximum to cope with new mRNA for induced gene
 - Ideally, massive amount of new protein will be made

Inducible promoters for bacterial gene expression

Promoter	Transcription regulation	Induction
Lac	Lac repressor	IPTG (lactose analogue)
Trp/lac	Lac repressor	IPTG
T7	Lac repressor, T7 RNA polymerase	IPTG
pBAD	AraC repressor	arabinose
Lambda P _L	Temperature sensitive lambda repressor	Shift from 30°C to 37°C

Fusion protein tags: protein purification/detection

- Tags can be added to the N- or C-terminus, or can be internal to the protein sequence (on a surface loop)
- Tags can be engineered to be removable
- More than one tag can be added to a protein, for “Tandem Affinity Purification”
- Peptide tags. Example: **FLAG tag**, detected by monoclonal antibodies
- Whole proteins. **Biotin carboxylase**: covalently attaches to biotin, biotin binds to streptavidin which can be immobilized on columns or beads

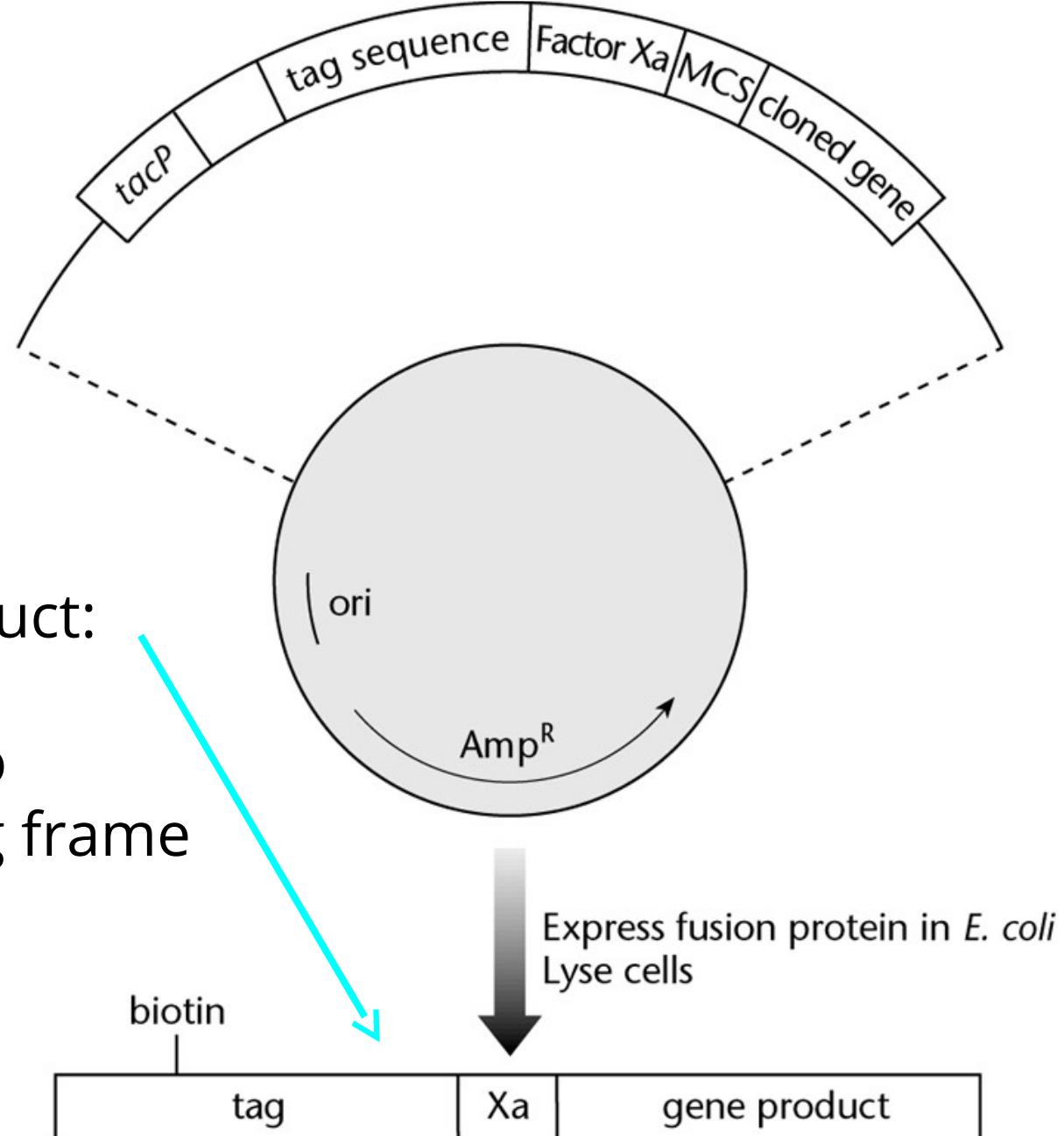
Epitope tags as purification handles

Protein tag	Size	What the tag sticks to	What causes release
Polyhistidine (6-10xHis)	6-10 a.a.	Immobilized Ni, Co, Cu, Zn ions	Imidazole >100 mM
Glutathione S-transferase (GST)	211 a.a.	Immobilized glutathione	Reduced glutathione
FLAG-tag: DYKDDDDK	8 a.a.	Anti-FLAG antibody	FLAG peptide or low pH
Protein A	280 a.a.	Any immunoglobulin G	Protease
Calmodulin binding protein	26 a.a.	Immobilized calmodulin	EGTA 2mM
Acceptor peptide (biotin added by biotin ligase)	15 a.a.	Avidin/streptavidin	Biotin

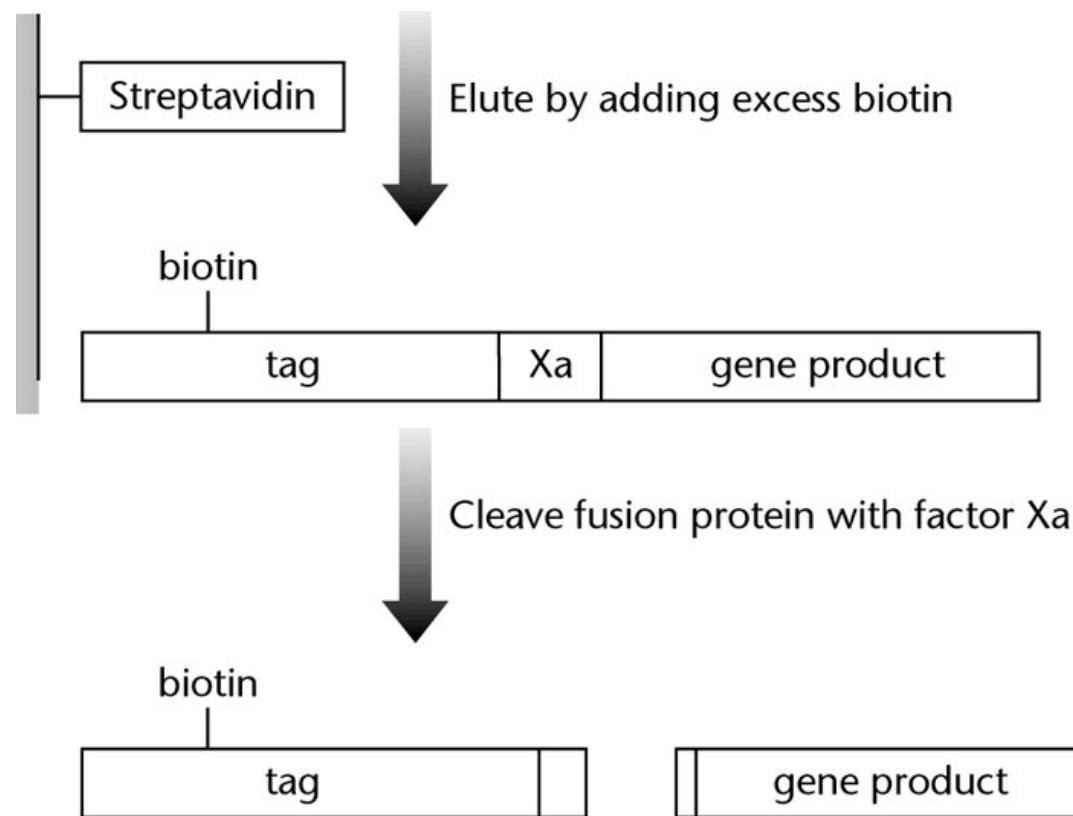
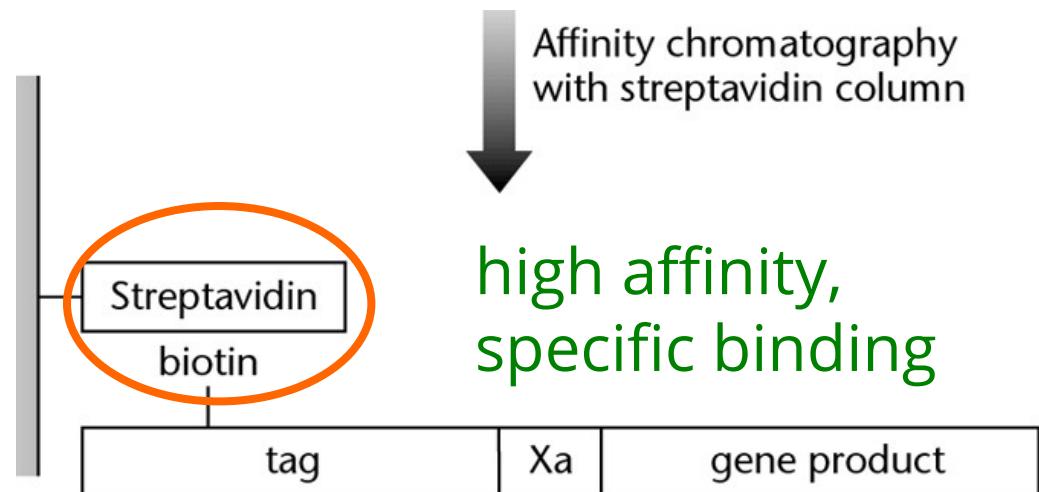
Tags for protein folding/solubility

Protein tag	Size	What the tag does	Additional facts
SUMO (small ubiquitin-like modifier)	98 a.a.	Improve protein folding/solubility	Can be removed by SUMO protease
MBP: maltose binding protein	396 a.a.	Helps folding/solubility AND sticks to cross-linked amylose for purification	Released from amylose by adding maltose
Trx: thioredoxin	109 a.a.	Improve protein folding/solubility	Extremely soluble, even at 40% of total cell protein, prevents inclusion bodies

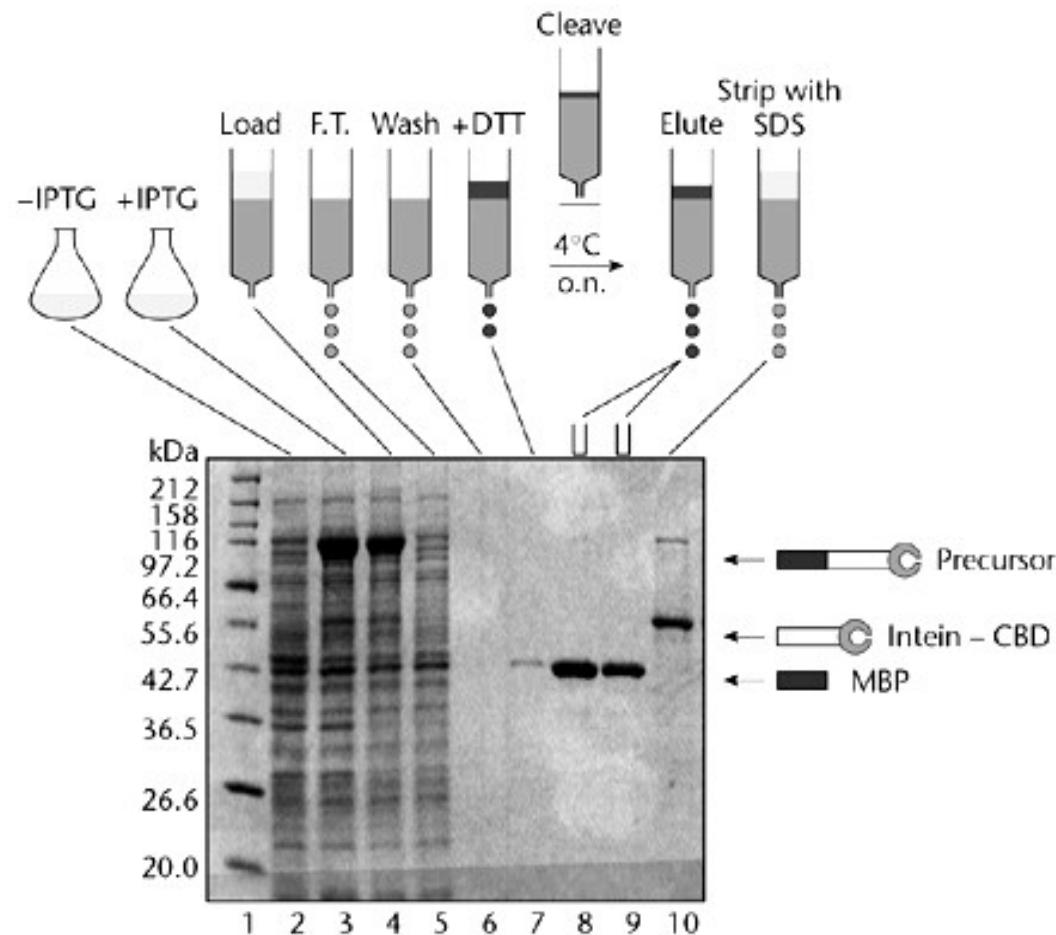
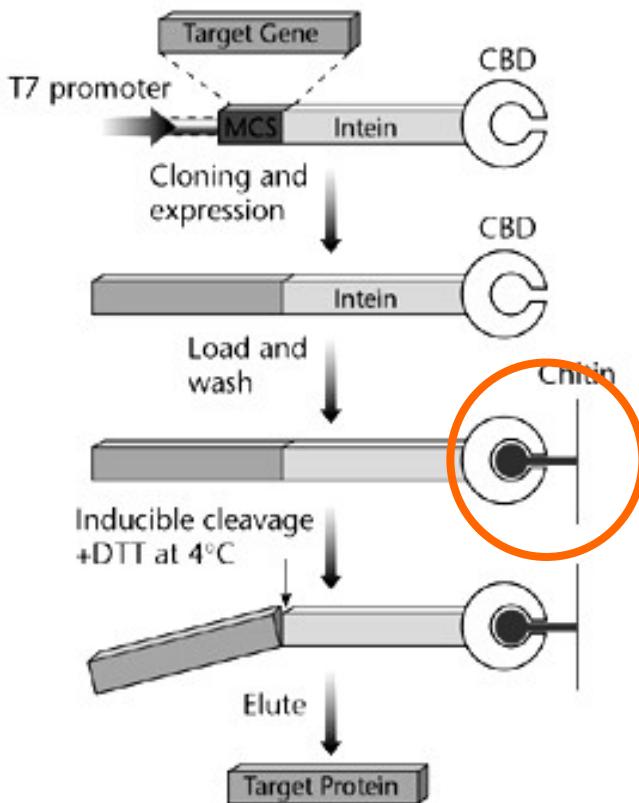
Biotin tag for protein purification



Biotin tag for protein purification



Another protein purification scheme--removable tag



Lane 1: Protein Marker.

Lane 2: Crude extract from uninduced cells.

Lane 3: Crude extract from cells, induced at 15°C for 16 hours.

Lane 4: Clarified crude extract from induced cells.

Lane 5: Chitin column flow through (F.T.).

Lane 6: Chitin column wash.

Lane 7: Quick DTT wash to distribute DTT evenly throughout the chitin column.

Lanes 8-9: Fraction of eluted MBP after stopping column flow and inducing a self-cleavage reaction at 4°C overnight.

Lane 10: SDS stripping of remaining proteins bound to chitin column (mostly the cleaved intein-CBD fusion).

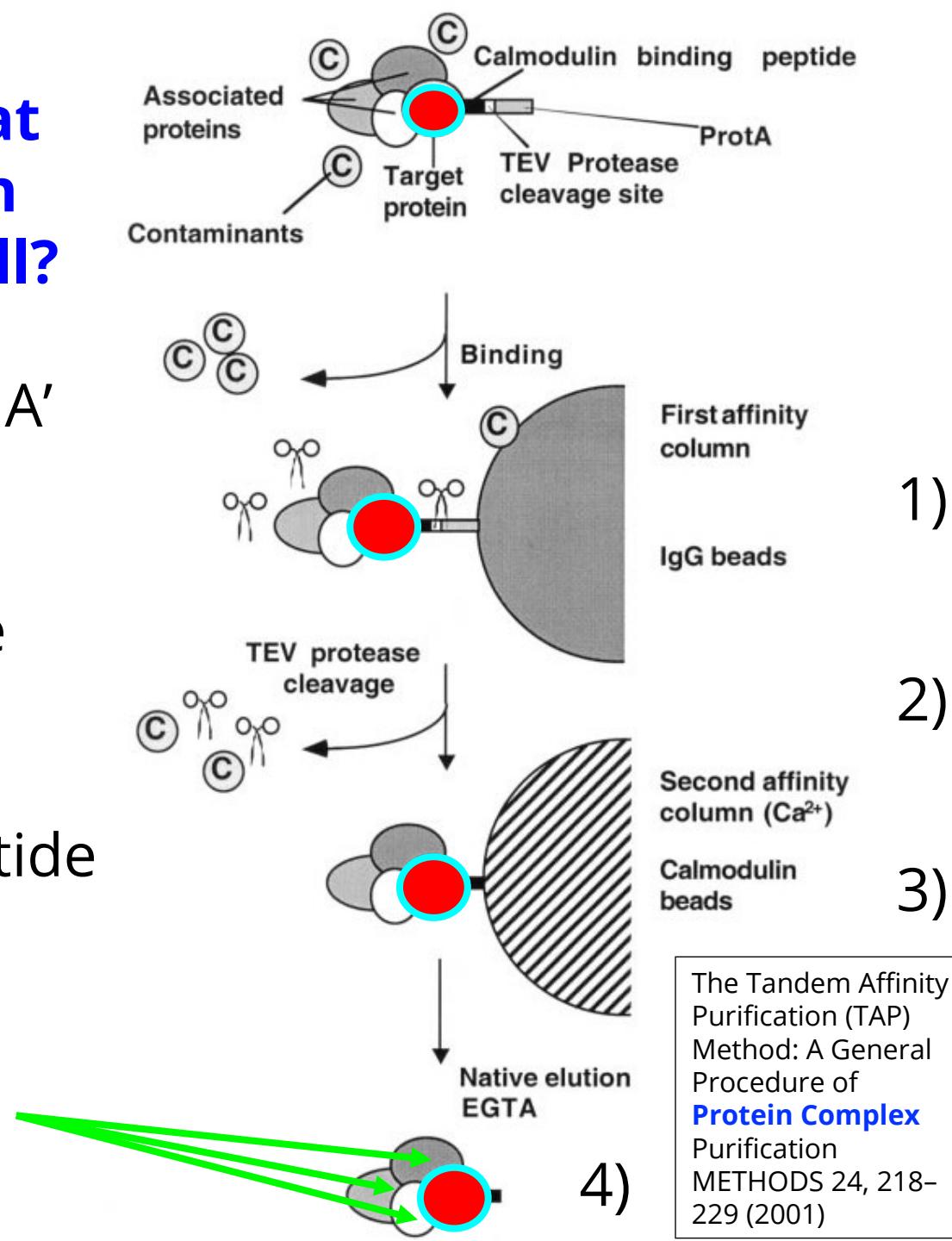
Tandem affinity purification (TAP): what does my target protein interact with in the cell?

1) IgG beads for 'protein A' tag

2) TEV protease cleavage

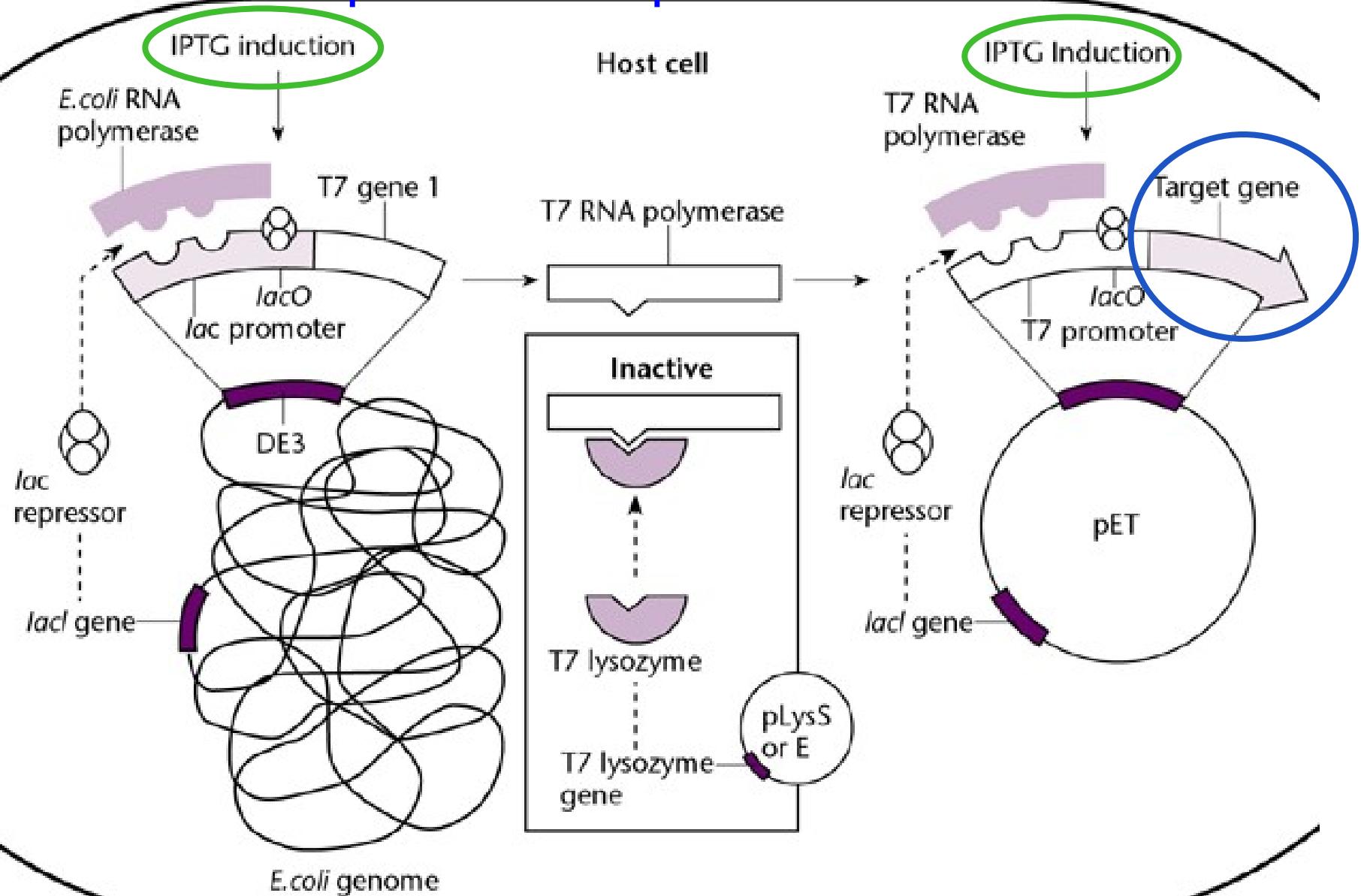
3) Calmodulin beads for calmodulin binding peptide

4) Pure protein and associated complex



The Tandem Affinity Purification (TAP) Method: A General Procedure of Protein Complex Purification
METHODS 24, 218–229 (2001)

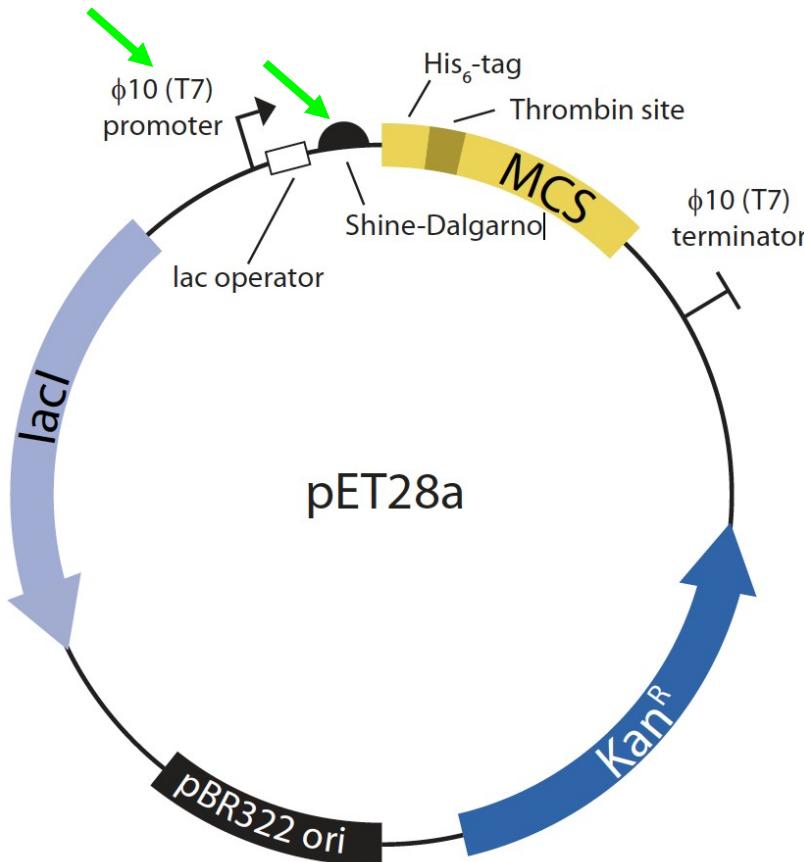
pET vectors: protein expression



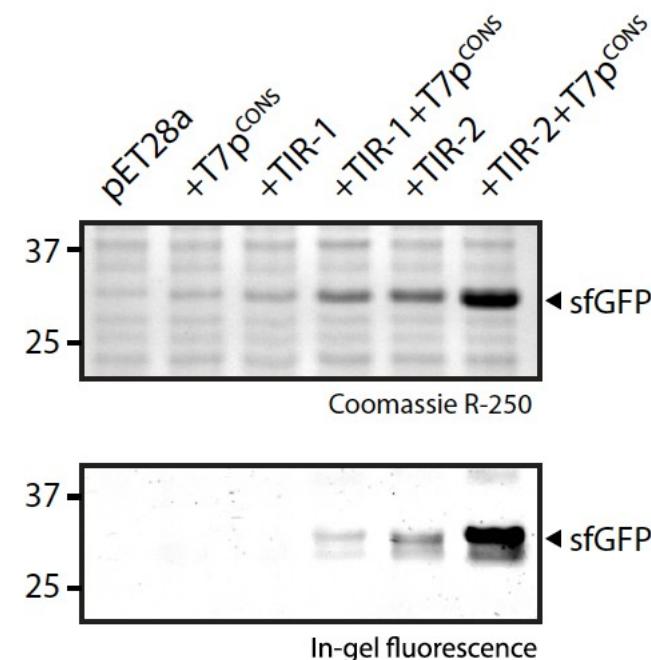
+IPTG = high T7 RNAP, high target gene expression

Updating pET vectors: new developments in 2020

How to get even more protein expression



- 1) Change T7 promoter to the $\phi 10$ consensus sequence
- 2) Alter translation initiation region (TIR) by searching through library of TIR mutants



Improved designs for pET expression plasmids increase protein production yield in *Escherichia coli*

Patrick J. Shilling¹✉, Kiavash Mirzadeh^{1,2}, Alister J. Cumming¹, Magnus Widesheim¹, Zoe Köck^{1,3} & Daniel O. Daley¹✉

COMMUNICATIONS BIOLOGY | (2020)3:214 | https://doi.org/10.1038/s42003-020-0939-8 | www.nature.com/commsbio

Purification of a 6xHis tagged protein, start to finish

- 1) Clone gene into pET vectors
 - Design primers, PCR
 - Choose vector, ligate gene in frame to vector
 - Sequence to verify the clone
 - Transform the plasmid into BL21 strain
 - T7 RNA polymerase gene
 - Lacks proteases
- 2) Overexpress the protein
 - Grow the cells to mid-log phase
 - Add IPTG to induce overexpression of the gene
 - Induction temperature may vary (15 – 42°C)
 - Concentration of IPTG varies (0.01 – 2.0 mM)
 - Time of induction varies (1 hour to overnight)

Purification of a 6xHis tagged protein, start to finish

3) Break open the cells

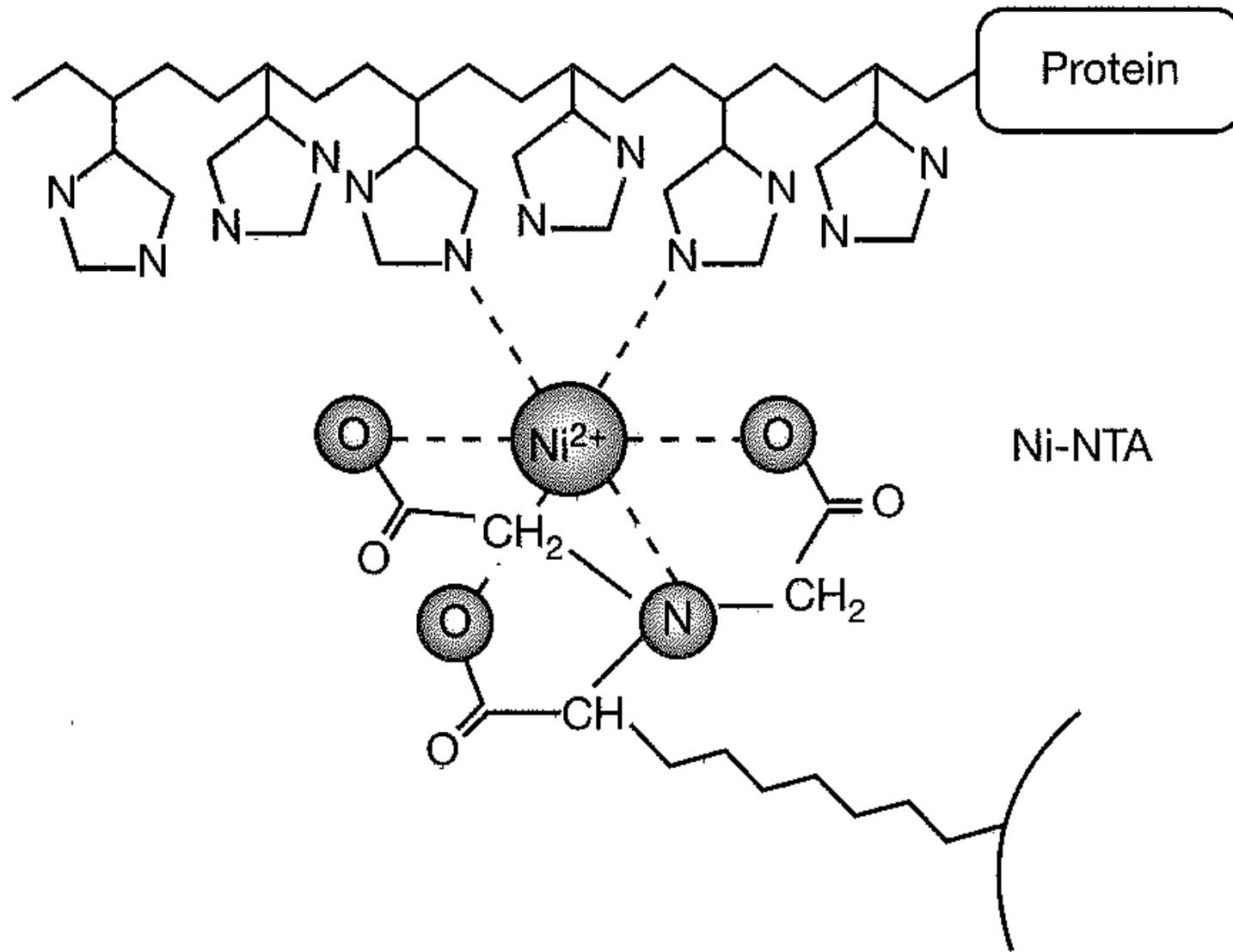
- lysozyme, sonication
- add protease inhibitors
- spin to get a clarified “lysate”

4) Affinity purification of protein

- apply lysate to a Nickel column or to Nickel beads
- wash with low concentration imidazole (20 mM)
- elute with high concentration imidazole (200 mM)

5) Check protein purity and concentration by SDS-PAGE

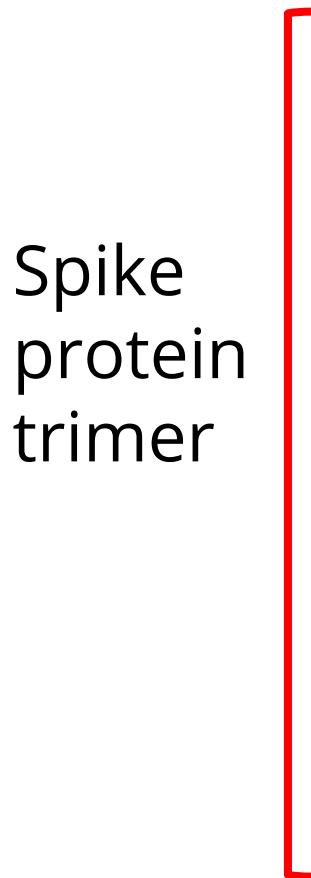
6xHis interaction with Ni⁺⁺



Protein purification and vaccines: SARS CoV-1 in 2009

- o Many different vaccines for SARS Cov-2 are in development now, for example:
 - Inactivated SARS CoV2 virus (Sinopharm, CoronaVac)
 - Vector-driven protein expression (Sputnik V)
 - Nucleic acid (RNA or DNA) vaccine (Moderna, Pfizer/BioNTech)
 - Antigenic proteins (especially the spike **protein**)
- o Production of antigenic proteins?
 - Example: SARS CoV-1, outbreak in 2003, thousands of infections, hundreds of deaths
 - No new cases since 2004
 - Can spike protein epitope (including only the receptor binding domain) confer protective immunity?
 - Du et al. 2009, doi:10.1016/j.virol.2009.07.018

Structure of SARS CoV-1 spike with ACE2 receptor



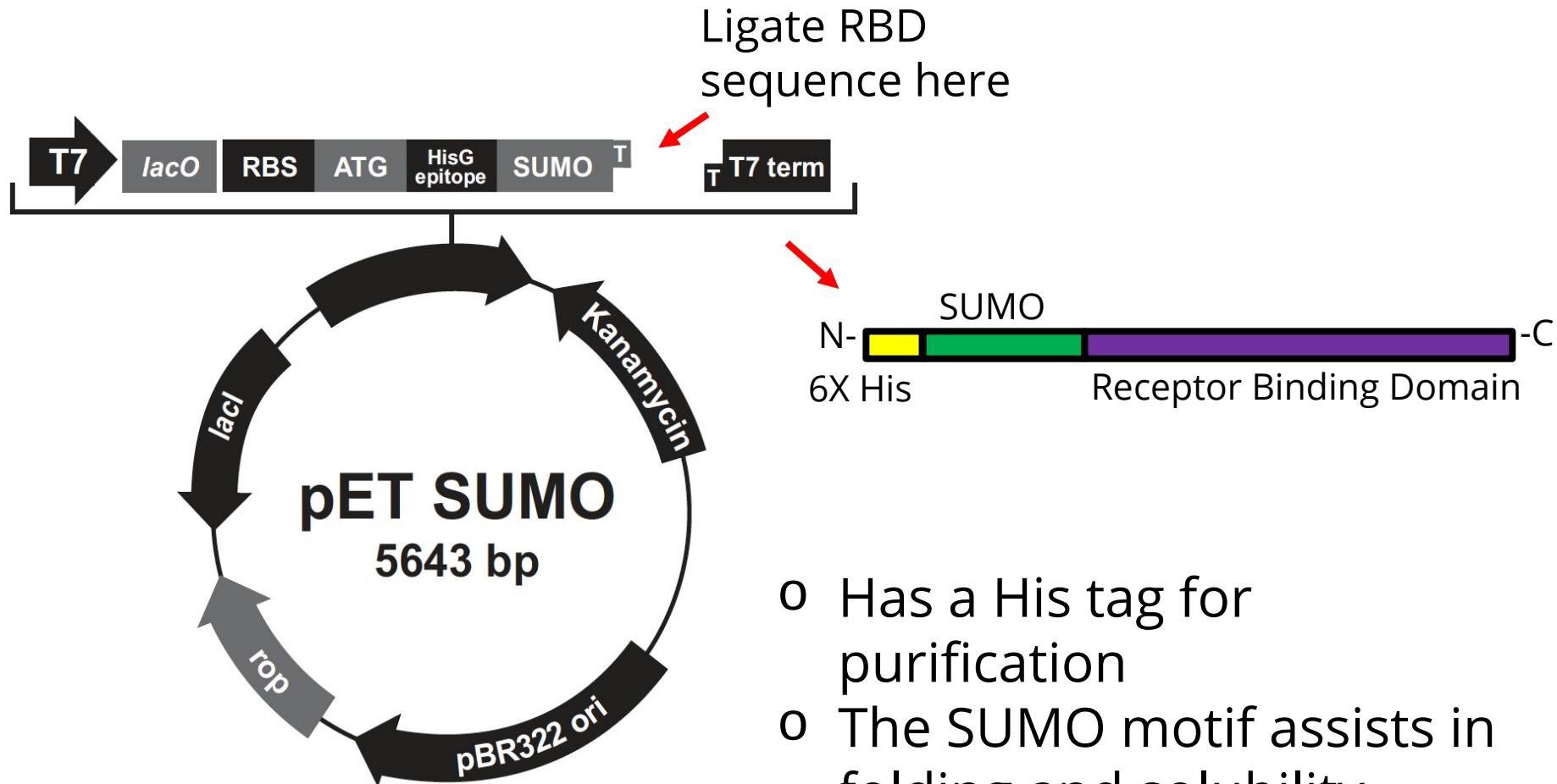
← Human ACE2 receptor

← RBD: receptor
binding domain, a.a.
318-510

PDB ID
6ACG

Can the RBD epitope
by itself elicit a
protective immune
response?

Making the RBD: protein expression in *E. coli*



- o Has a His tag for purification
- o The SUMO motif assists in folding and solubility
- o SUMO motif gets cut away by SUMO protease

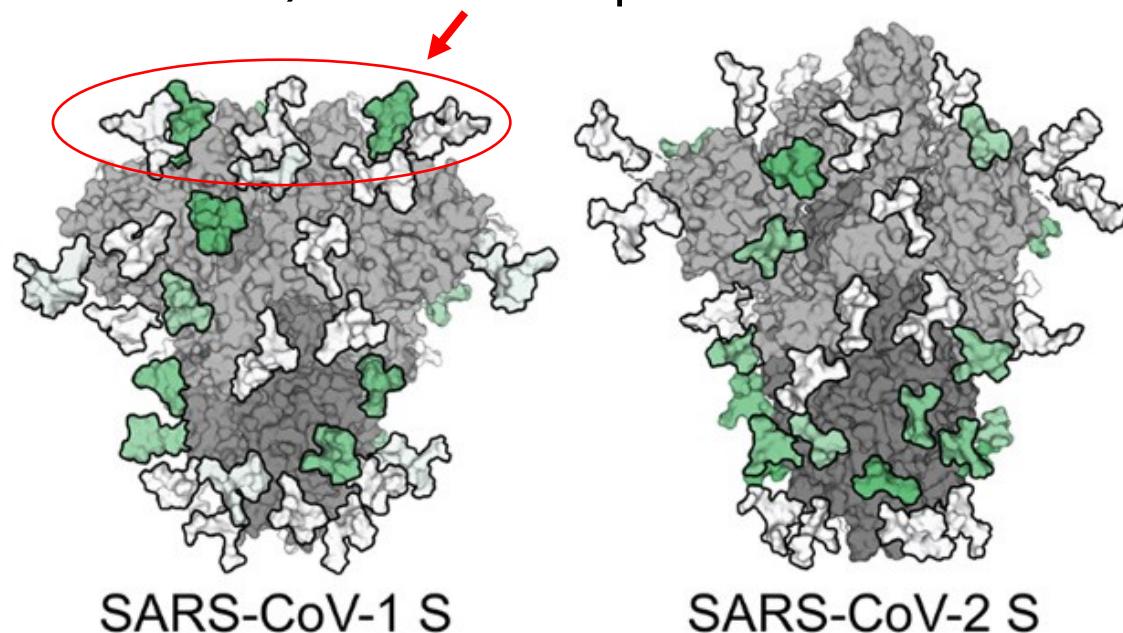
Comparison of *E. coli* produced RBD to baculovirus and human cell line expressed versions

- o RBD was also expressed in human cell line, as well as in baculovirus infected insect cell line
- o All three versions were used to vaccinate mice
 - 20 micrograms purified RBG, followed by two boosts of 10 micrograms each in 3 week intervals
 - The mice were tested for:
 - humoral immune response (antibodies)
 - protective immunity
- o All three yielded humoral (antibody-mediated) response, although less so with *E. coli*-made RBD
- o All three vaccines conferred similar protective immunity (in mice) to SARS CoV-1 infection

Advantage and disadvantage to *E. coli* expression?

- Using *E. coli*, lots of doses of the vaccine could be made cheaply & easily
- However, the protein would not be glycosylated in *E. coli*, this may explain the lower humoral response in mice

Glycosylation map shows sites (in green and white) that overlap the RBD



2020-21 update:

- Spike (RBD) protein currently being tested as low cost (~\$1/dose) vaccine alternative, see Hotez & Bottazzi 2020
- Vaccine testing in progress in India
- Both SARS CoV and SARS CoV2 RBD proteins being tested
- Proteins produced using the yeast *Pichia pastoris*
 - can do glycosylation
 - Proteins secreted into growth medium, so easier to purify

Protein overexpression: pitfalls and limitations

1) Low or no gene expression

- a) mRNA synthesis and stability:
 - i. Premature transcription terminator
 - ii. Unstable mRNA
- b) Protein synthesis and stability
 - i. Gene has lots of rare codons: if a low abundance codon for *E. coli* is used a lot in gene, translation rate will be slow (limited [tRNA])
 - i. Check codon usage in *E. coli* at:
<http://people.mbi.ucla.edu/sumchan/calton.html>
 - ii. Protein product is toxic to the cell: gene deletions occur
 - iii. Protein is degraded by cellular protease

E. coli rpoB
ATGGTTTACTCTTACCGAGAAAAAACGTATTCTGTAAGGATTGGAA
ACGTCCCAAGCTTCTGGATCTCTTACGCCGTTCTACAGCTTGACT
CGTTTCGAAATTATCAGCAAGATCTGAAGGGCAGTATGGTCTGGAA
GCTGCTTCGTTCTGGATCTTCCCGATTGAGCTACAGCGGTAATCCGA
GCTGCAATACCGTCACTGCCGCTTGGCAAGCGTGTGGCTCAGCAG
AATGTCAAATCCGTGGCTGACCTTCCGACCGCTGCGCGTTAAACTG
CGTCTGGTATCTAGGCCGGAAGGCCGGAAGGCCGCTAAAGACAT
TAAAGAACAAAGAAGTCTACATGGCGAAATTCTCGCTCATAGCACAG
GTACCTTGTATCAACGGTACTGAGCGTTATCGTTCAGCTGCAC
CGTAGCTGGGGCTCTTCTGACTGCCGAAAGAAGTAAACCCATACTTC
GGGTAAGTGTGATCTAGGCCGTTATCATCCCTACCGTGGCTCAG
TGGACTTCGAATTCTGATCCGAAGGACAACCTGTTGAGTATCGACCGT
CGCGTAAAGTCTGCCGACCATCTTCCGGCCCTGAACATACACCC
AGAGCAGATCTCGACAGTCTTGGAAAAGTATCTTGAATCTCG
ATAAACAGCTGAGATGGAACCTGGTCCGGAAACGCTGCGTGGTAAACC
GCACTTTCGATCAAGGAACTGGTAAAGTGTAGCTGGAAAAGGCCG
CCGTATACCTGCCGCGCACATTGCCAGCTGGAAAAGACGACGCTAAC
TGATCGAAGTCCCGGTTGAGTACATCGCAGGTAAGTGTGCTAAAGAC
TATATTGAGTGTAGCTACCGGGAGCTGATCTGCCAGCGAACATGGAGCT
GAGCCTGGATCTGGCTGAACTGGCAGCTGCCATCAACAGCTATCG
AAACCGCTTACCAAGCATGCTGGATCACCGGCTTATATCTCTGAAACC
TTACGTTGTCGACCCAACGTTACCGGCTGCTGAGCCACTGGTAAAGTCTA
CCCGATGATGCCCTTGGCAAGCCGGCAGACTGTTGAGCAGCTGAAAGCC
TGTGAGAAGCTTCTTCTCGAAGACGGTTAGTGTGTTCTGGGTT
GGTGTGATGAAGTTCACCGTCTCTGTCGCGGAAGAAATCGAAGGTT
GGTACCTTCGACCAAGACATCATTGATGTTGATGAAAGGCTCATCG
ATATCGTCAAGGTAAGGGCAAGTGTGATATCGACCACTCCGGCAC
CGTGTATCCGTTCCGTTGGCAATGGCGAAACCGATTCGGCGTTGG
CTCTGGTACGGTGTAGGCTGGTAAAGGCGCTGCTGCTGGGAGATC
TGGTACCCCTGATGCCCCAGGATATGCAACGCCAACGGCTTCCCA
GCAGTGAAGAGTCTCGGTTCCAGCCAGCTGCTCAGTTATGGACCA
GAACAGGAGCTTCTGGTCAAGGATATTGCCAACGCTGATCTCCGACTCG
GCCAGGGCTGCTGGCTGAACTGGTCTCGAGGTTGAAGTGGAGCTGG
TAACCTGGCTAGCAAGGGCAATCCAGCTGTTGACGCCGACCGGTT
GACTATGGAGCTTACCCAGCAGGGTGTATCGCTGGCTGGCTCT
GATCCGTTCTGGAAACAGTGGCAGCCCAACCGTGGCATGGTGGTGG
ACATGCAACGTCAGGCCCTGGACTCTGGCTGCTGATAAGGCCGCTGGT
GGTACTGTTATGGACAGCTGTTGGCTGACTCCGGTGTAACTGGGT
AGCTAAAGCTGGTGTGCTGTTAGTACGGTGTGCTCCGTATGTTA
TCAAAAGTAAAGCAAGACGAGTGTACCTGGTGAAGCAGGTACGACATC
TAAACCTGCAACAAATCACCGCTTCAACCGAACACCTGTATCAACCA
GATCCGTTGTGTCCTGGTGAACCGGTTGAACGTGGCAGCTGCTGG
CAGAGCTGCCCTGCCAGGCCACTGGTGAACCTGGCTGGTCAAGGATC
CGCGTAGCTTGTGATGCCGTTGAATGGTACCTGGTCAAGAACCTC
CGTATCCGAAGCTGTTGTCAGGAAGACGGTTTACCCACATC
AGGAACCTGGCTGTGCTGGTGAACCAAGCTGGCTGGCAAGAGATC
ACCGCTGACATCCGAAGCTGGTGAAGCTGCCGCTTCAACCGATGTA
ATCCGTTATGTTACCTGGTGGCAAGTGGCTGGGACATTCTGG
TTGGTAAAGTAAAGCCGAAAGTGAACACTGACCCAGAAAGAAAAA
CTGGCTGGTGTGATCTGGTGAAGAAAGCTCTGACGCTTAAAGACTCTC
TCTGGCTGACCAAAAGCTGTATCCGGTACGGTTATCGACGTTAGGGT
TAACTCCGATGGCTGGTGAAGAAAAGACAAACGCTGCCGCTGGAAATC
ATGCAAGCTTCAACCGGAGAAGAACGACTGGTCAAGAACAGTGG
AAGGTATATCTGGCTGTTAACCGGCTATCCAGCTGGTGAACAGATGG
AGGTGCTGACGTAACAGGTTGAAATTCTCGATCAACCGGATCGAAG
ATATGCCCTACGATGAAAGAACGTTACGCCGAGACATCTGACTGAAACCG
CTGGGCTGACCTGGTGTGATGAACATGCCGATAGCTCCGAAACCCACT
GGGTATGGCTGGCAAGGATGATGCCGAGAACATGACCGCTGCAAC
AGAGCAGTGGCTGTTAGGAGAAGAAGTAAAGTAAAGTACTGTA
GAGGAGACGGGAGATGCCGAACTGGTAAAGTAAAGTAAAGTAA
GAGGATGCCGAGAACATGGCTGTTCAAGGAGTGGCTGGGAGAG
GTGTCATTGGACTTATGTGCCCCGAGGAGACATGCCCTGGGAGAG
GTGTCATTGGACTTATGTGCCCCGAGGAGACATGCCCTGGGAGAG

GGAAATTGTACCCGATCTCATAGTTAATCCACCGGTATACCATCAAGAAT
GACCCTGGTCAGCTTGAAGCTTGTGGTAAGGGTGTCTTCACTTA
AGGGAAAGGAGGGTGTGGAAACAGCATTTCATGGAAACCCAGAAGAAC
CTAAGAAAAGAGCTAGAAGAGCTTGATTCAAGCATACGGGTAGAGAAAAT
AATGTATGATGGTTAACAGGGGAAGAGATTAGGGCTGATATTCATTG
GTGTGATCTACTATCAGAGGCCCTGGACATGGTTGCAGATAAGATACAT
GCACCTTCAAGAGGGCTCTGTCCAAGTTCTAACTAACGAAACCCGAGGAAGG
TAGAGCAAGAGAGGGGTGCTAAAGATTGGAGAAATGGAGAGAGCGTTC
TTGGTGGCCATGGAGCTGCATGTTACTAACTAGAGCGTTTACTTGGAGAG
AGCGATAAGAACGTAAGATTGGTGTGGAGAACTGGACACATGACT
AGAGGACAAGAGAAGAAGAAGAGTCTACTGTCTGTGGAGAAGAAC
AAAGAATTAGCAAAGTAGAGATGAGCTACCGCTTCAATTGTTGCTTGT
GAGTTGAAGGCCATGGTTATTAGACCTAAGTTAACCTCAGAGGGGT
GTGA

Protein overexpression: pitfalls and limitations

2) Protein product is not functional or hard to purify

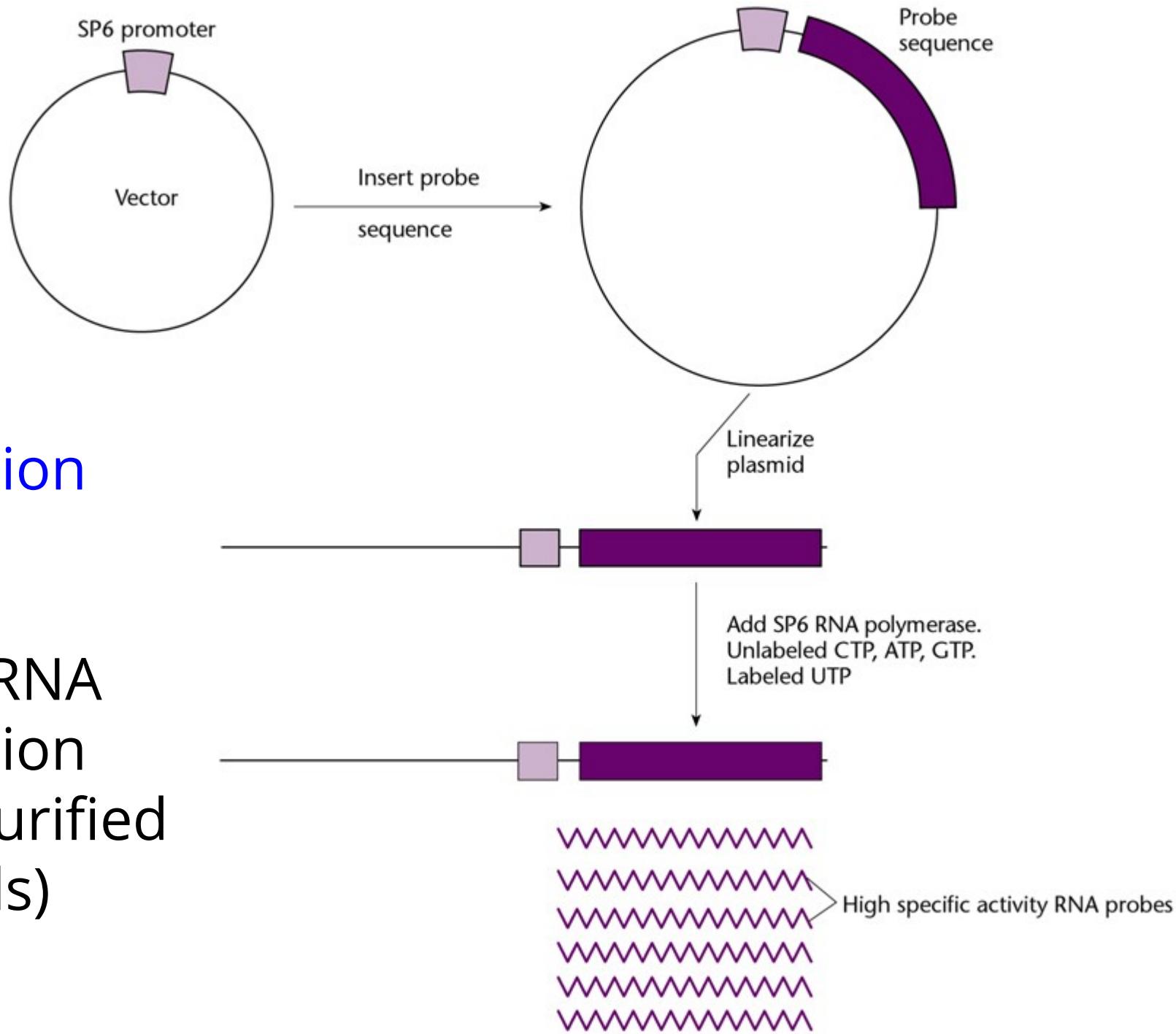
a) *Inactive protein*

- i. Lacking necessary cofactor
- ii. Does not fold properly
- iii. Needs posttranslational modification
 - Glycosylation
 - Acetylation
 - Phosphorylation
- iv. Functions as part of a multi-protein complex

b) *Protein purification or detection difficulties:* hard to separate or distinguish from host proteins

- Add epitope tag

RNA
expression
vectors:
in vitro RNA
expression
(from purified
plasmids)



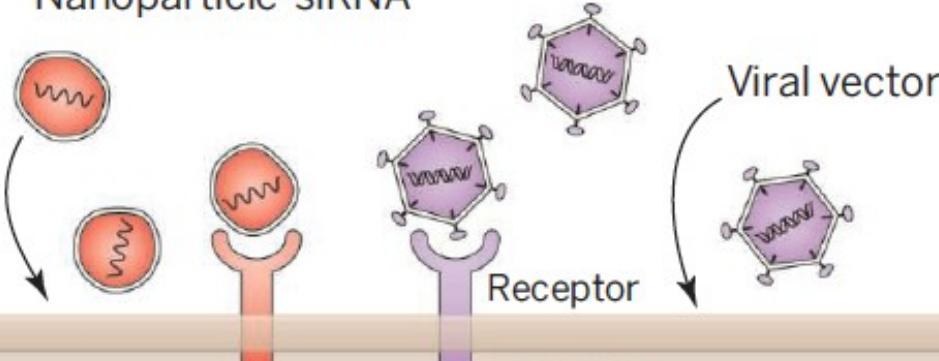
Therapeutic RNAs: RNA interference (RNAi)

- Micro RNAs and small interfering RNAs can be used to knock down expression of disease-causing genes.
- The current best RNAi therapeutic target is the liver
 - because of the liver's physiological role in removing particles from circulation, the RNAi therapeutic agent is very likely to reach it
- Some diseases that can be targeted at the liver:
 - Transthyretin, causing amyloid neuropathy
 - Hepatitis B
 - Liver cancer
 - Hepatic porphyria (heme biosynthesis disorder)

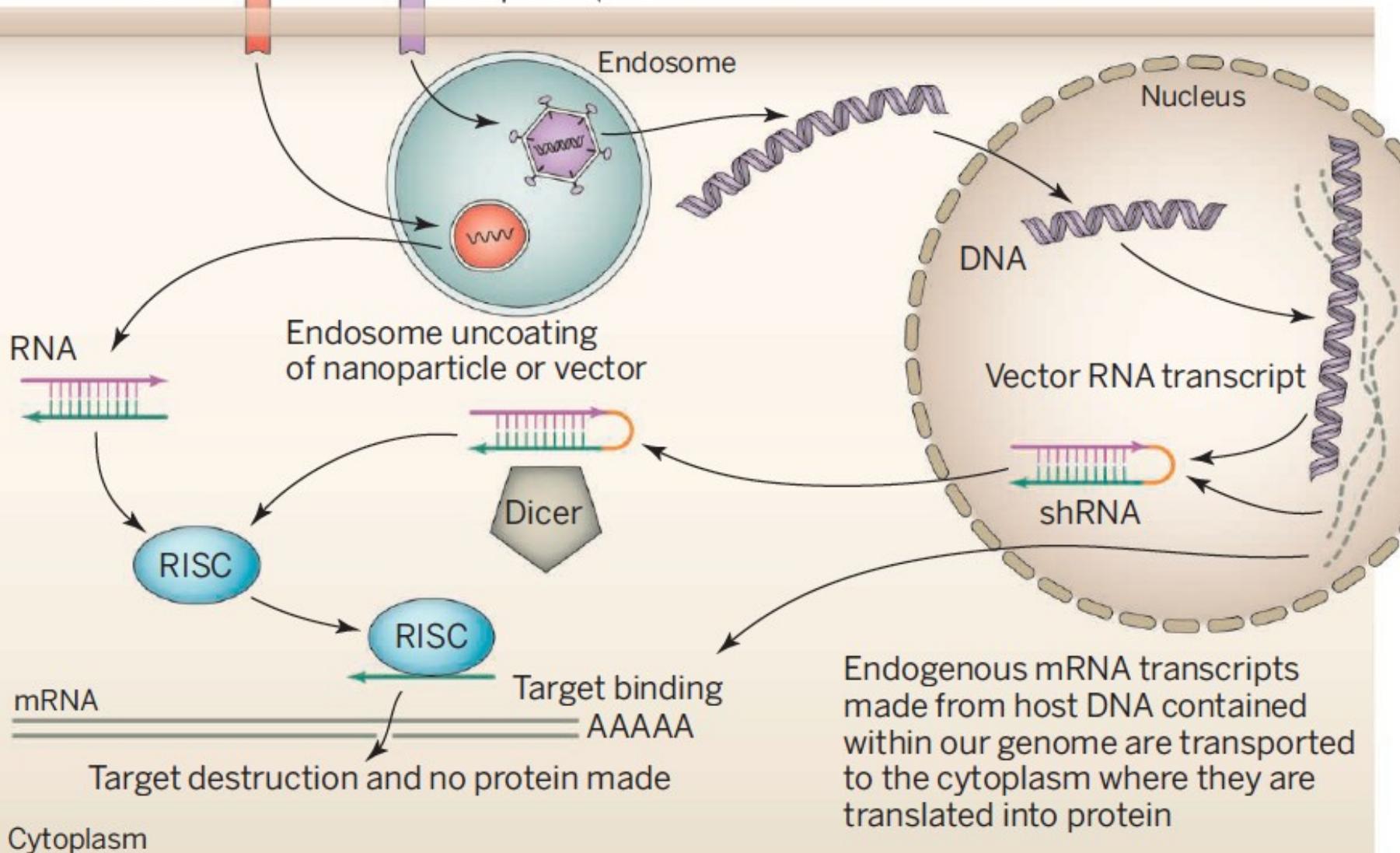
Two modes of delivery of therapeutic RNAi

- 1) Premade RNAi precursors produced by *in vitro* transcription or chemical synthesis
 - Lipid nanoparticles (LNPs)
 - GalNac -RNA conjugates, which induce receptor mediated endocytosis
- 2) RNA expression vectors (recombinant virus) that make RNAi precursors upon entry into cells
 - Theoretically better for life-long treatments or hard to reach tissues
 - Precise control of expression is hard to achieve

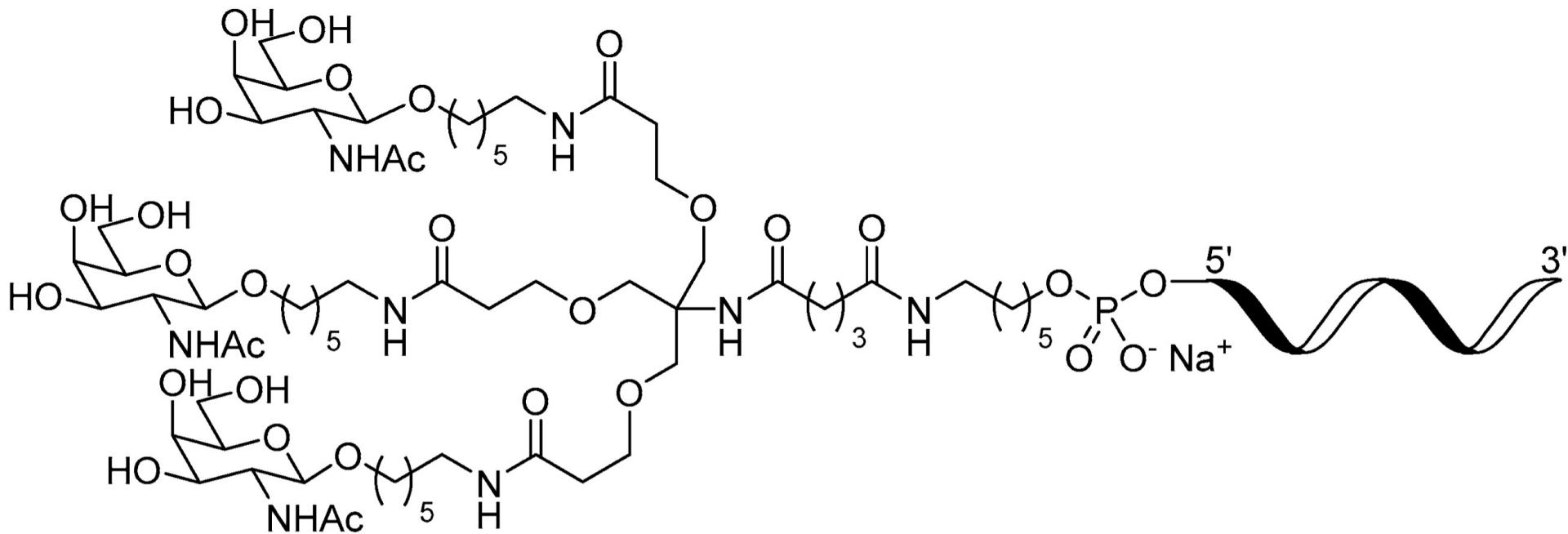
Nanoparticle-siRNA



Two delivery modes
for RNAi therapy:
nanoparticle versus
viral vector

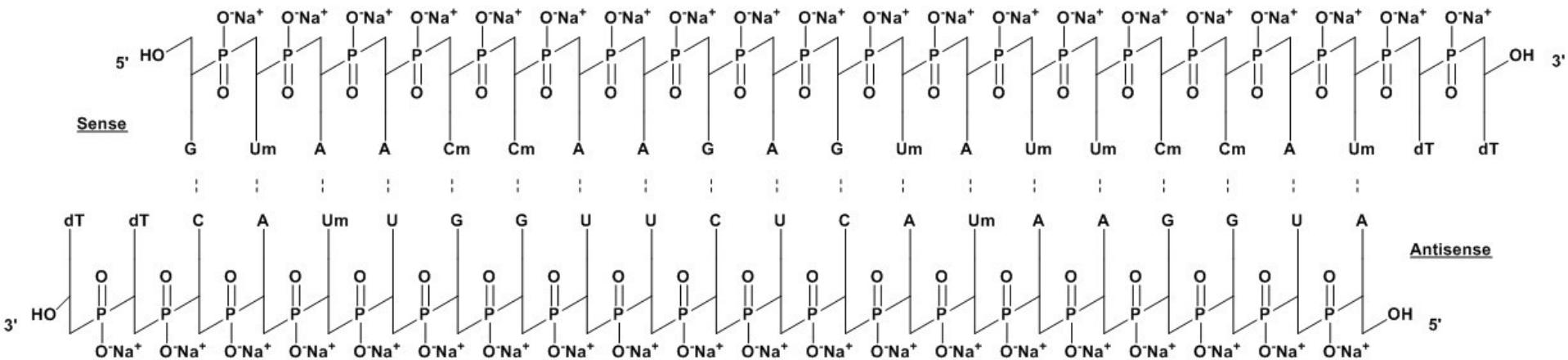


Nanoparticle (NP) example: N-acetylgalactosamine (GalNac)/RNA conjugate



- Binds to the asialoglycoprotein receptor that is highly expressed on hepatocytes
- Results in rapid endocytosis

Patisiran: first RNAi drug approved by FDA (2018)

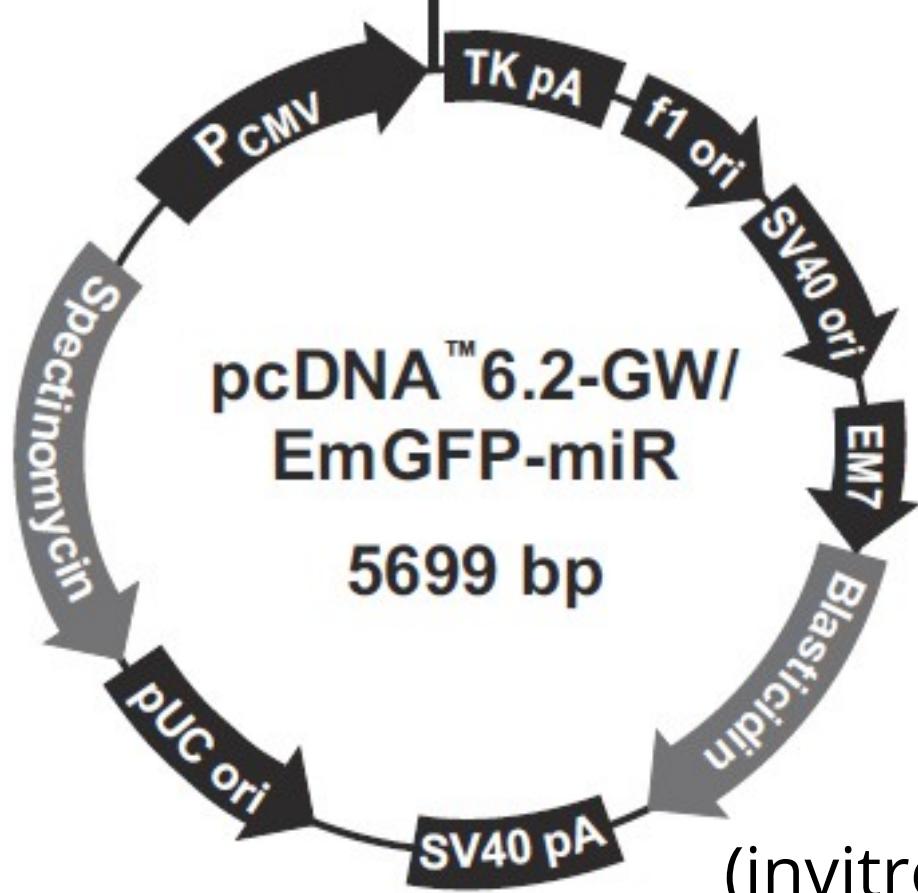


A, adenosine; C, cytidine; G, guanosine; U, uridine; Cm, 2'-O-methylcytidine; Um, 2'-O-methyluridine; dT, thymidine (modified nucleotides help tame dendritic cell immune response)

- Formulated as lipid NPs for delivery to hepatocytes
- RNAi specifically binds to a genetically conserved sequence in the 3'-untranslated region (3'-UTR) of mutant and wild-type transthyretin (TTR) messenger RNA (mRNA)
- Treats transthyretin amyloidosis (peripheral neuropathy)

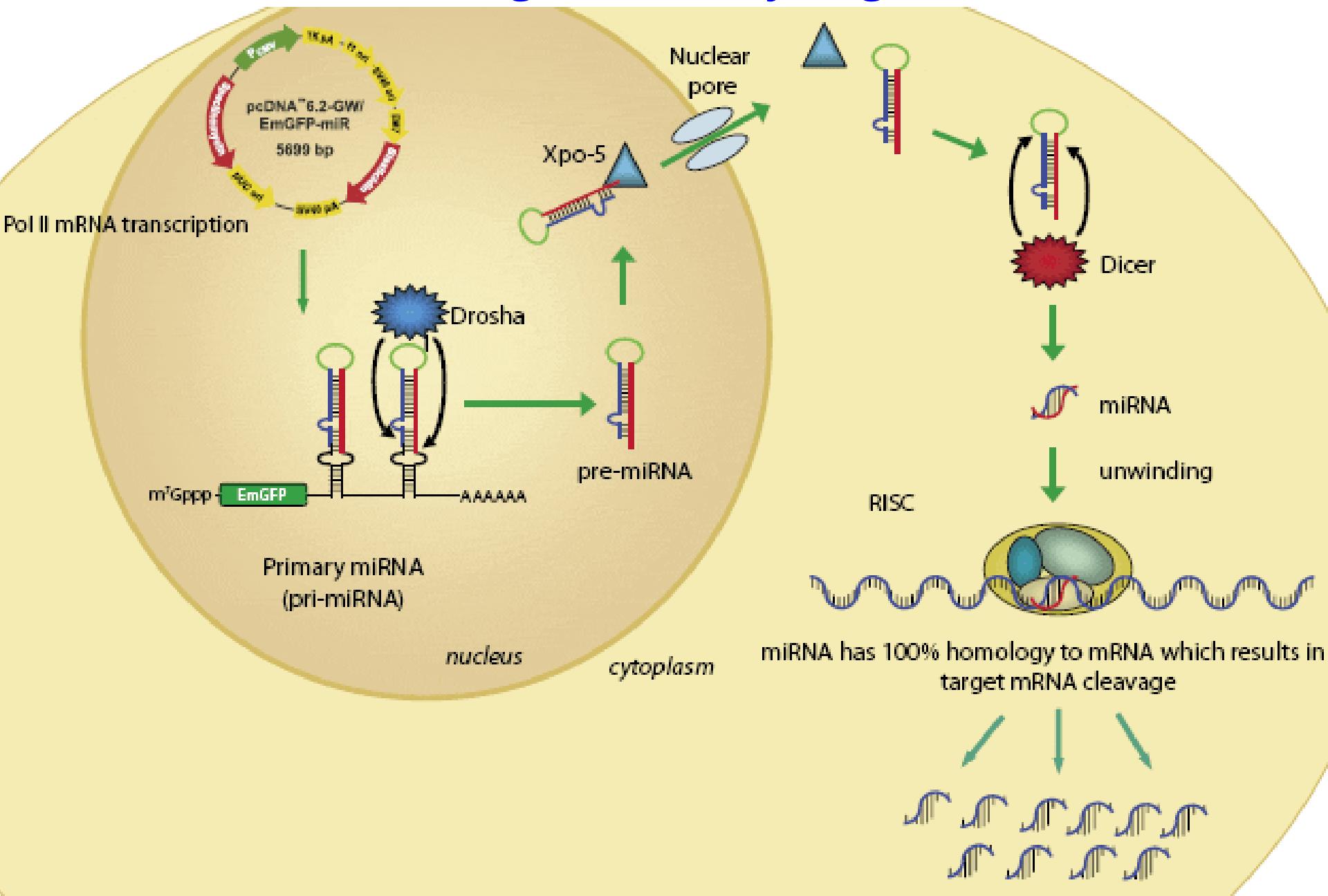


Eukaryotic
vector for
micro RNAs: *in*
vivo RNAi



(invitrogen.com)

Directed mRNA degradation by engineered miRNA



Gene cloning: Expression of protein and RNA

Controlled gene expression

I. proteins

- o Various tags and their functions
- o A protocol for purification of 6xHis tagged proteins expressed in *E. coli*

II. RNA

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?

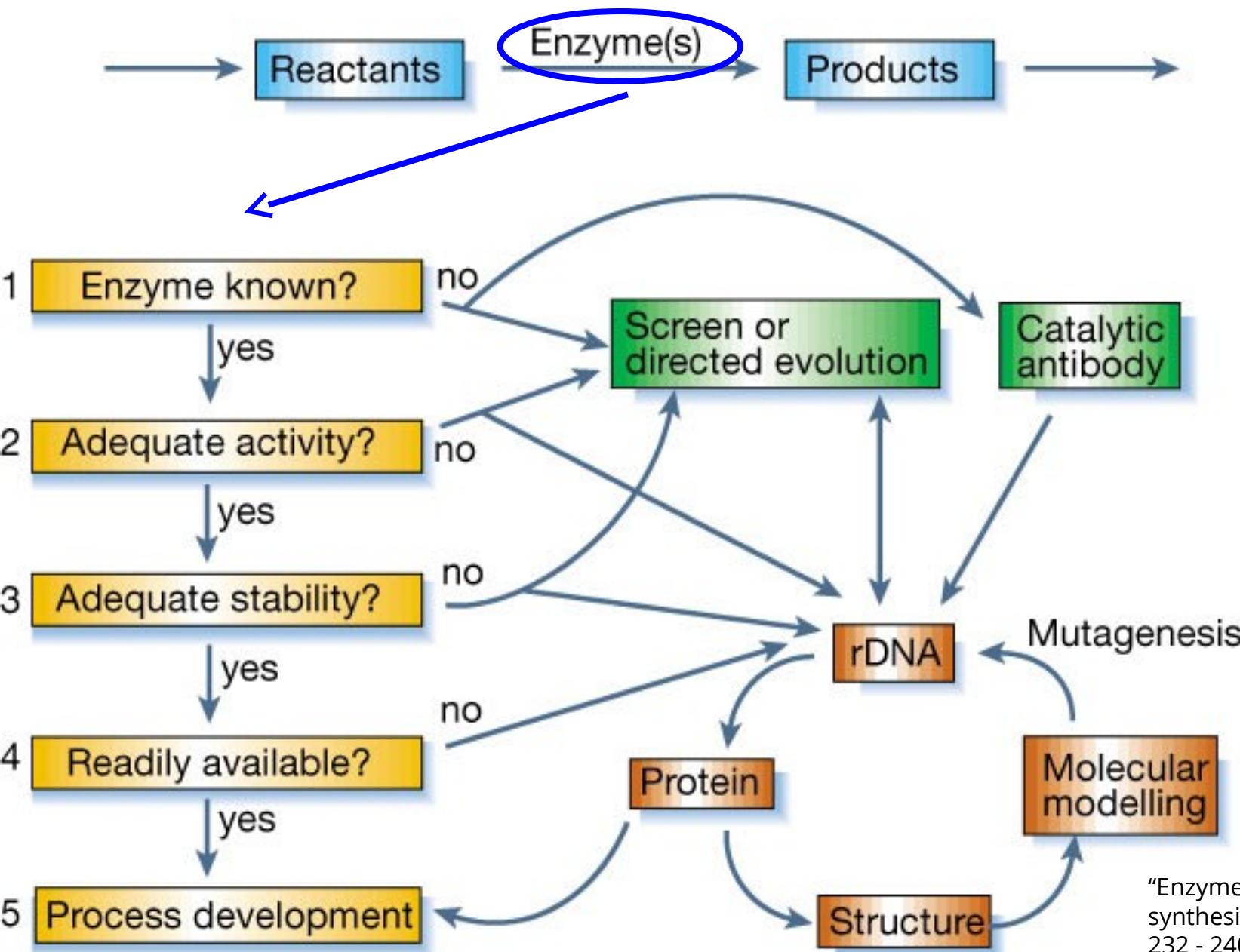
Many enzymes are commercially important, e.g.:

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

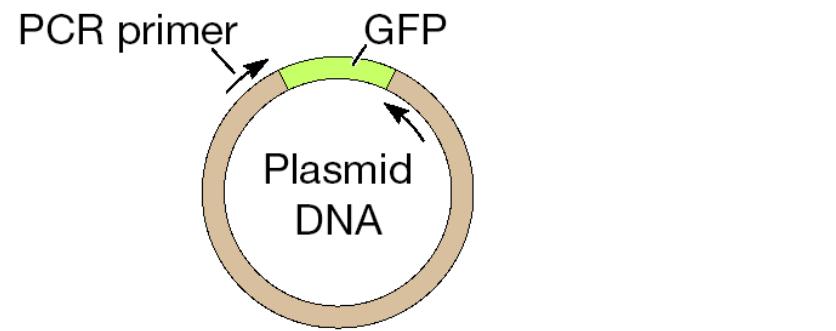
Obtaining and improving useful enzymes



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

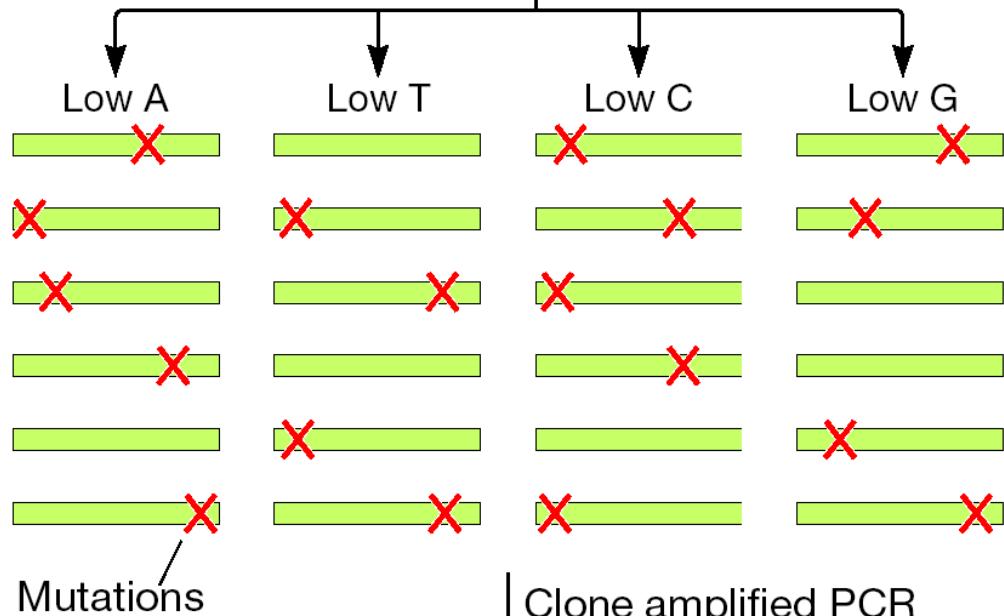
- a) General mutagenesis: expose the organism or its DNA to chemical, UV, or radiation mutagen
- b) PCR mutagenesis: base misincorporation
 - Include Mn²⁺ in reaction
 - Reduce concentration of one dNTP
- c) 'Cassette' mutagenesis
 - Partly randomized oligonucleotides used in cloning
- Many possible 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)



Four PCR reactions, with each nucleotide deficient

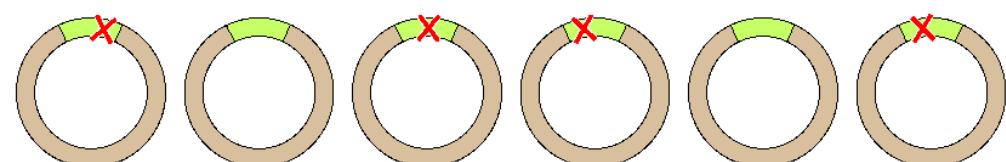
Mutations in
PCR products



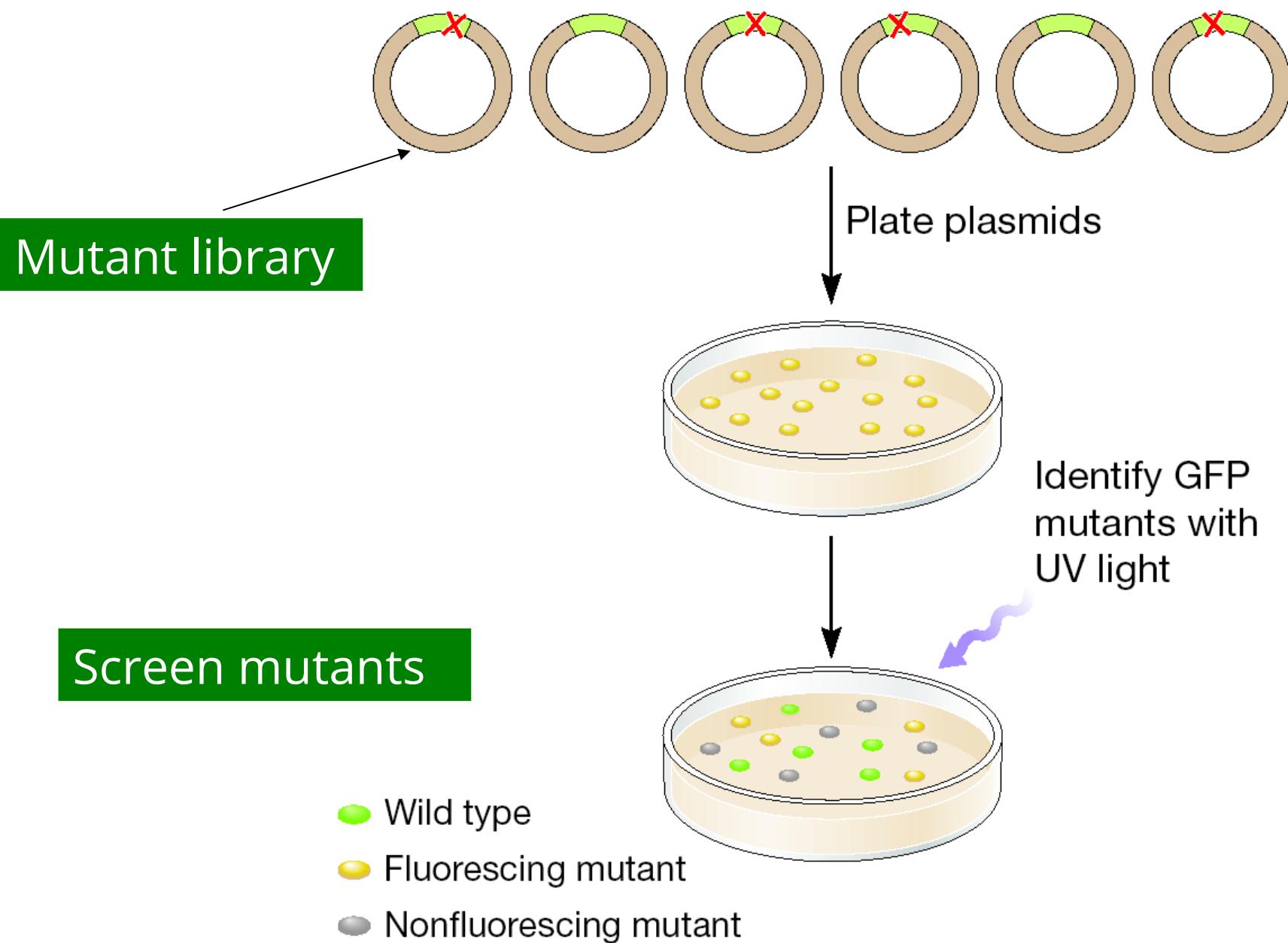
Mutations

Clone amplified PCR
products containing
mutations into plasmids

Mutant library

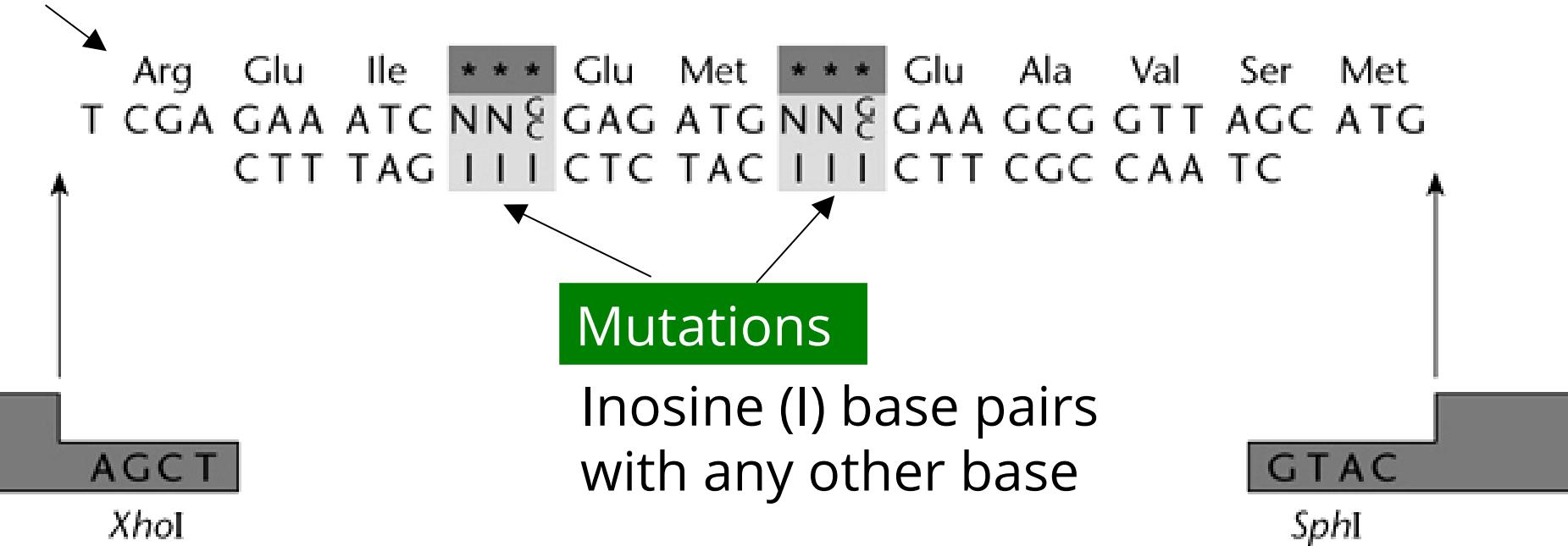


Random mutagenesis by PCR: the Green Fluorescent Protein



Cassette mutagenesis (semi-random)

Translation of sequence



Synthetic cassette DNA strands are synthesized, and annealed

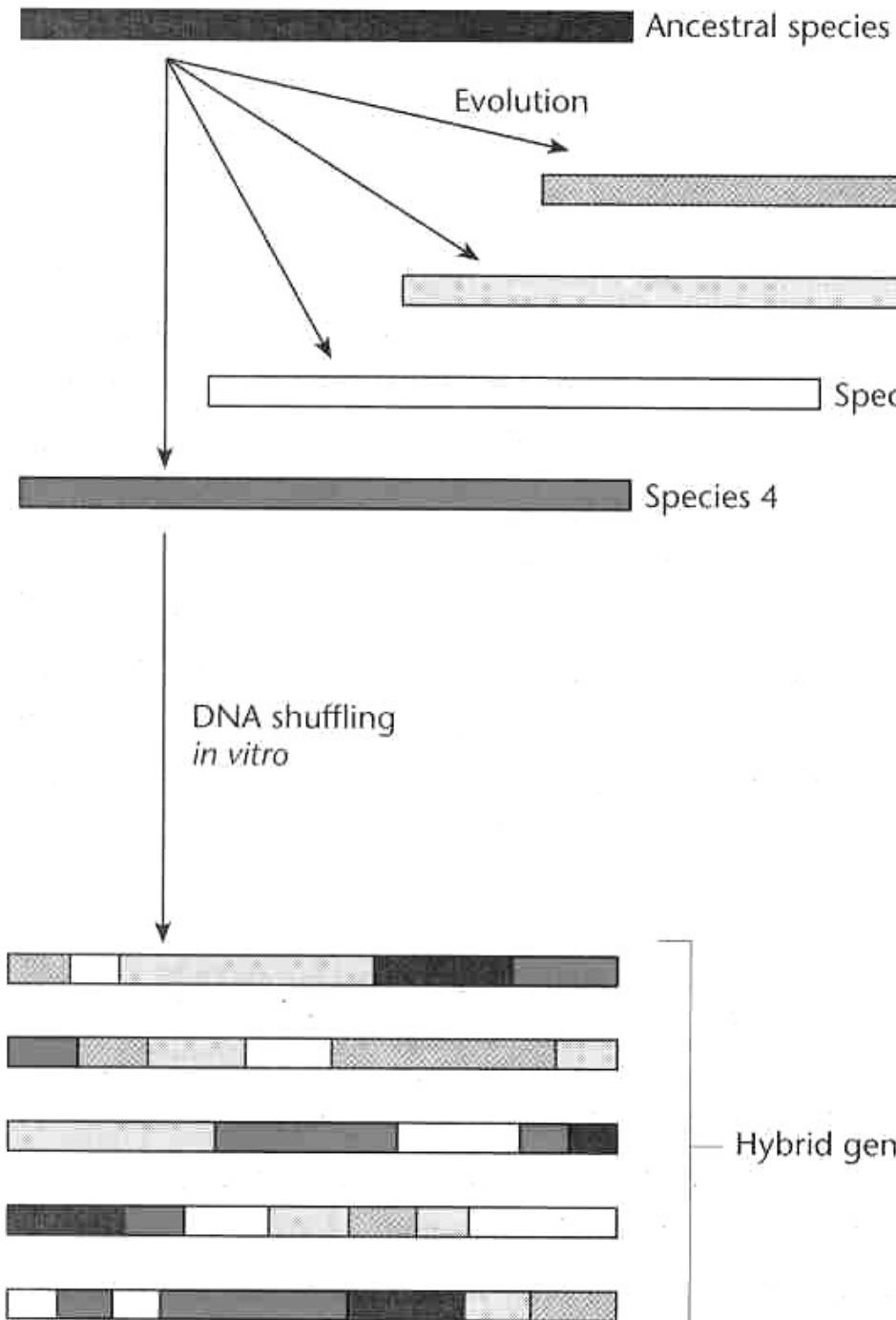
The cassette is ligated to make library, then screened

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*

*	20	*	40	*
401	MIEVKPIN A EDTYEIRHRILRPNQPLEACMYETDLLGGAFHLGGYYRGKL			
B6	MIEVKPIN A EDTYEIRHRILRPNQPLEACKYETDLLGGFHLGGYYRDRL			
DS3	MIEVKPIN A EDTYEIRHRILRPNQPLEACMYETDLLGGFHLGGYYRGKL			
60	*	80	*	100
401	ISIASFH A EHSELEGEEQYQLRG M ATLEGYREQKAGSTLIRHAEELLRK			
B6	ISIASFH O AEHSELEGOKQYQLRG M ATLEGYREQKAGSTLIRHAEELLRK			
DS3	ISIASFH N AEHSELEGOKQYQLRG M ATLEGYREQKAGSTLIRHAEELLRK			
*	120	*	140	
401	KGADLLWCNARTSVSGYYEKLGFSEQGEVYDIPPIGPHILMYKKLT			
B6	KGADLLWCNARTSVSGYYKKLG F SEQGGVYDIPPIGPHILMYKKLT			
DS3	KGADLLWCNARTSVSGYYEKLGFSEQGGVYDIPPIGPHILMYKKLA			

Sites of natural variation are shaded

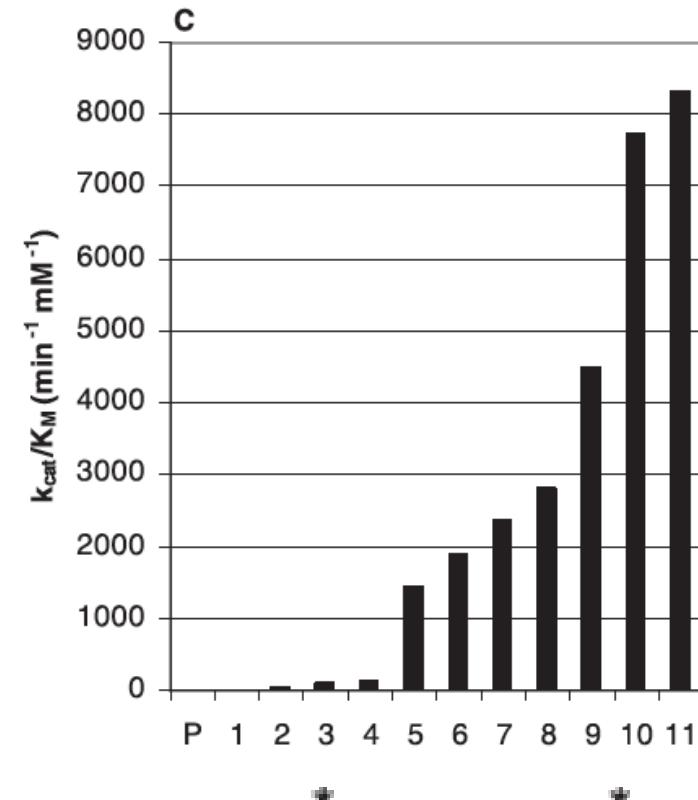
Discovery and Directed Evolution of a Glyphosate Tolerance Gene

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. Lassner¹

Mutant creation, screening, and assays

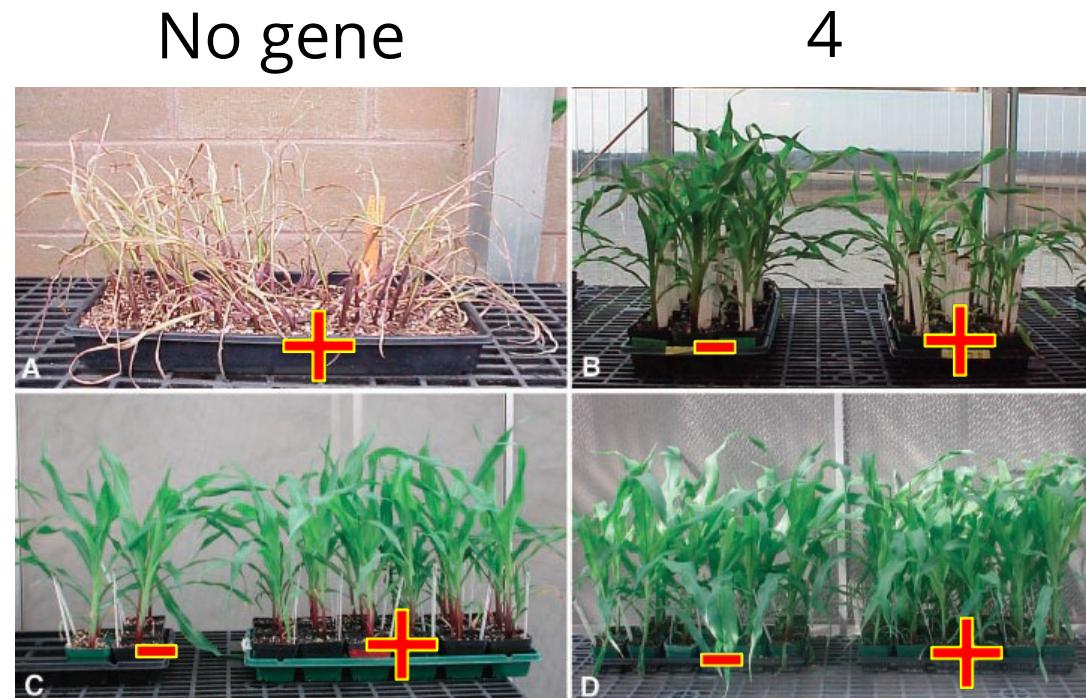
- The three GAT genes were fragmented and shuffled by PCR over 11 rounds
- Additional mutations were added at rounds 4 and 8, based on sequences of related genes in other *Bacillus* species
- After each shuffle, libraries were created
- 5000 gene variants were screened for GAT activity
- Top performing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

Top performer (K_{cat}/K_m) for each round



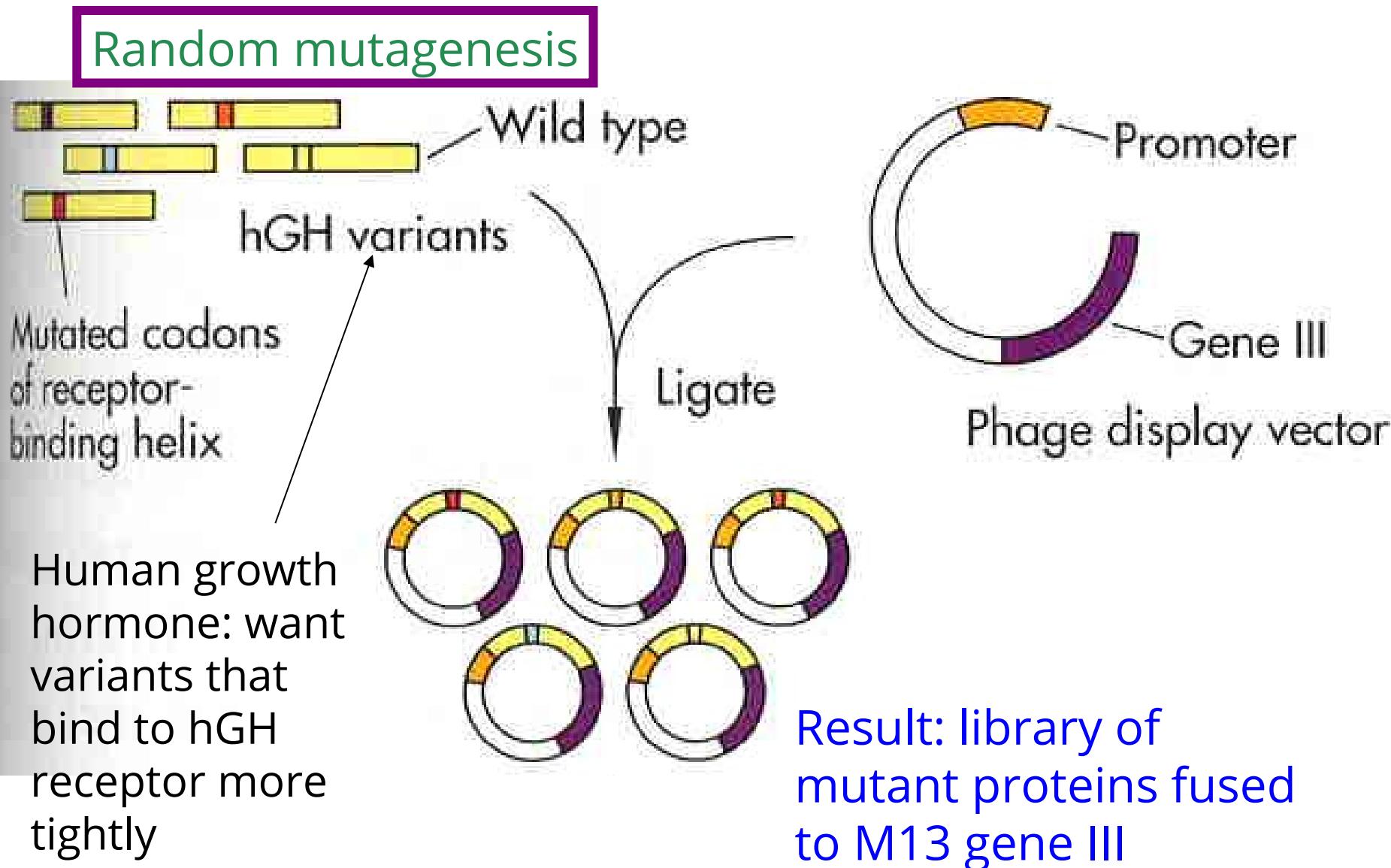
MIEVKPINNAEDTYD**L**RHRV**L**RPNQPIEACMF**E**SDLTRSAFHLGGF**Y**GGKLISVASFHQAE
 HSEL**O**G**K**K**Q**YQLRG**V**ATLEGY**R**EQKAGSS**L**V**K**HAEE**I**LRKRG**A**DMI**W**CNART**S**ASGYY**R**K
 LGFSE**Q**GE**V**F**D**T**P**PG**G**PHIL**M**YKRIT

* 146

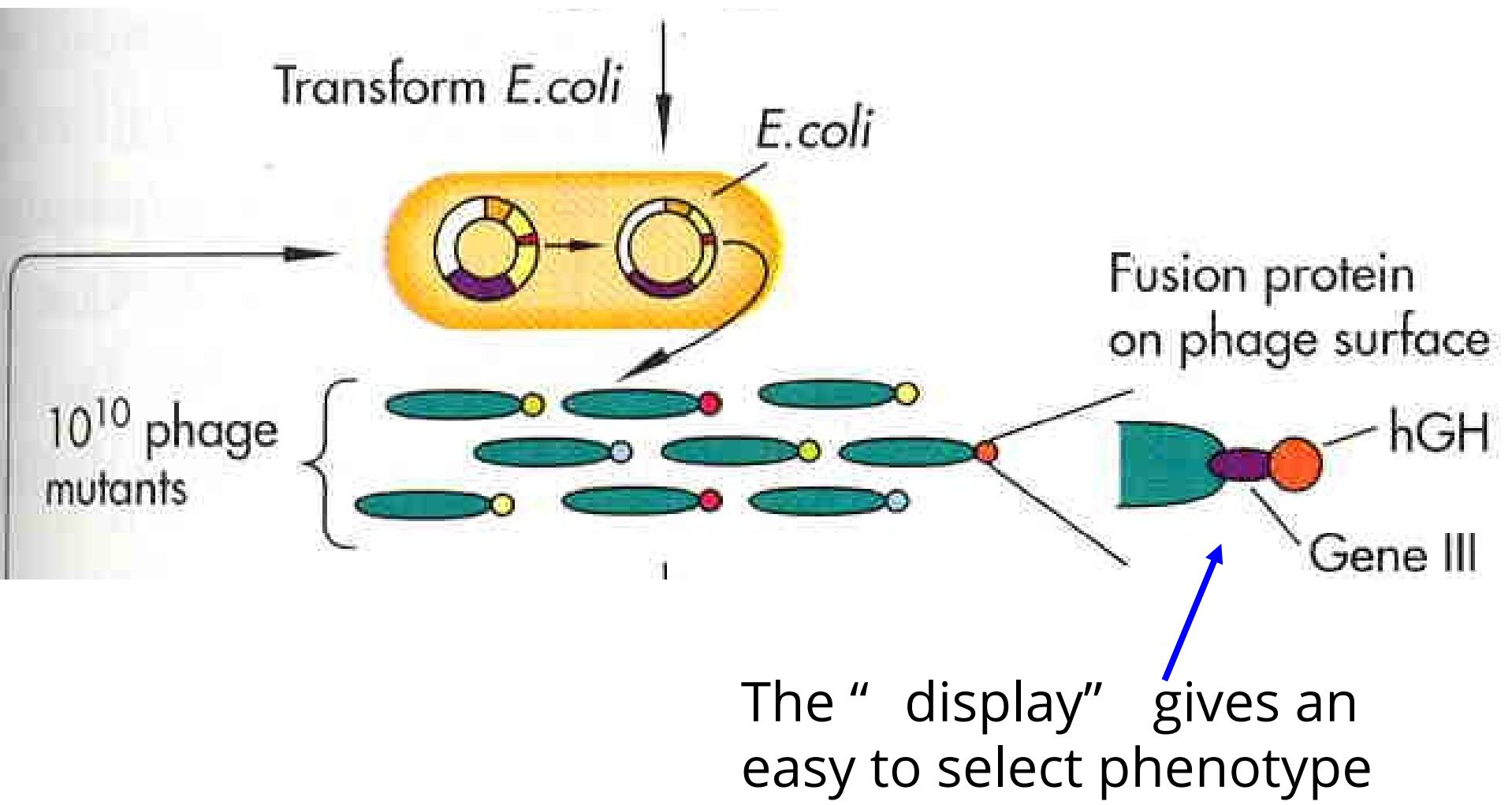


Blue: original diversity, Green: added rd.4, Pink: added rd. 8, Orange: random occurrence

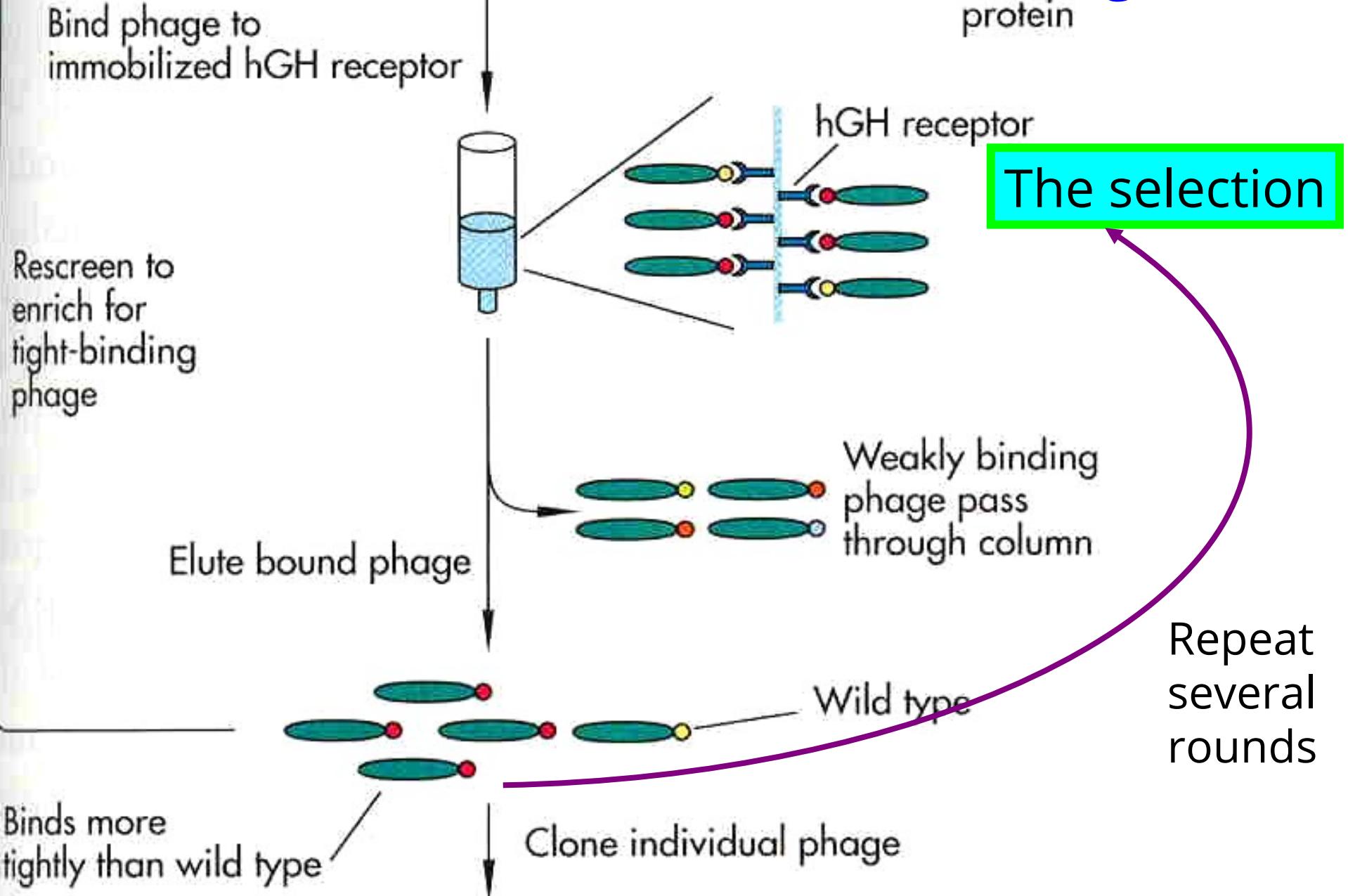
Screening huge (10^9) libraries: phage display



Phage display: production of recombinant phage



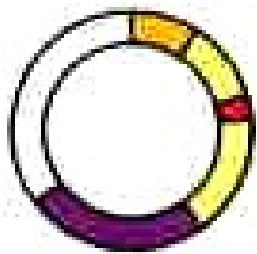
Phage display: collect tight-binding phage



Which sequences were selected?



Isolate phage DNA

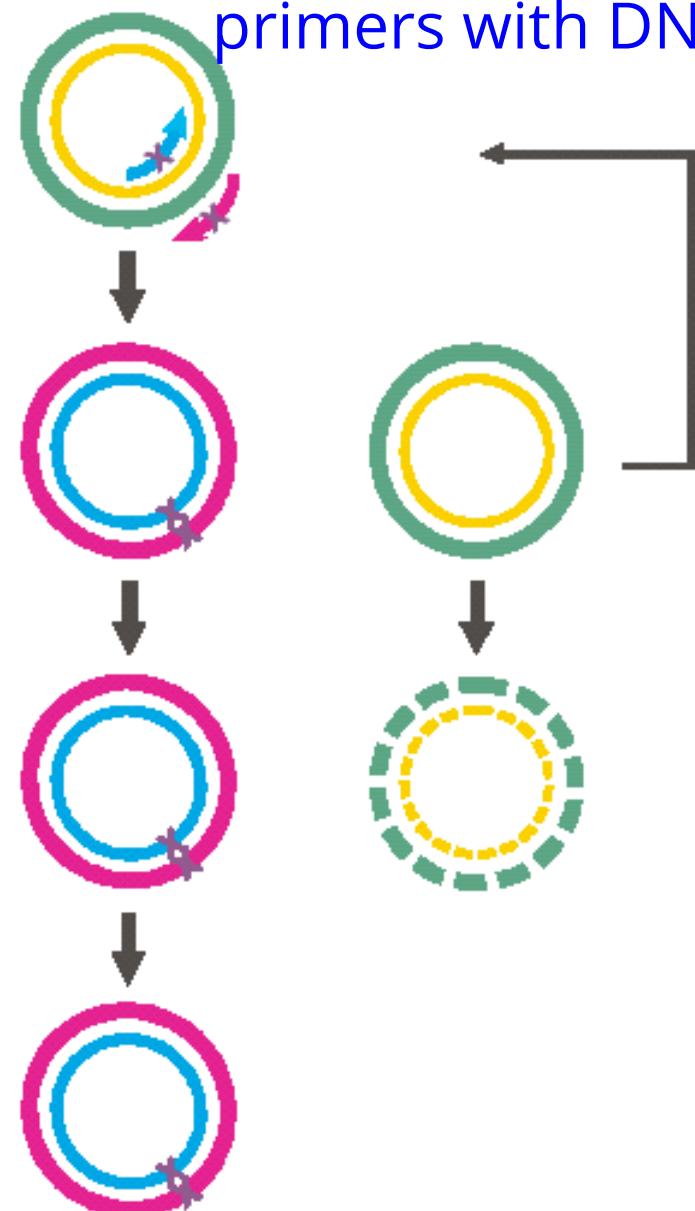


Determine sequence of
tight-binding mutant

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

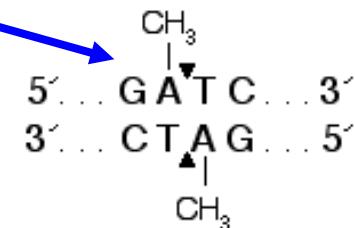
Template DNA is destroyed, only mutated DNA remains

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

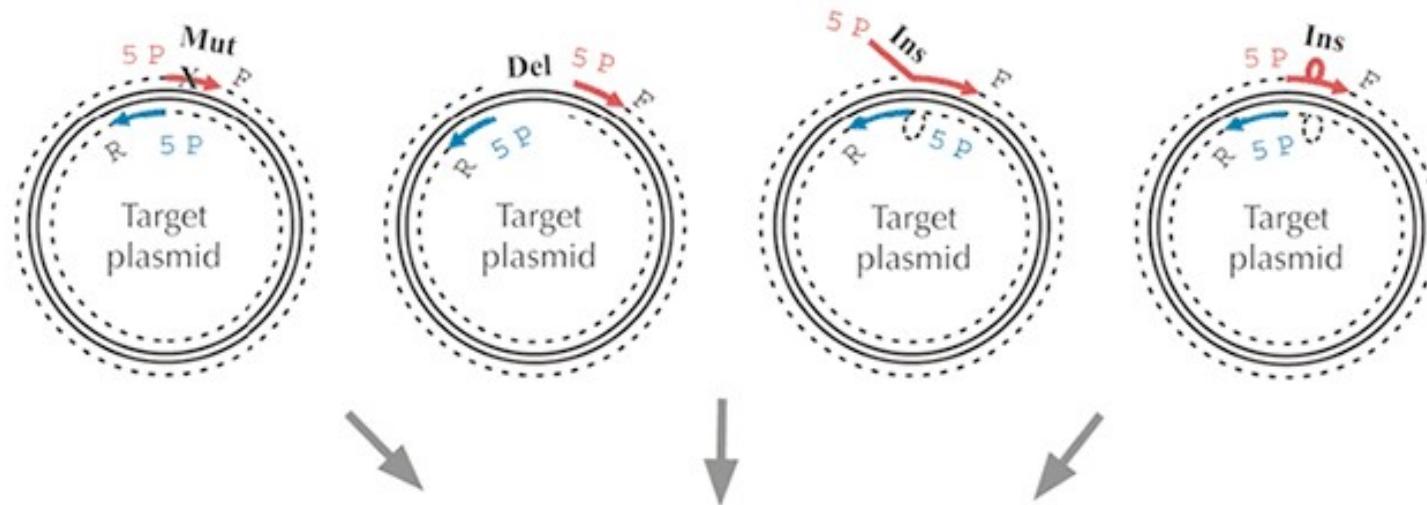
Site directed mutagenesis: plasmid PCR/ligation/transformation

Point mutation

Deletion

Insertion option 1

Insertion option 2



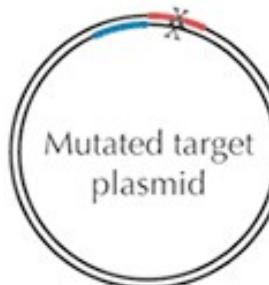
Linear amplified target plasmid
with desired mutation



Step 1.
Amplification of target
plasmid with two
phosphorylated primers.

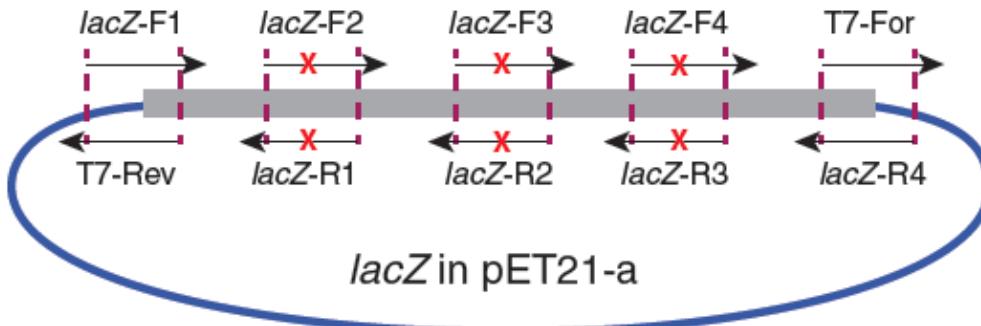
<http://www.thermoscientificbio.com/mutagenesis/phusion-site-directed-mutation-kit/>

Step 3.
Transformation
into *E. coli*.



Step 2.
Plasmid circularization
by ligation.

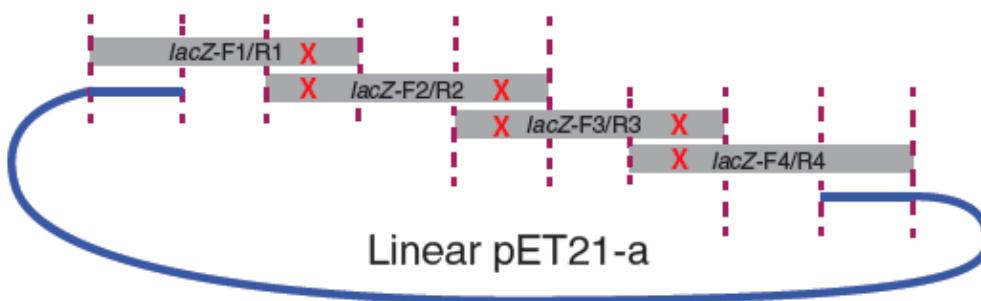
Site directed mutagenesis: Gibson assembly



lacZ in pET21-a



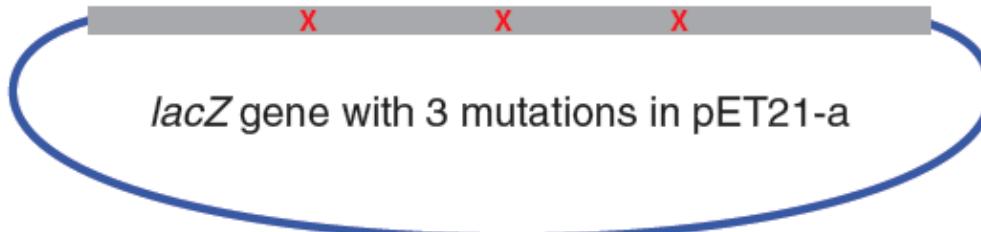
PCR to generate fragments with designed mutations for assembly.



Linear pET21-a



Gibson Assembly Master Mix to join fragments at 50°C.



lacZ gene with 3 mutations in pET21-a

Multiple, overlapping DNA fragments, each with a mutation

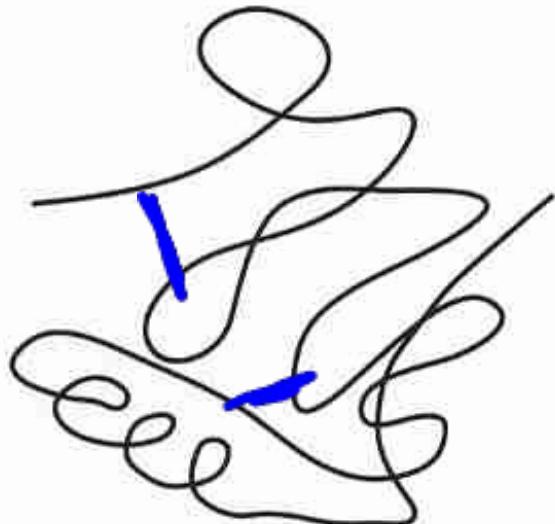
Gibson assembly stitches the fragments together

Rational site-directed mutagenesis

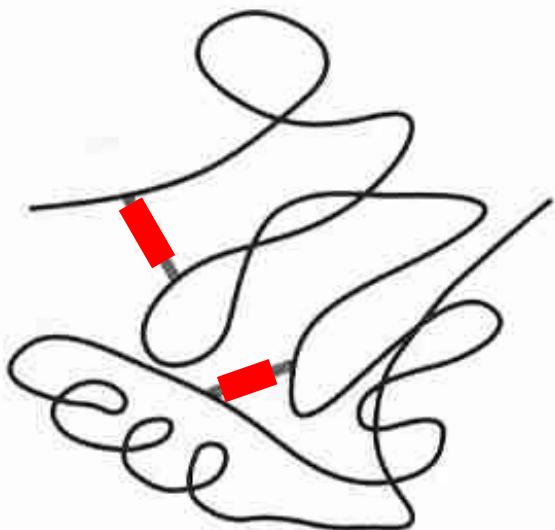
Re-program a protein by making deliberate changes in amino acid sequence, based on structures or other prior knowledge

- 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



Native protein



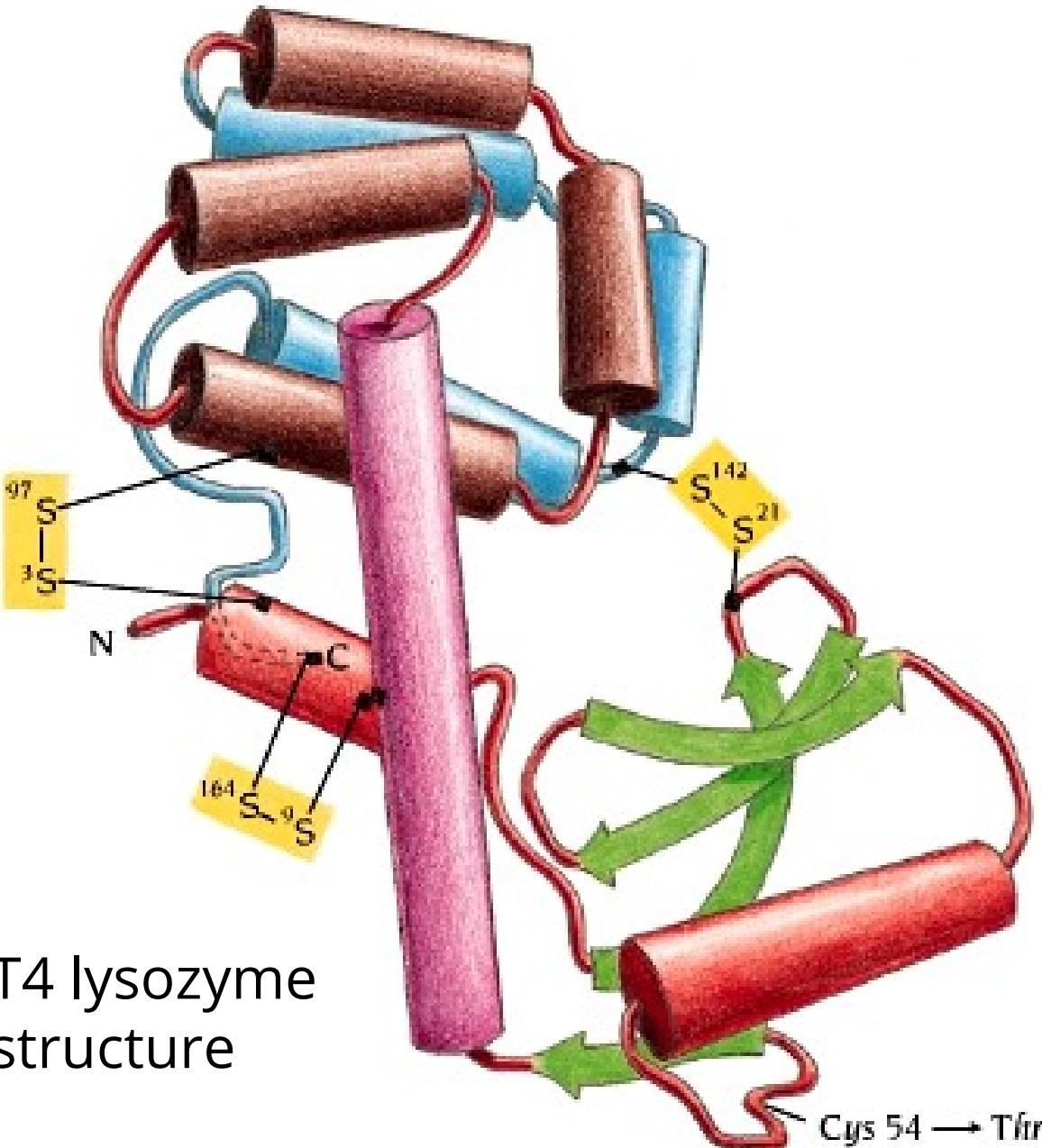
Engineered protein

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

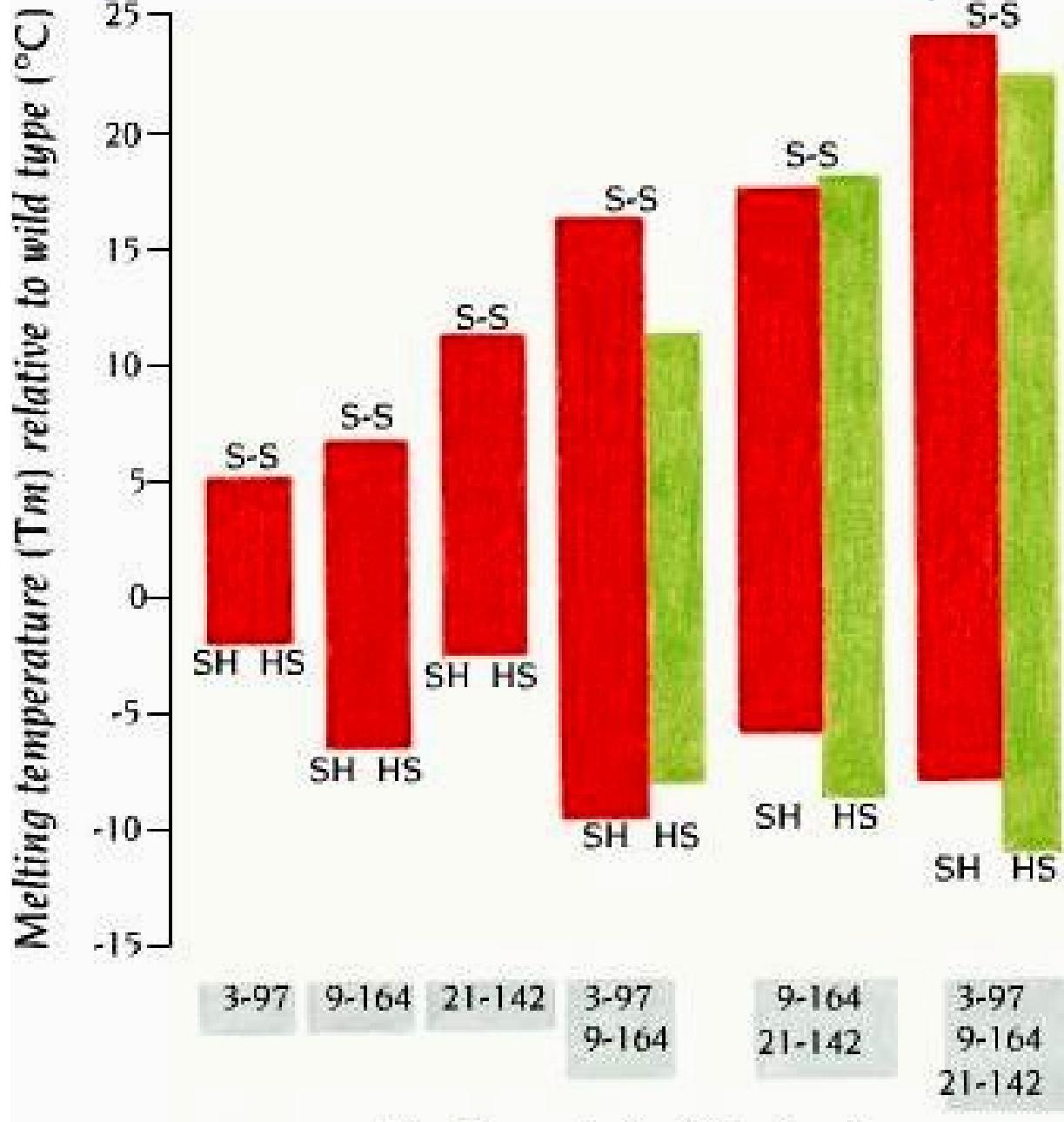
Gene was
mutagenized to
add cysteine
residues in close
proximity

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 enzyme 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

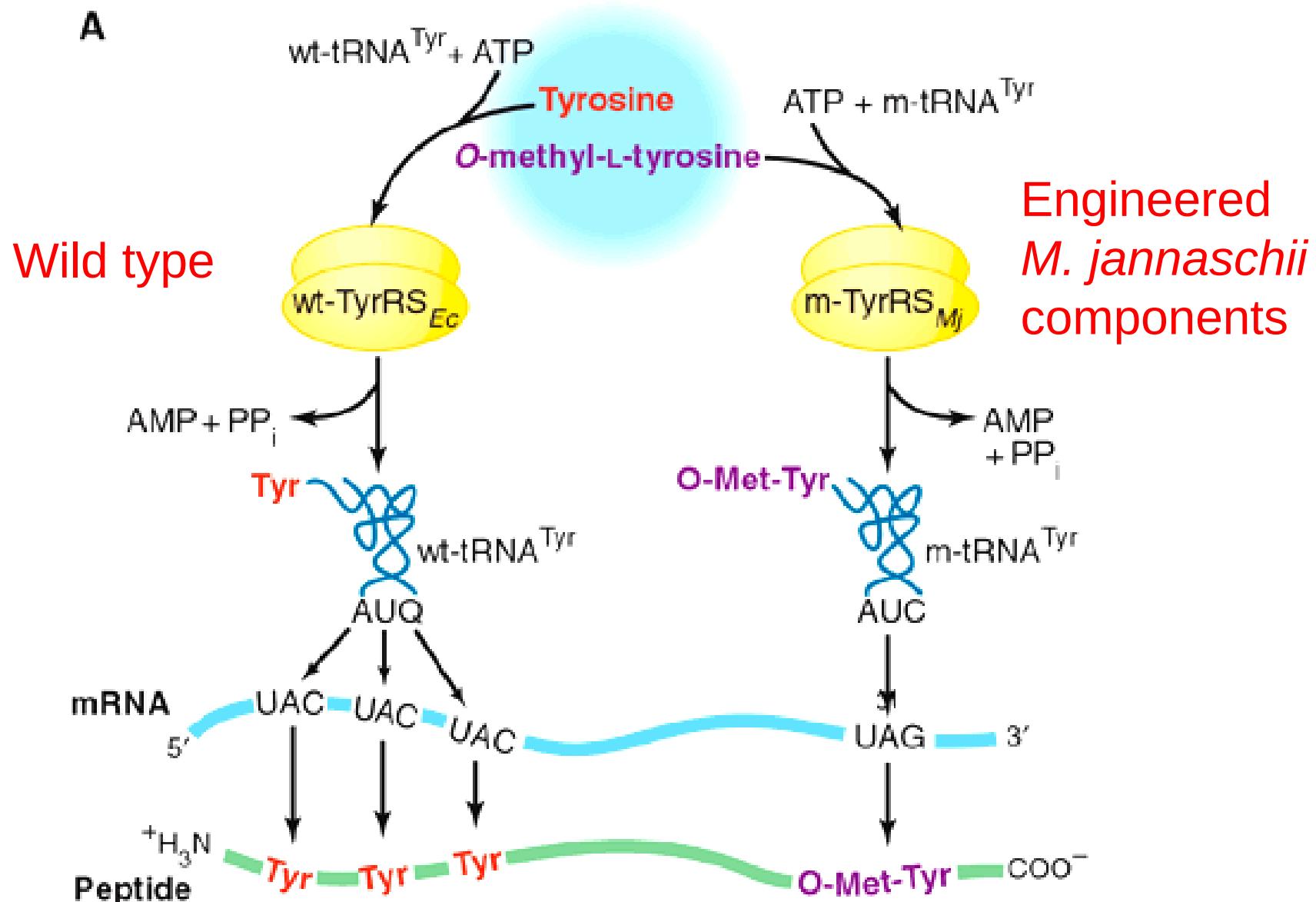
2 key mechanisms of fidelity

- 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

	Second letter					
	U	C	A	G		
First letter	UUU } Phe UUC } UUA } Leu UUG }	UCU } Ser UCC } UCA }	UAU } Tyr UAC } UAA } Stop UAG }	UGU } Cys UGC } UGA }	UGG } Trp UGA }	U C A G
C	CUU } CUC } CUA } Leu CUG }	CCU } CCC } CCA }	CAU } His CAC } CAA }	CGU } CGC } CGA }	CGG }	U C A G
A	AUU } AUC } Ile AUA }	ACU } ACC }	AAU } Asn AAC }	AGU } AGC }	AGA }	U C A G
G	GUU } GUC } GUA }	GCU } GCC }	GAU } Asp GAC }	GGU } GGC }	GGG }	U C A G
	Val GUG }	Ala GCG }	GAA } Glu GAG }	Gly GGA }		

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)

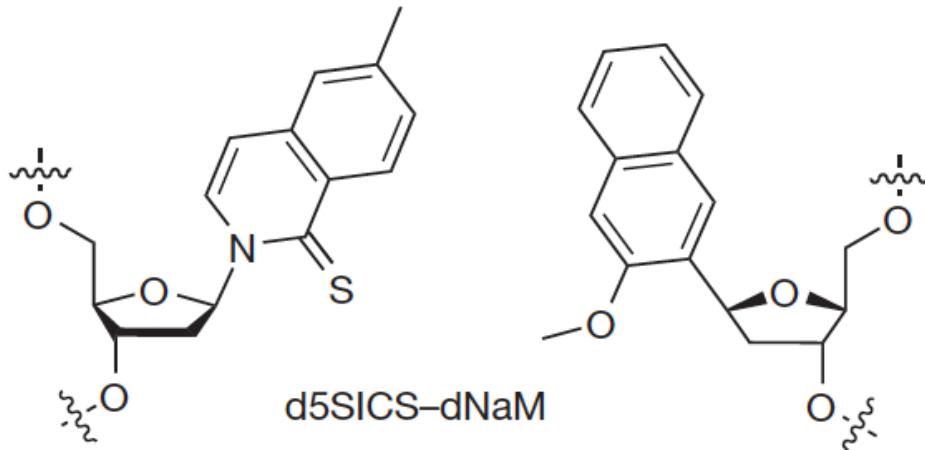
Quadruplet tRNAs charged with unnatural amino acids cause insertion of the unnatural amino acid at the quadruplet

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

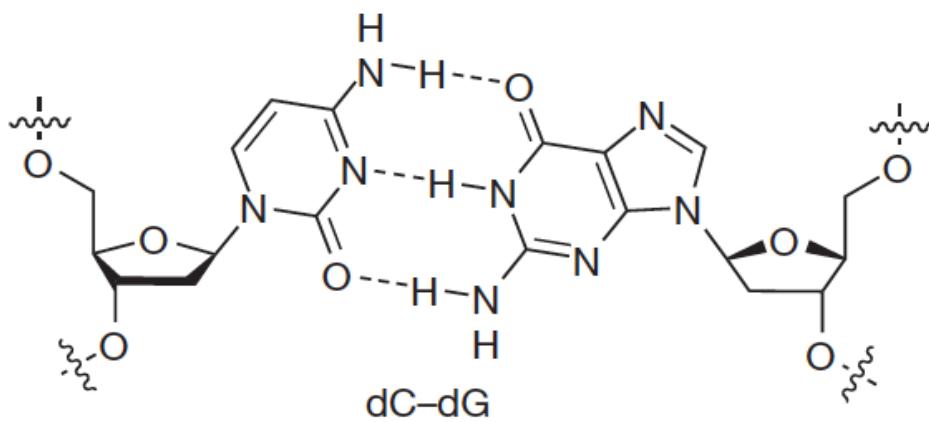
256 blank codons available (as opposed to 3 stop codons), so it could be possible to encode many different unnatural amino acids into proteins

"...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. Malyshev¹, Kirandeep Dhami¹, Thomas Lavergne¹, Tingjian Chen¹, Nan Dai², Jeremy M. Foster², Ivan R. Corrêa Jr²
& Floyd E. Romesberg¹

(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

Applied mutagenesis: pathway engineering and synthetic biology

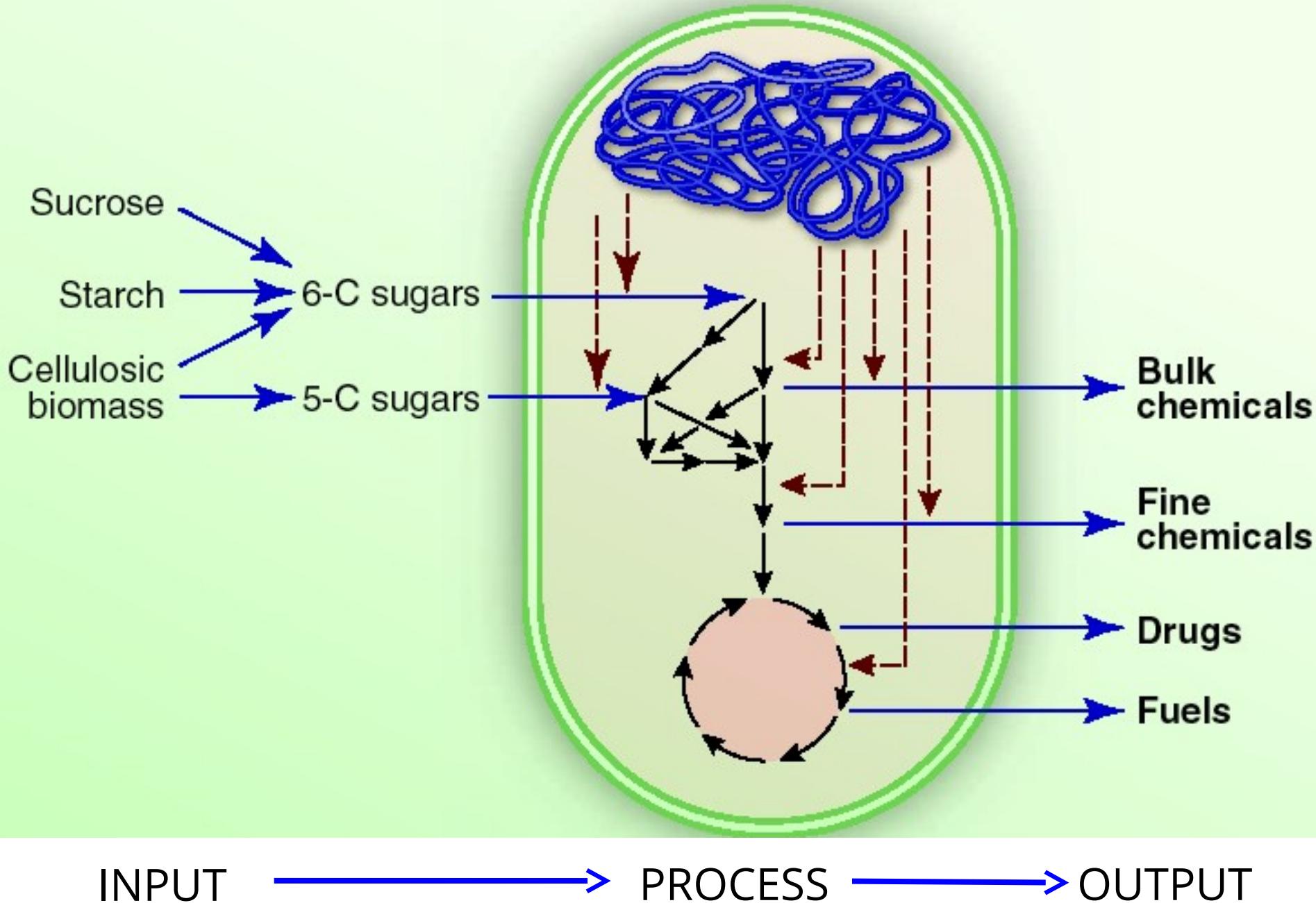
Increase biological production of useful molecules

- Random screening for overproducing strains (genome shuffling)
- Rational engineering of pathways and organisms

Guide to readings

- 1) Metabolic engineering short review 2010.
- 2) Original genome shuffle. A paper reporting a recombination approach for improvement of antibiotic production
- 3) MAGE 2009. A paper describing a method to rapidly introduce genetic variation to targeted regions of the genome
- 4) Synthetic artemisinin. A paper describing metabolic engineering of *S. cerevisiae* to produce the antimalarial compound artemisinin.
- 5) Implications of synthetic genomes (2010). Perspective on the first completely synthesized bacterial genome.
- 6) Biocontainment of genetically modified bacteria (2015)

The cell as a bioreactor



Biologically derived molecules for sale:

- Antibiotics
- Vitamins
- Metabolic by-products (ethanol, lactic acid)
- Amino acids and derivatives (indigo, aspartame)
- “secondary metabolites” from plants – e.g. alkaloids (caffeine, theobromine, etc.)
- Hydrocarbons for fuel
- Synthesis often requires multiple steps and enzymes, making these molecules difficult to synthesize chemically

Increased production of antibiotics: Classical Strain Improvement

- 1) Obtain organism that produces the compound of interest
-- the original strain of *Penicillium* mold made penicillin at micrograms per liter of culture
- 2) Random mutagenesis followed by screen for increased production.
- 3) With top producer, repeat mutagenesis and screen
- 4) Outcome: milligrams of penicillin per liter of culture (1000-fold increase in production)
 - Time consuming and expensive process!

Genome shuffling: an alternative to Classical Strain Improvement

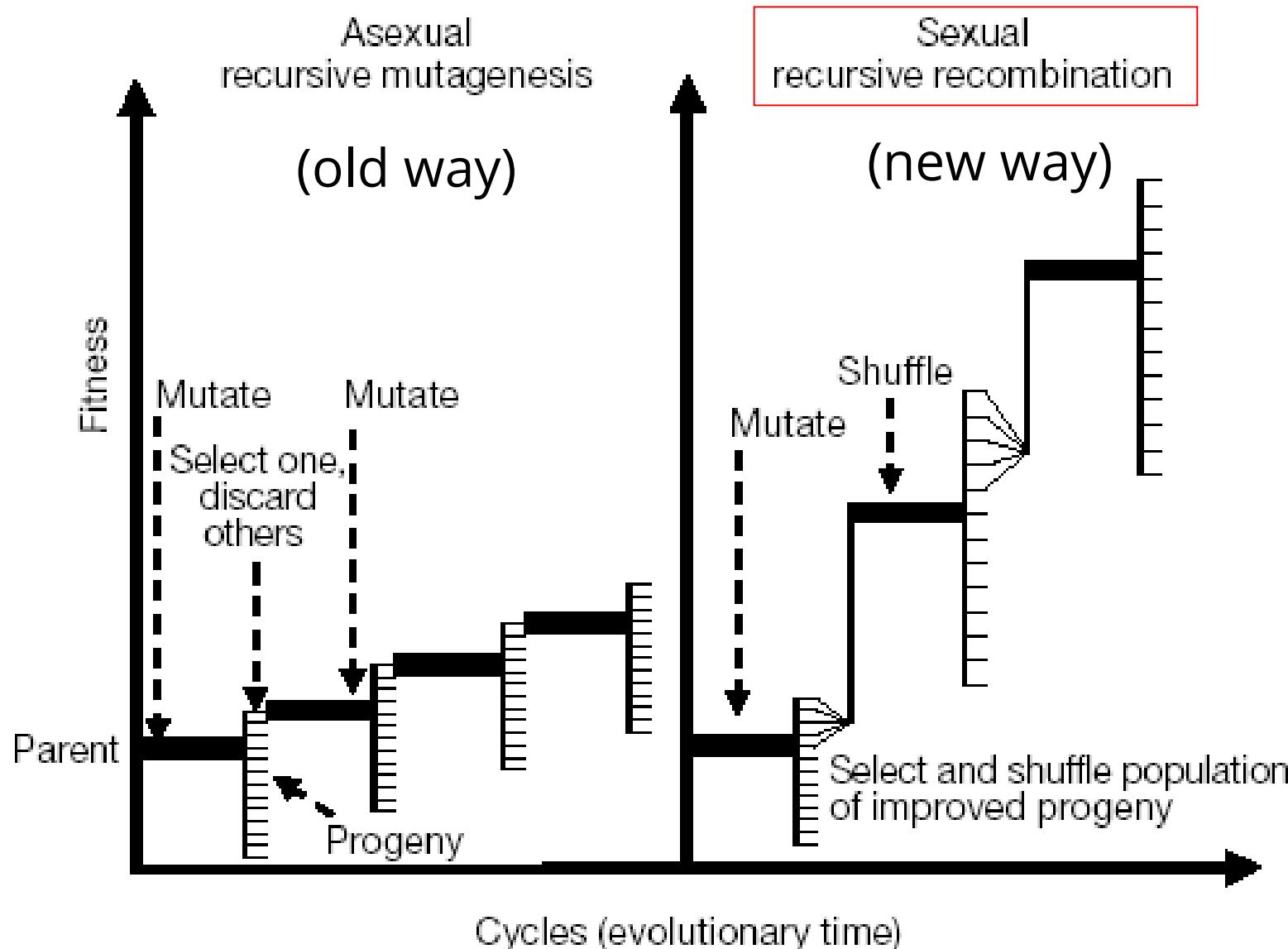
Genome shuffling leads to rapid phenotypic improvement in bacteria

**Ying-Xin Zhang*, Kim Perry*, Victor A. Vinci†, Keith Powell*,
Willem P. C. Stemmer* & Stephen B. del Cardayré***

* Maxygen, 515 Galveston Drive, Redwood City, California 94063, USA

† Eli Lilly and Company, Lilly Corporate Center, Indianapolis, Indiana 46285,
USA

The shuffling advantage: simultaneous recombination of entire genomes (breeding) with multiple parents



Testing recursive shuffling

- Compare classical strain improvement (CSI) to genome shuffling
- *Streptomyces sp.*: produce polyketide antibiotics
- Induce recombination by recursive protoplast fusion:
 - Fuse protoplasts
 - Regenerate cell walls, grow as a population (F1)
 - Make protoplasts with F1, repeat until F4
- Test with 4 auxotrophy markers (next page)
- Test for increased antibiotic production

Test of recursive shuffling

4 parental strains

Supplements required:

Strain	Genotype
<i>S. coelicolor</i> 2684	<i>proA1 argA1 cys^{wt} uraA1</i>
<i>S. coelicolor</i> 2685	<i>proA1 arg^{wt} cysD18 uraA1</i>
<i>S. coelicolor</i> 2686	<i>pro^{wt} argA1 cysD18 uraA1</i>
<i>S. coelicolor</i> M124	<i>proA1 argA1 cysD18 ura^{wt}</i>

pro, arg, ura (not *cys*)

pro, cys, ura (not *arg*)

arg, cys, ura (not *pro*)

pro, arg, cys (not *ura*)

“ Shuffle” (recombine) all 4 strains

Can *progeny* be isolated that can *grow without pro, arg, ura, and cys supplementation* (indicating progeny with all 4 genes wild type)?

Shuffling: increased efficiency of recombination



Table 1 Distribution of phenotypes in a four-strain cross of *S. coelicolor*

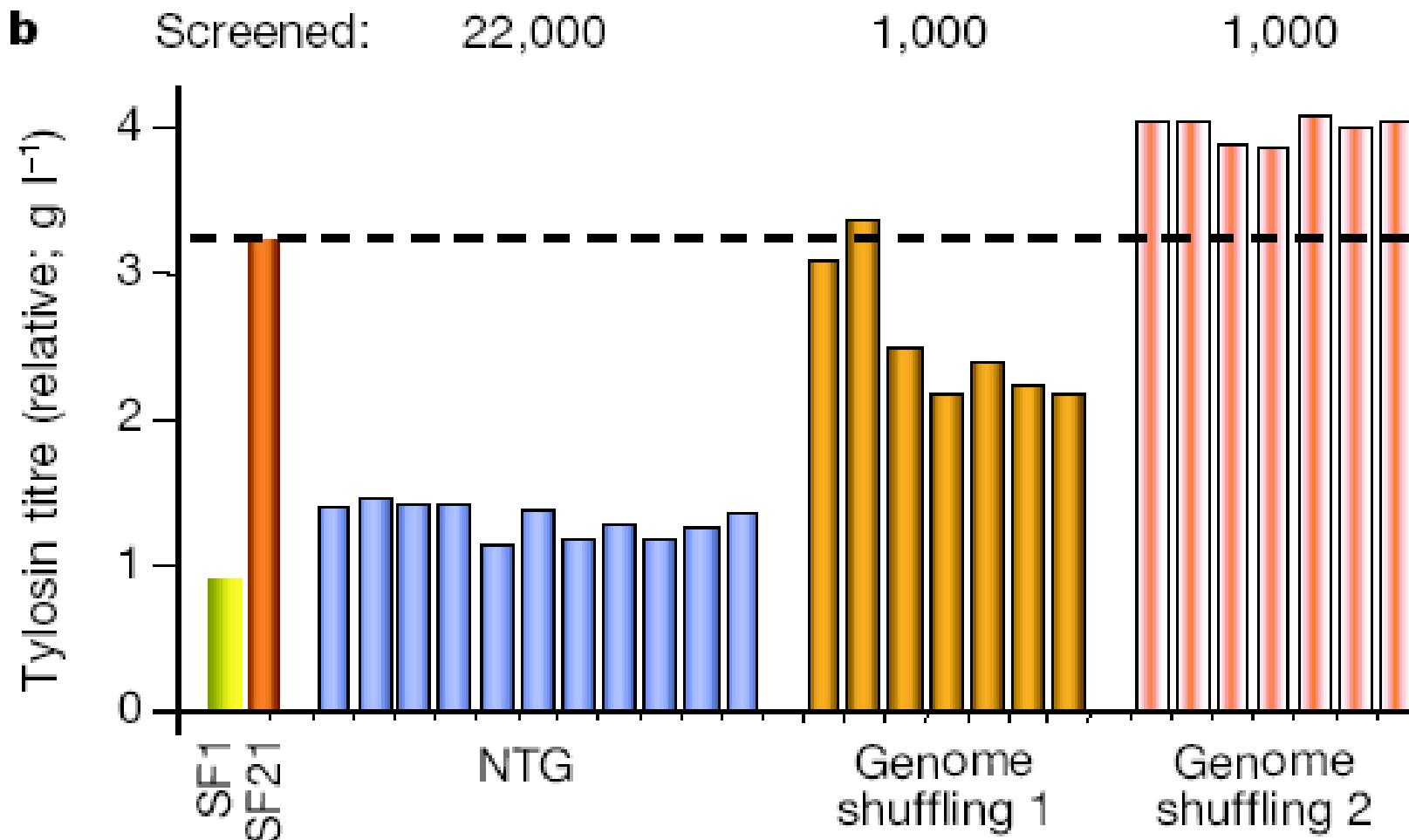
Phenotype	Single fusion*	Recursive fusion†
Two markers	8.4%	60%
Three markers	0.73%	17%
Four markers	0.000045%	2.5%

The distribution of phenotypes from each fusion is reported in Supplementary Information.

* Phenotypes were determined from colony counts on defined medium containing 16 combinations of the four supplements (See Methods). Each phenotype is corrected for dilution and the presence of prototrophic markers, and divided by the total colonies growing on completely supplemented medium. The value shown represents the sum of the frequencies from each phenotypic class.

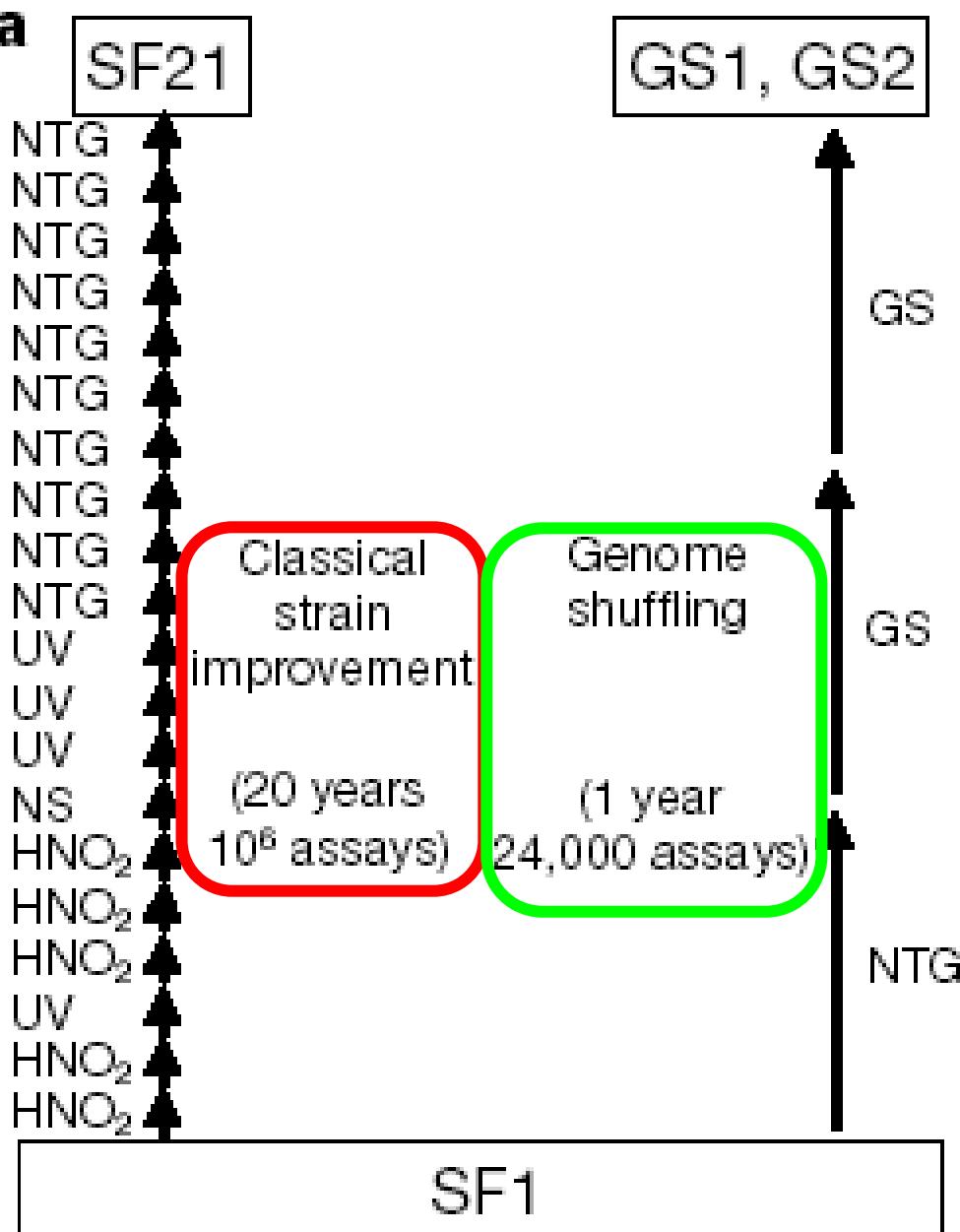
† The distribution of 483 individual colonies characterized for marker phenotype.

Test case: increase tylosin production by *S. fradiae*



SF1 was treated with NTG, 11 strains selected (22,000 screened), those 11 strains were shuffled once (GS1) and then again (GS2)

Comparison: CSI versus genome shuffling

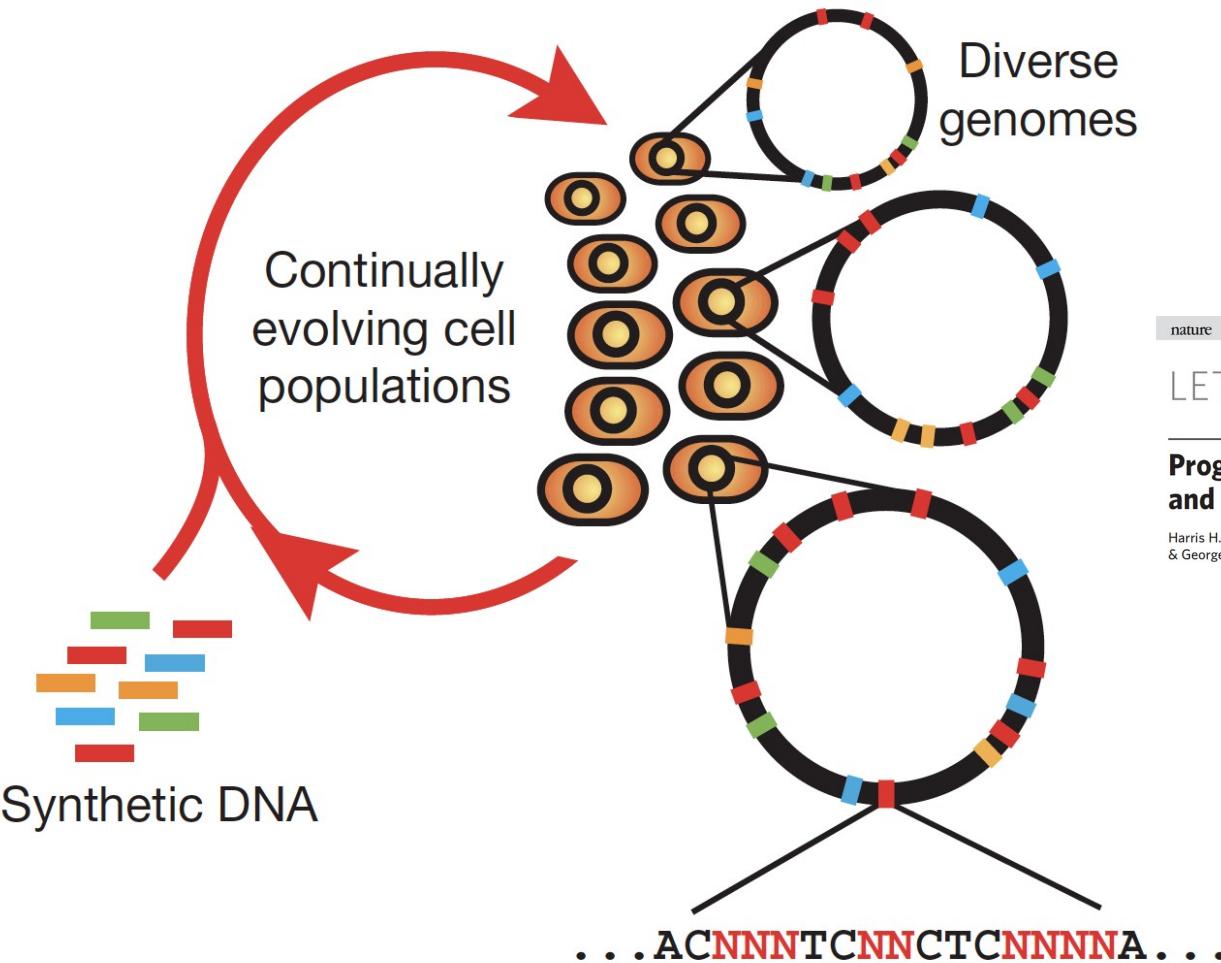


c

Strain	Titre (rel. g l ⁻¹)
SF1	1.0 ± 0.1
SF21	6.2 ± 2.4
GS1	8.1 ± 1.2
GS2	6.2 ± 1.2

Similar results &
much faster with GS

Targeted genome evolution by MAGE (Multiplex Automated Genome Engineering)



Cells are repeatedly transformed with synthetic oligonucleotides that recombine with genome and increase sequence diversity

nature

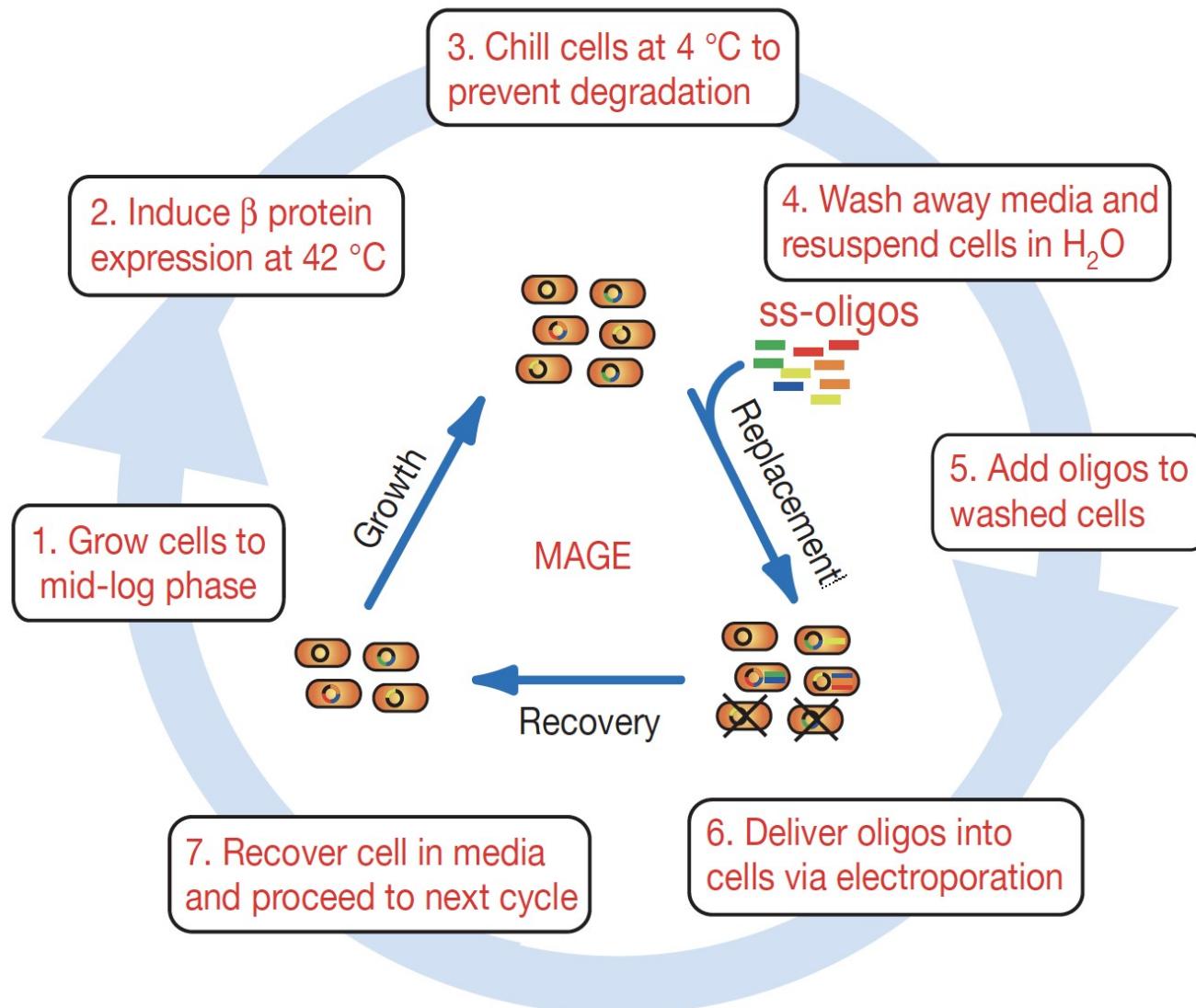
Vol 460 | 13 August 2009 | doi:10.1038/nature08187

LETTERS

Programming cells by multiplex genome engineering and accelerated evolution

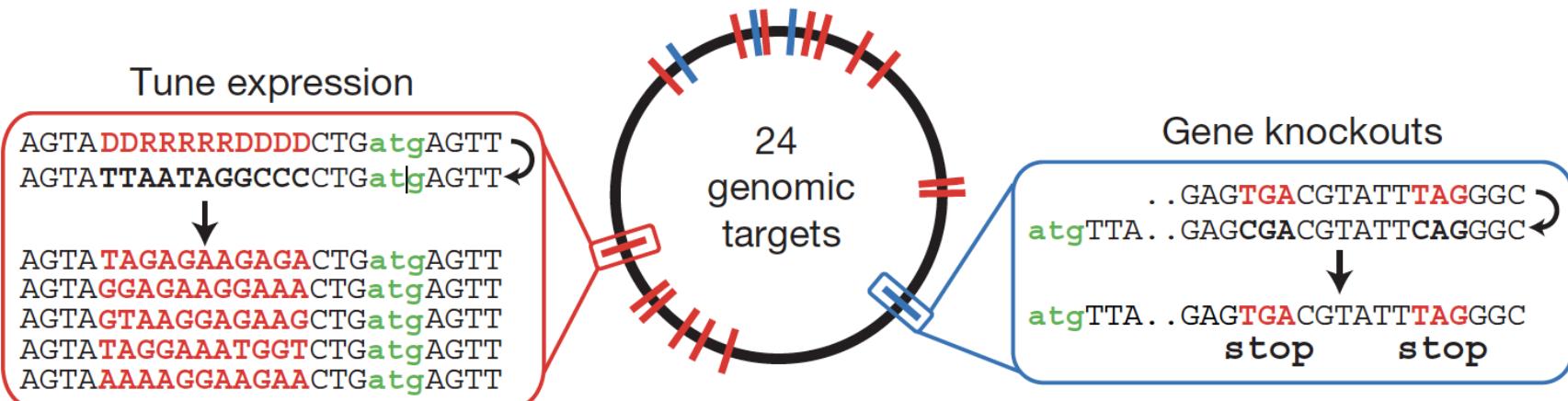
Harris H. Wang^{1,2,3*}, Farren J. Isaacs^{1*}, Peter A. Carr^{4,5}, Zachary Z. Sun⁶, George Xu⁶, Craig R. Forest⁷ & George M. Church¹

One automated cycle (takes ~2.5 hours)



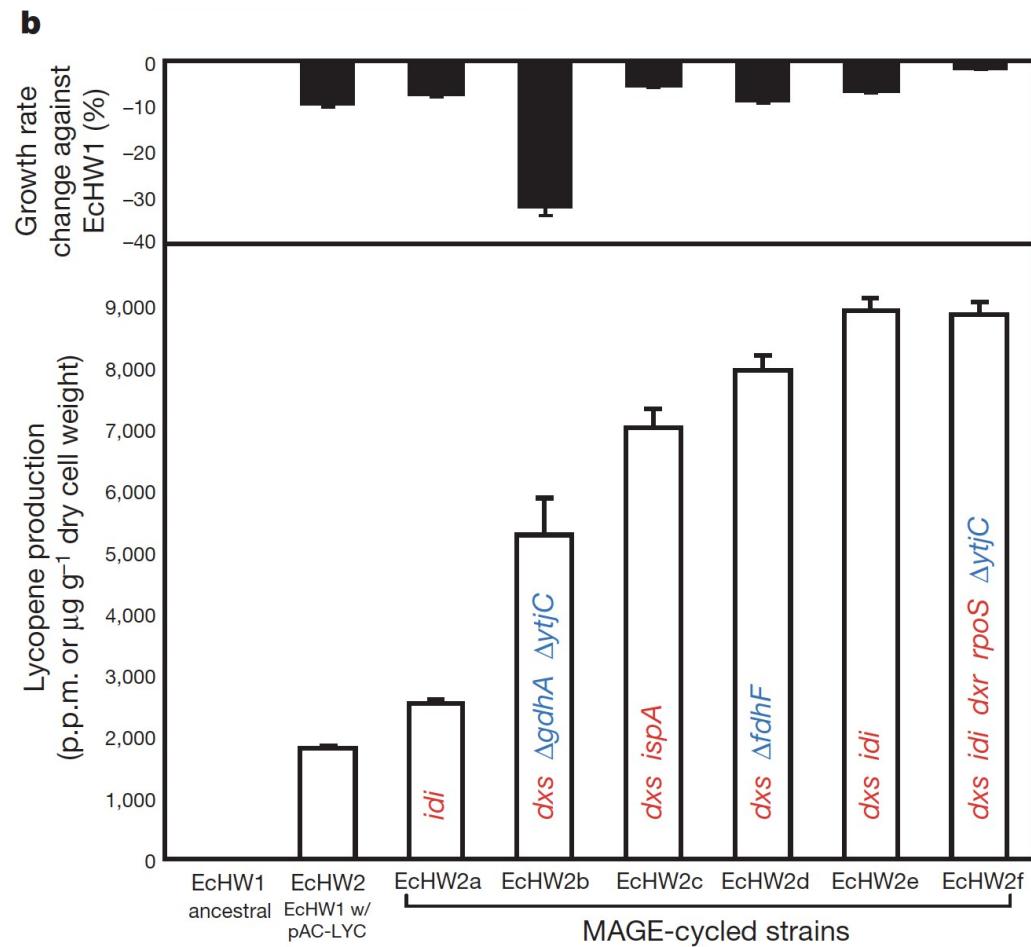
After 5 cycles, measured average of 3.1 base changes per cell

Test of method: lycopene production



- 20 genes in *E. coli* known to enhance lycopene production, and 4 genes known to divert resources away from the lycopene pathway
- DNA was added to randomize the 20 genes' ribosome binding sites, as well as to knock out the diverting genes
- MAGE cycling was done and colonies were selected for analysis on the basis of increased red color (lycopene)

Optimized translation of a subset of the genes in pathway, and knock out of 3 of the 4 diverting genes, provided up to 5X increases in lycopene production



Increasing production of a biological compound: rational design

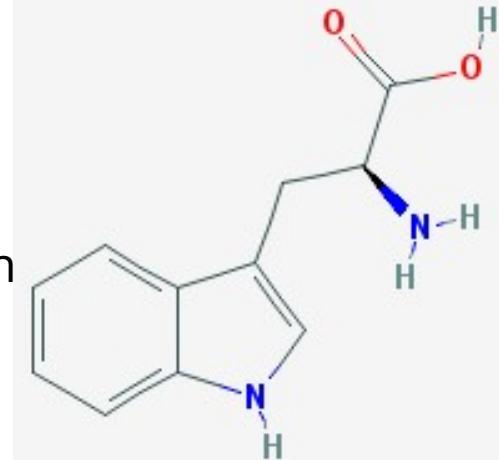
- 1) Increase production of a naturally produced commercial compound
 - Modify existing genes
- 2) Obtain a new organism that can convert an existing compound into a commercial compound
 - Introduce new genes
 - Modify existing genes

natural source of indigo: woad [*Isatis tinctoria*]



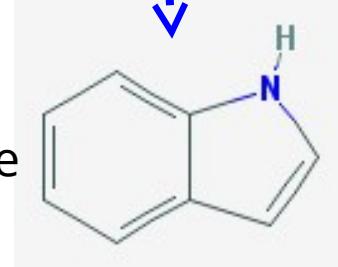
woad

tryptophan



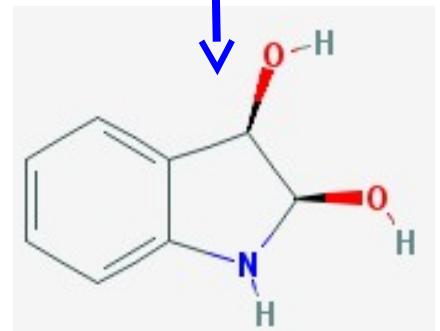
Tryptophanase (*E. coli*)

indole



Naphthalene dioxygenase (cloned)

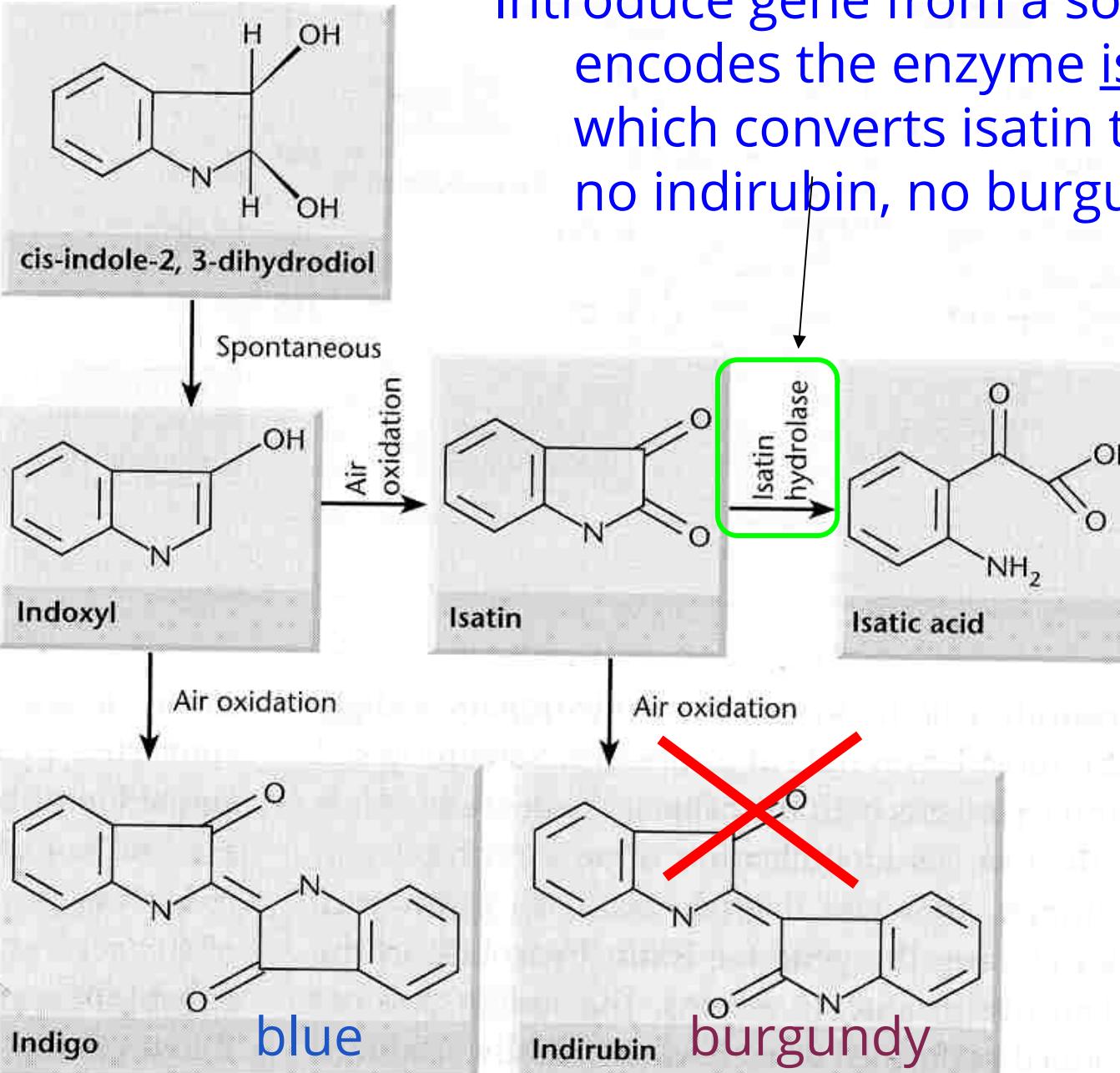
cis-indole-2,3-dihydrodiol



Engineering *E. coli* to produce indigo

- Mutate tryptophan synthase complex to release indole
- Introduce naphthalene dioxygenase (from *Pseudomonas putida*)

Introduce gene from a soil microbe that encodes the enzyme isatin hydrolase which converts isatin to isatic acid (so no indirubin, no burgundy color)

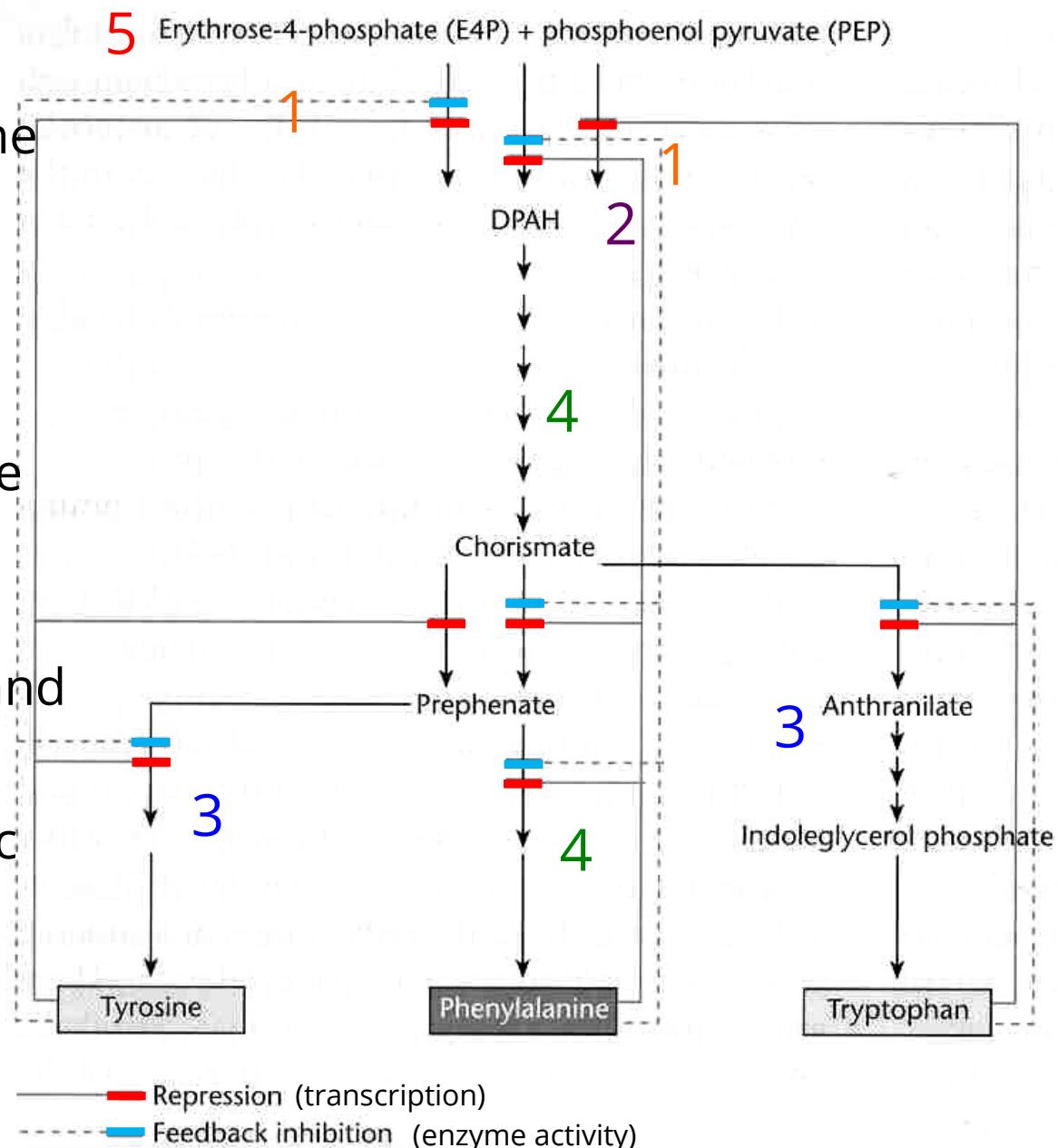


Metabolic engineering: genetic manipulations to maximize synthesis of biological molecules

- Remove transcription, translation repressors
- Defeat enzyme feedback control
- Speed up rate-limiting enzymes
- Block competing pathways
- Funnel carbon to the pathway of interest
- Increase transport of compound out of cell

How to overproduce phenylalanine in *E. coli*

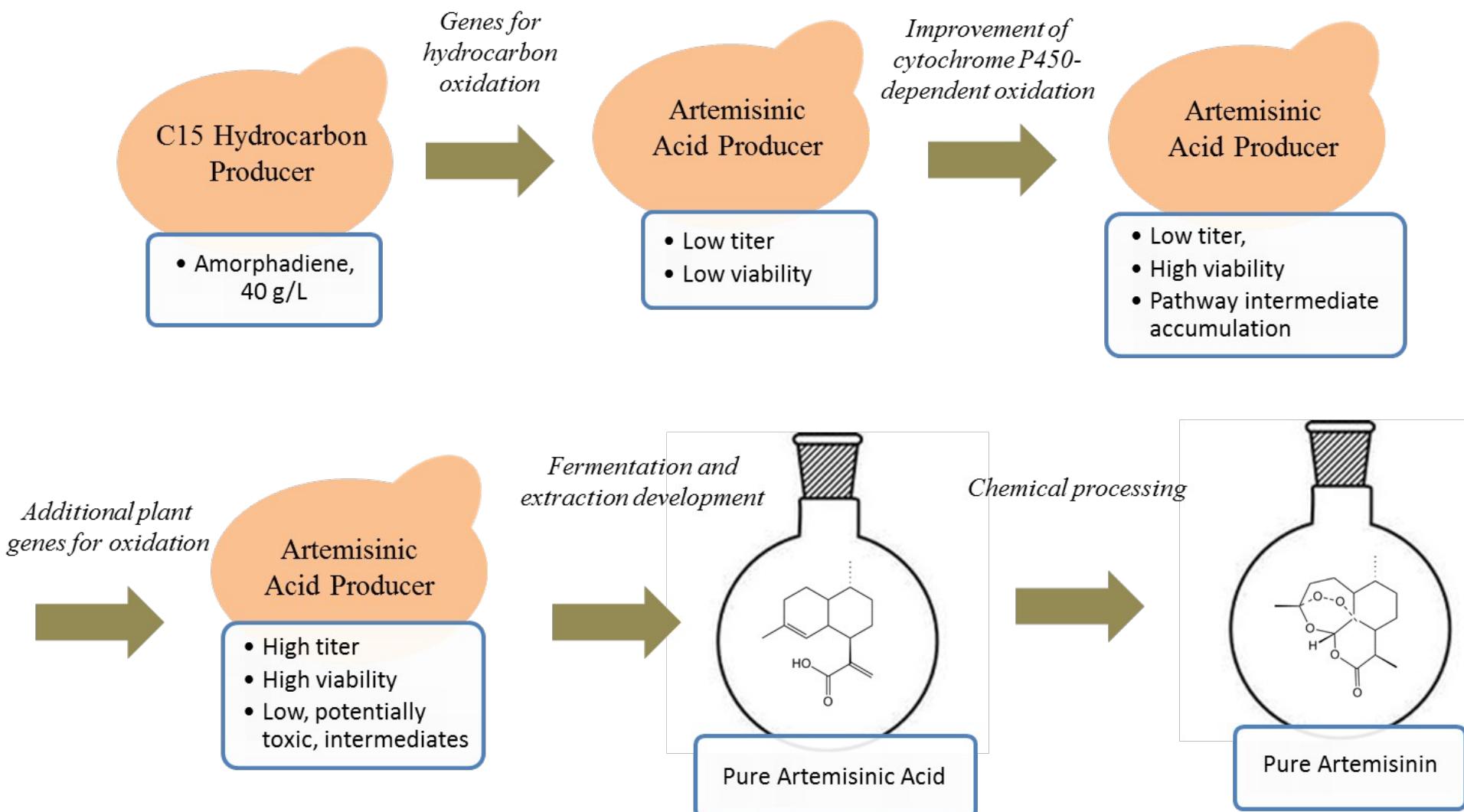
- 1) Remove feedback inhibition (select strains resistant to phenylalanine analogue feedback inhibitors)
- 2) Avoid transcriptional repression (place genes under control of non-phe controlled promoters)
- 3) Remove pathway competition (delete tyr and trp specific genes)
- 4) Overexpress phe-specific genes
- 5) Increase E4P and PEP synthesis



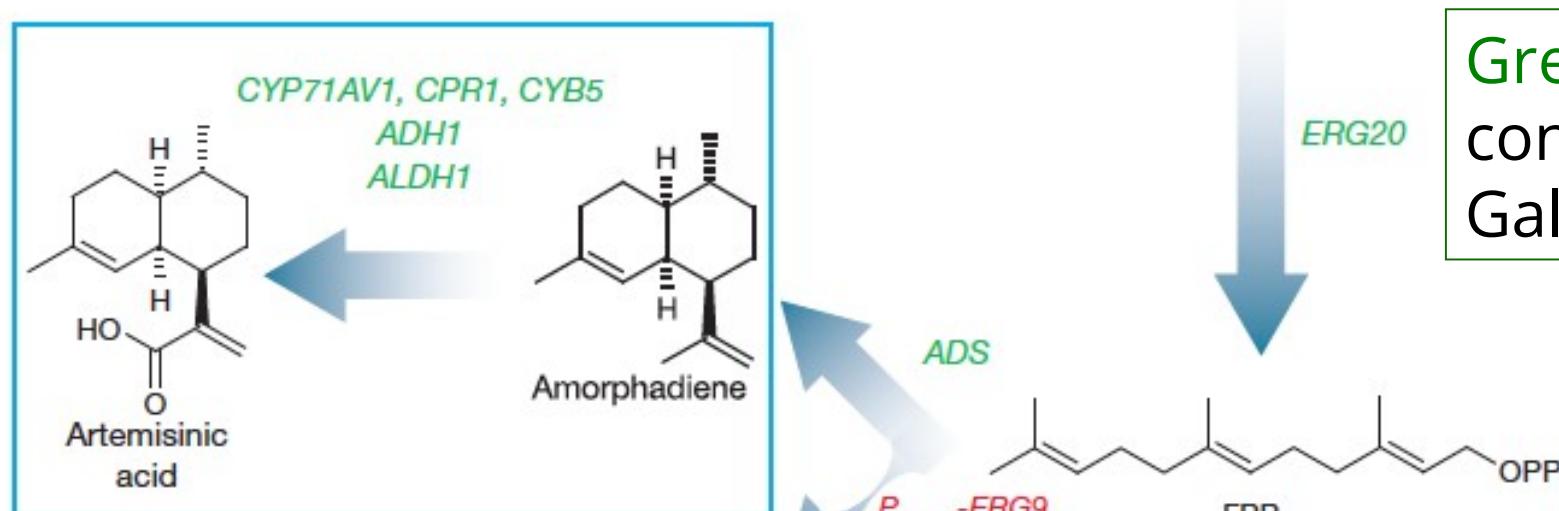
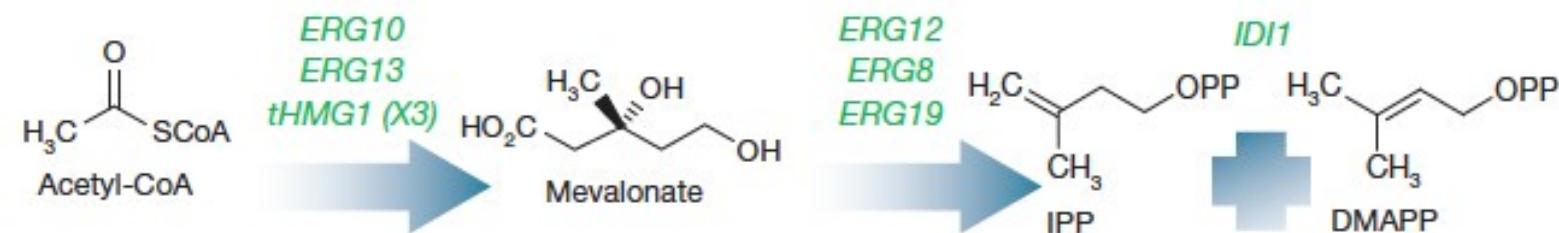
S. cerevisiae engineered to produce artemisinic acid

- Artemisinin is a primary medicine for treating malaria, which caused 660,000 deaths worldwide in 2010
- Production: the plant *Artemisia annua*
- The supply line depends on robust crop yields, so off- years could cause drug shortages
- The “semi-synthetic” production of artemisinin through engineering of yeast was recently reported
- Expected production of artemisinin by this approach: 50-60 tons/year, or 80-150 million doses, produced relatively cheaply

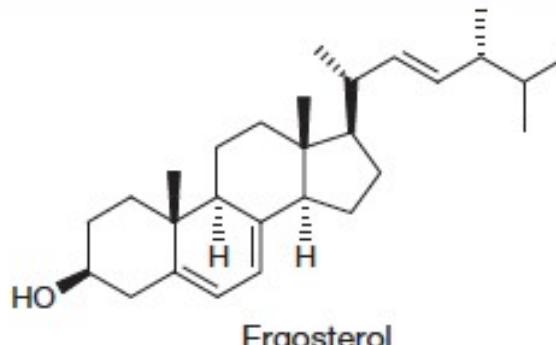
Synthesis of the antimalarial drug artemisinin (originally produced only by the wormwood plant, *Artemisia annua*)



Pathway for synthesis of artemisinic acid

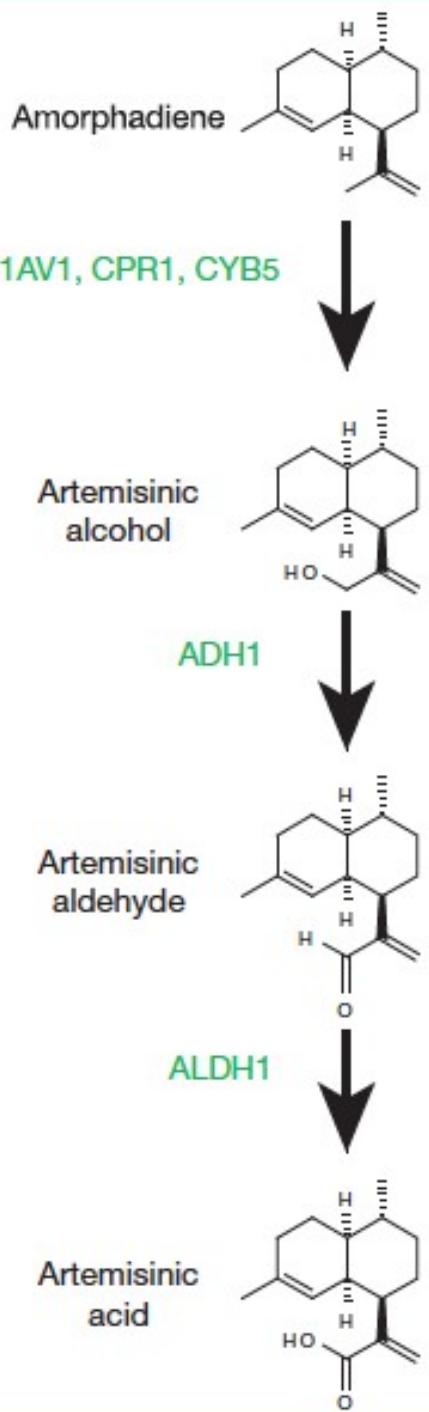


Green genes:
controlled by Gal induction



ERG1
ERG7
ERG11
ERG24
ERG2
ERG25
ERG6
ERG2
ERG3
ERG5
ERG4

Red genes: engineered to
be repressible by addition
of Cu⁺⁺ or methionine



Genes isolated and cloned from *Artemisia annua* were the key

However: farmers who grow *Artemisia* (in Asia and Africa) could lose buyers – can they adapt?

Other potentially disruptive synthetic biology projects on the horizon:

- Vanilla
- Vetiver
- Patchouli
- Rubber
- Coconut
- Saffron
- Opioids

<http://www.guardian.co.uk/global-development/poverty-matters/2013/apr/12/synthetic-malaria-compound-artemisia-farmers>

Rational metabolic engineering

- Requires at least some knowledge of the biochemical pathway required for compound synthesis
- High degree of control
- Trial and error approach can be time consuming
 - try something
 - see if it works
 - find out where the new block to production is
 - change it, too
 - and so on...

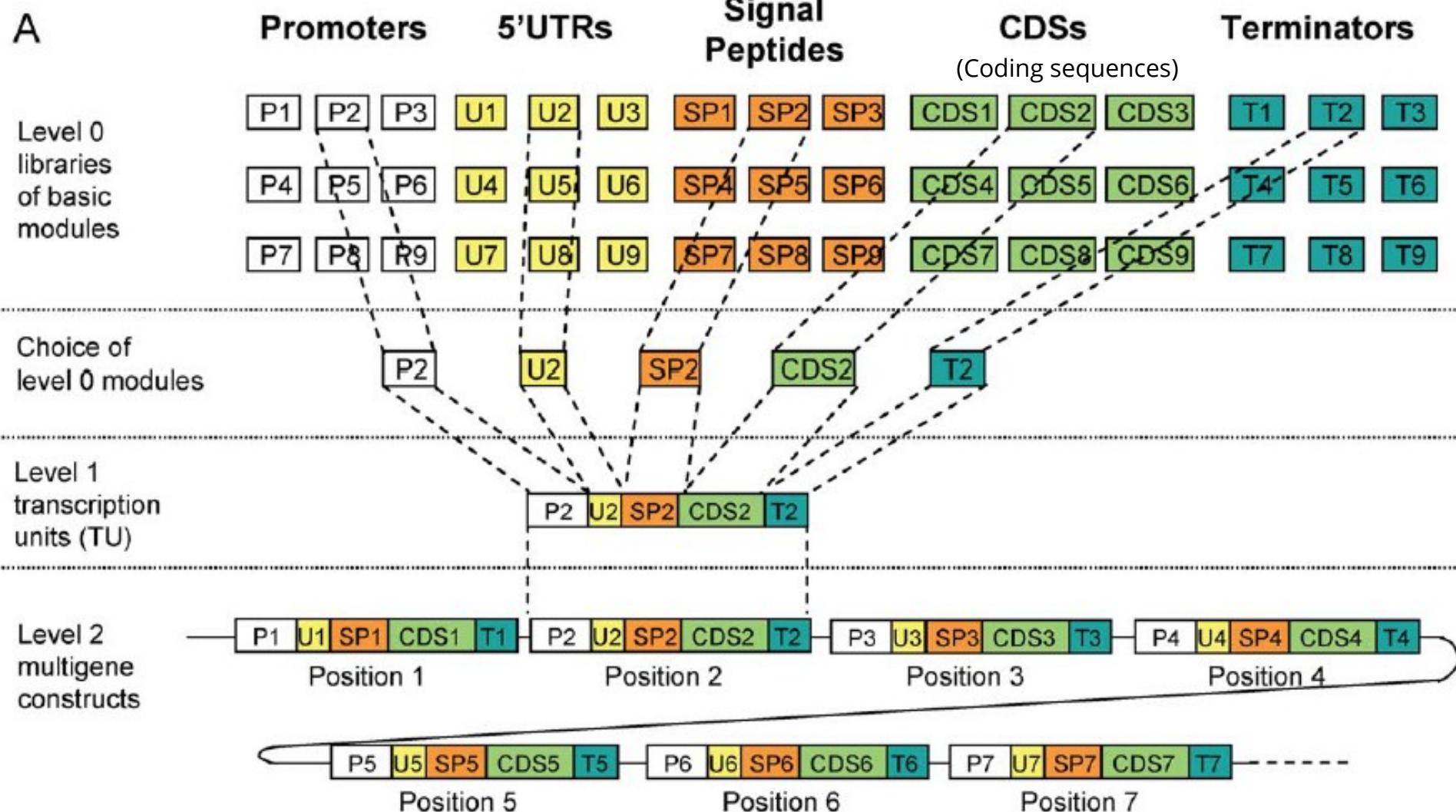
Synthetic biology:

Construction of cellular systems from component parts to reprogram an organism, or to create a new organism

- engineering principles are followed
- synthesis of modules, followed by ordered assembly
- complete, synthetic genome construction is possible

However: it can be difficult to predict how the modules will interact with each other or with the environment in a living system

The biobrick approach: mix and match modules



A Modular Cloning System for Standardized Assembly of Multigene Constructs

PLoS ONE | www.plosone.org

Ernst Weber¹, Carola Engler¹, Ramona Gruetzner, Stefan Werner, Sylvestre Marillonnet^{*}

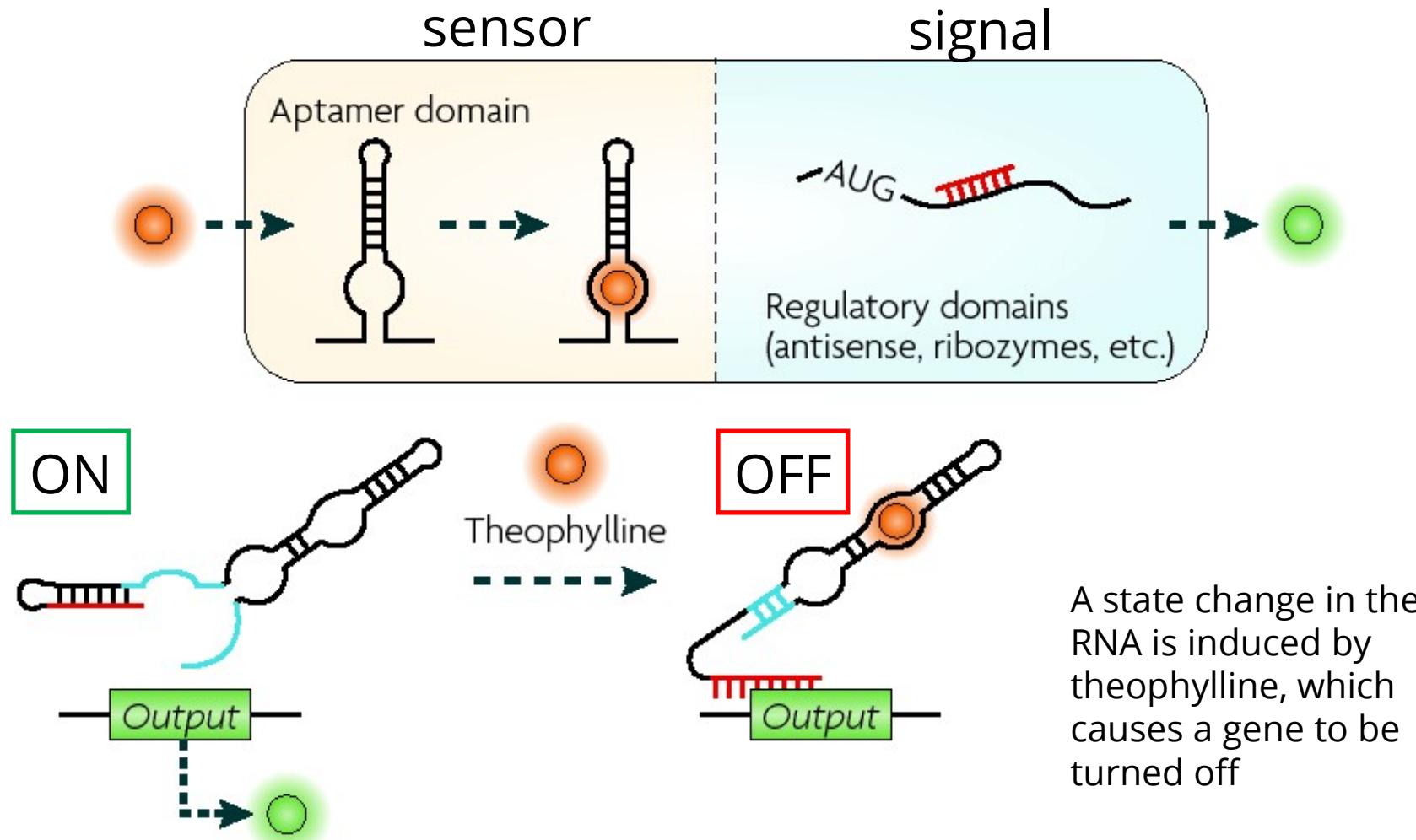
Icon Genetics GmbH, Halle/Saale, Germany

February 2011 | Volume 6 | Issue 2 | e16765

Synthetic biology module:

An engineered genetic element that performs a specific task

BIOSENSOR: a genetic control module. Presence of a small molecule shuts off signal



Registry of biological parts:

http://parts.igem.org/Main_Page

Registry of Standard Biological Parts



tools catalog repository assembly protocols help search

BBA_



iGEM 2020: An exceptional year

This is not a normal year. Together, we are facing the biggest pandemic in the past 100 years. Based on your feedback, we are adapting the iGEM experience - we have examined each part of the competition to make it even better this year. This is iGEM in the time of a pandemic. It will be different but it will be worth it.

[See the full announcement here](#)

Take a look at some of the changes on our [New for 2020](#) hub.

Add and Document Parts

Start [adding and documenting](#) your parts now! Your parts should be well characterized and measured, and follow the Registry's requirements.

Sample Submissions

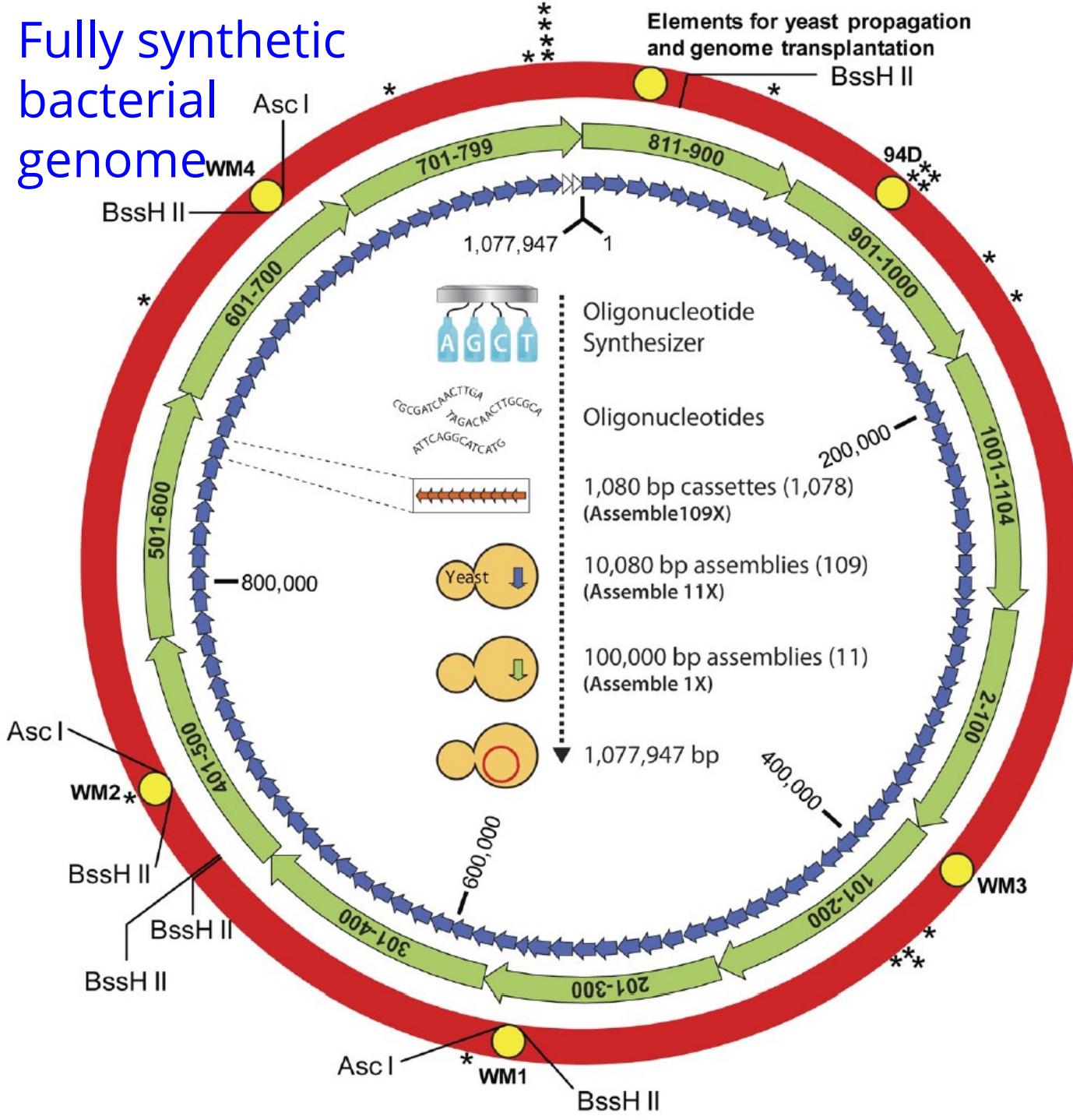
iGEM teams do not need to send samples of their parts for 2020. We want teams to focus on the documentation of their parts! Teams must follow 2020 requirements for parts, including [BioBrick RFC10](#) or [Type IIS compatibility](#).

Update: Distribution Kit

As a result of the current pandemic, we will not be able to manufacture and ship our [2020 DNA Distribution Kits](#) to teams and labs this year. We sincerely apologize for any inconvenience this may cause.

iGEM: International Genetically Engineered Machine

Fully synthetic bacterial genome



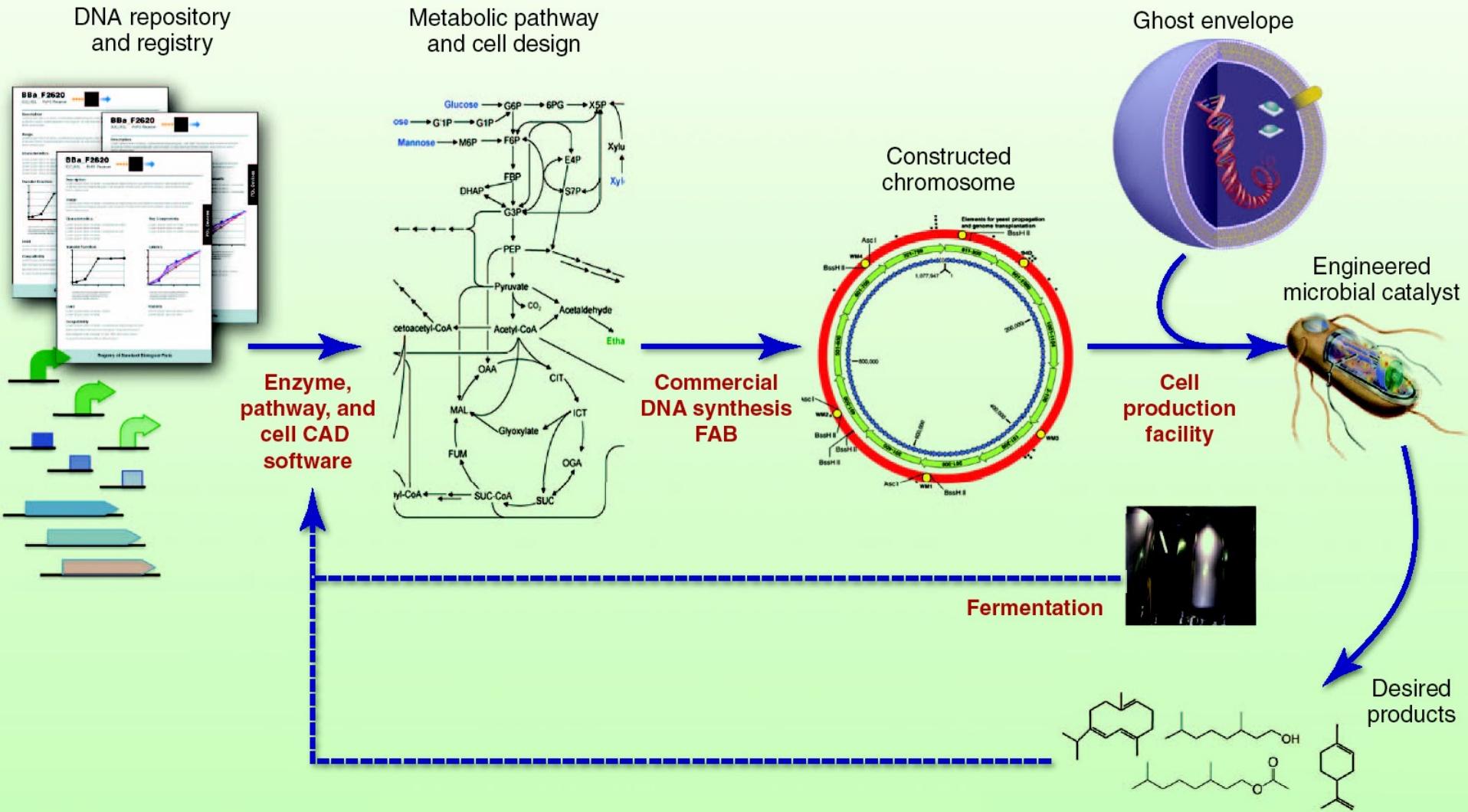
1080 bp synthetic cassettes assembled into sets (10 per set, 109 sets total) by recombination in yeast

10,800 bp fragments assembled into sets (10 per set, 11 sets total)

11 ~100 kb fragments assembled into complete genome

~ \$40M
20 people
>10 years

The future of bioengineering?



Risks? How to manage them?

How do synthetic life forms interact with various potential environments?

How long do synthetic organisms persist, and how well do they compete with non-synthetic organisms?

How quickly can the synthetic organism evolve?

Can synthetic organisms transfer their DNA to natural organisms and vice versa?

“Once released, synthetic organisms cannot be retrieved”

Genetic engineering for bio-containment

- Make the organism dependent on one or more non-canonical amino acids
 - Genetic code is re engineered
 - Requires feeding with non-canonical amino acid
 - Genetic code differs from other organisms: genetic exchange can't readily occur
- Make the organism dependent on added small molecules not found in nature
 - 'riboregulators' based on riboswitches controlling many essential genes in an organism
 - Addiction modules encode nucleases that kill the cell if the antidote isn't made, and the antidote is controlled by a small molecule

Applied mutagenesis: pathway engineering and synthetic biology

Increase biological production of useful molecules

- Random screening for overproducing strains (genome shuffling)
- Rational engineering of pathways and organisms