LOSCHMIDT LABORATORIES

Advanced molecular biology tools

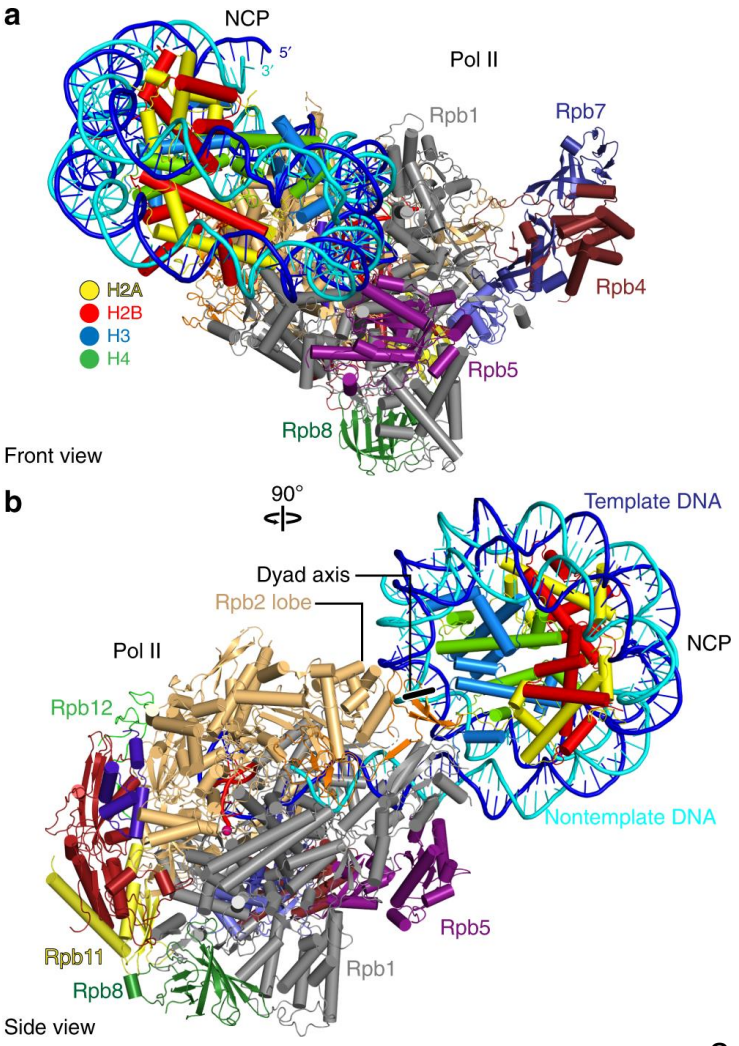
Molecular cloning and gene assembly in biotechnology, biomedicine and basic research

Martin Marek
Loschmidt Laboratories
Faculty of Science, Masaryk University
Kamenice 5, bld. C13, room 332
martin.marek@recetox.muni.cz

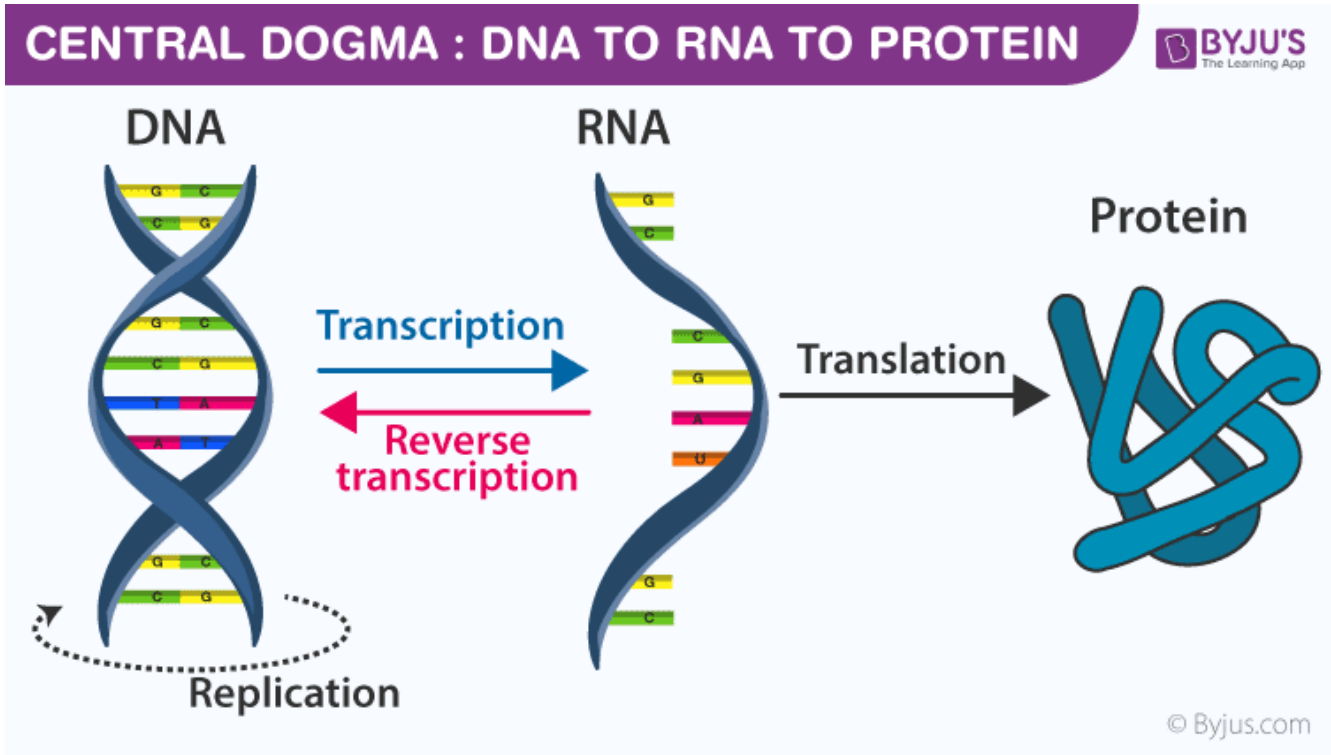
Molecular Biotechnology (2021)

What is molecular biology?

The molecular biology studies biological macromolecules and the molecular mechanisms found in living systems, such as the molecular nature of the gene and its mechanisms of gene replication, mutation and expression.

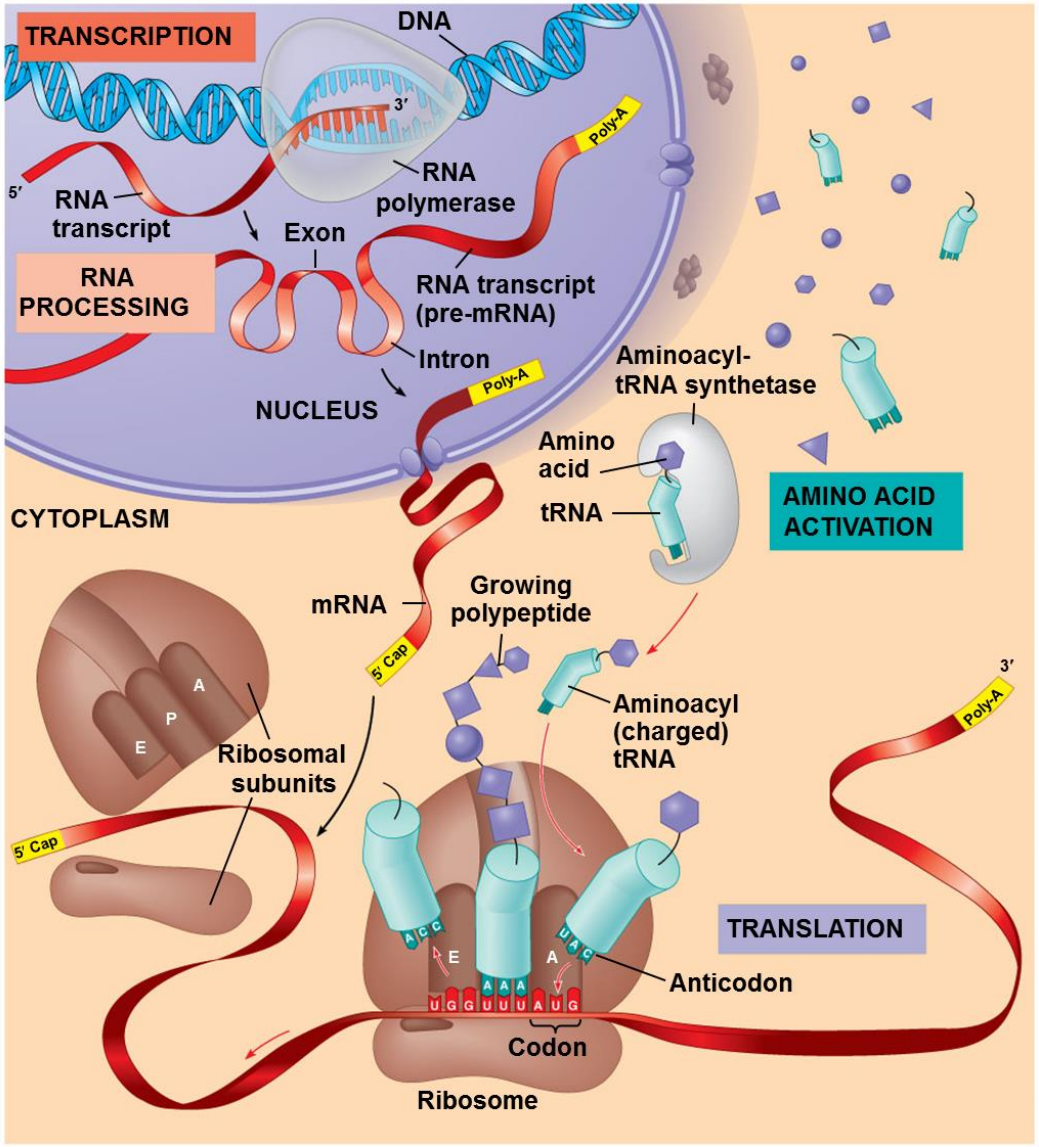


Central dogma of molecular biology



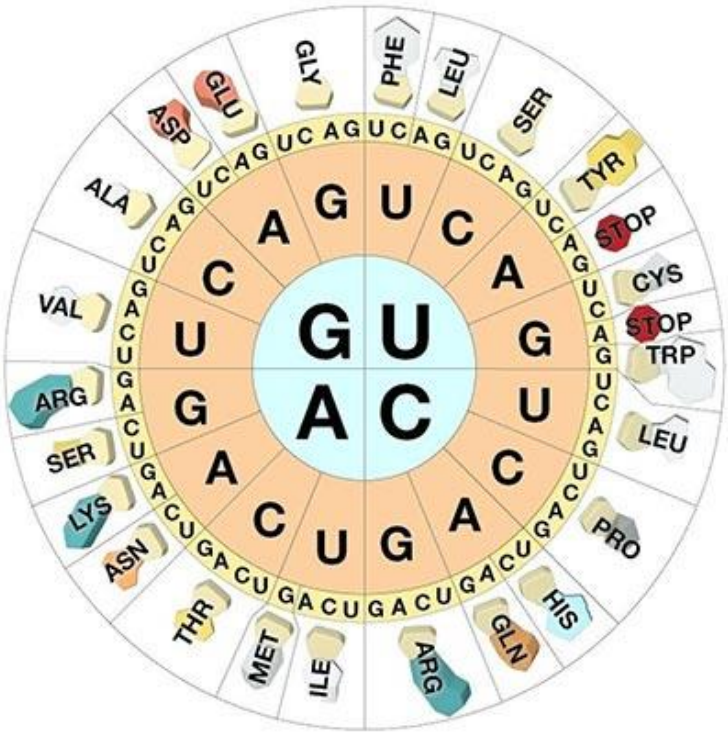
- The central dogma of molecular biology states that DNA contains instructions for making a protein, which are copied by RNA.
- RNA then uses the instructions to make a protein.
- In short: DNA → RNA → Protein, or DNA to RNA to Protein.

How DNA directs protein synthesis



Redundancy of the genetic code

- Degeneracy of codons is the redundancy of the genetic code, exhibited as the multiplicity of three-base pair codon combinations that specify an amino acid
- The genetic code is degenerate mainly at the third codon position
- The genetic code consists of 64 triplet codons specifying 20 canonical amino acids and 3 stop signals



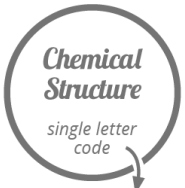
Second Letter

First Letter	T	C	A	G	Third Letter
T	TTT Phe, TTC Phe, TTA Leu, TTG Leu	TCT Ser, TCC Ser, TCA Ser, TCG Ser	TAT Tyr, TAC Tyr, TAA Stop, TAG Stop	TGT Cys, TGC Cys, TGA Stop, TGG Trp	T C A G
C	CTT Leu, CTC Leu, CTA Leu, CTG Leu	CCT Pro, CCC Pro, CCA Pro, CCG Pro	CAT His, CAC His, CAA Gln, CAG Gln	CGT Arg, CGC Arg, CGA Arg, CGG Arg	T C A G
A	ATT Ile, ATC Ile, ATA Ile, ATG Met	ACT Thr, ACC Thr, ACA Thr, ACG Thr	AAT Asn, AAC Asn, AAA Lys, AAG Lys	AGT Ser, AGC Ser, AGA Arg, AGG Arg	T C A G
G	GTT Val, GTC Val, GTA Val, GTG Val	GCT Ala, GCC Ala, GCA Ala, GCG Ala	GAT Asp, GAC Asp, GAA Glu, GAG Glu	GGT Gly, GGC Gly, GGA Gly, GGG Gly	T C A G

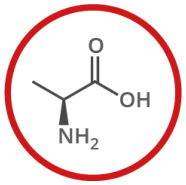
A GUIDE TO THE TWENTY COMMON AMINO ACIDS

AMINO ACIDS ARE THE BUILDING BLOCKS OF PROTEINS IN LIVING ORGANISMS. THERE ARE OVER 500 AMINO ACIDS FOUND IN NATURE - HOWEVER, THE HUMAN GENETIC CODE ONLY DIRECTLY ENCODES 20. 'ESSENTIAL' AMINO ACIDS MUST BE OBTAINED FROM THE DIET, WHILST NON-ESSENTIAL AMINO ACIDS CAN BE SYNTHESISED IN THE BODY.

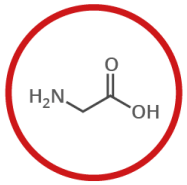
Chart Key: ALIPHATIC · AROMATIC · ACIDIC · BASIC · HYDROXYLIC · SULFUR-CONTAINING · AMIDIC · NON-ESSENTIAL · ESSENTIAL



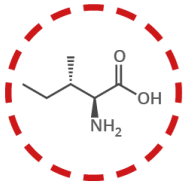
NAME (A)
three letter code
DNA codons



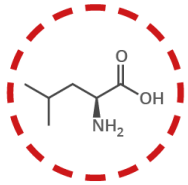
ALANINE (A)
Ala
GCT, GCC, GCA, GCG



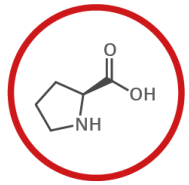
GLYCINE (G)
Gly
GGT, GGC, GGA, GGG



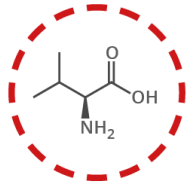
ISOLEUCINE (I)
Ile
ATT, ATC, ATA



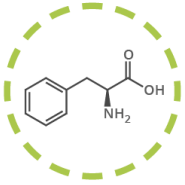
LEUCINE (L)
Leu
CTT, CTC, CTA, CTG, TTA, TTG



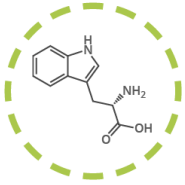
PROLINE (P)
Pro
CCT, CCC, CCA, CCG



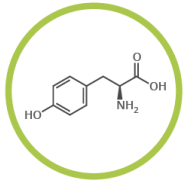
VALINE (V)
Val
GTT, GTC, GTA, GTG



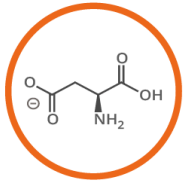
PHENYLALANINE (F)
Phe
TTT, TTC



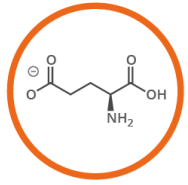
TRYPTOPHAN (W)
Trp
TGG



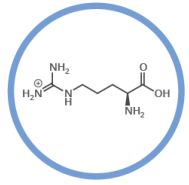
TYROSINE (Y)
Tyr
TAT, TAC



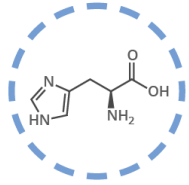
ASPARTIC ACID (D)
Asp
GAT, GAC



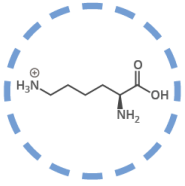
GLUTAMIC ACID (E)
Glu
GAA, GAG



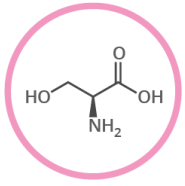
ARGININE (R)
Arg
CGT, CGC, CGA, CGG, AGA, AGG



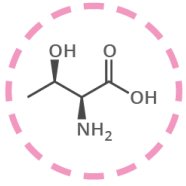
HISTIDINE (H)
His
CAT, CAC



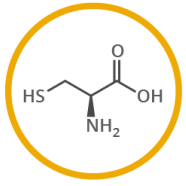
LYSINE (K)
Lys
AAA, AAG



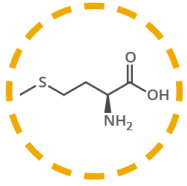
SERINE (S)
Ser
TCT, TCC, TCA, TCG, AGT, AGC



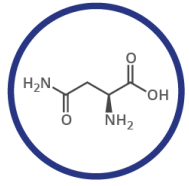
THREONINE (T)
Thr
ACT, ACC, ACA, ACG



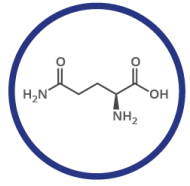
CYSTEINE (C)
Cys
TGT, TGC



METHIONINE (M)
Met
ATG



ASPARAGINE (N)
Asn
AAT, AAC

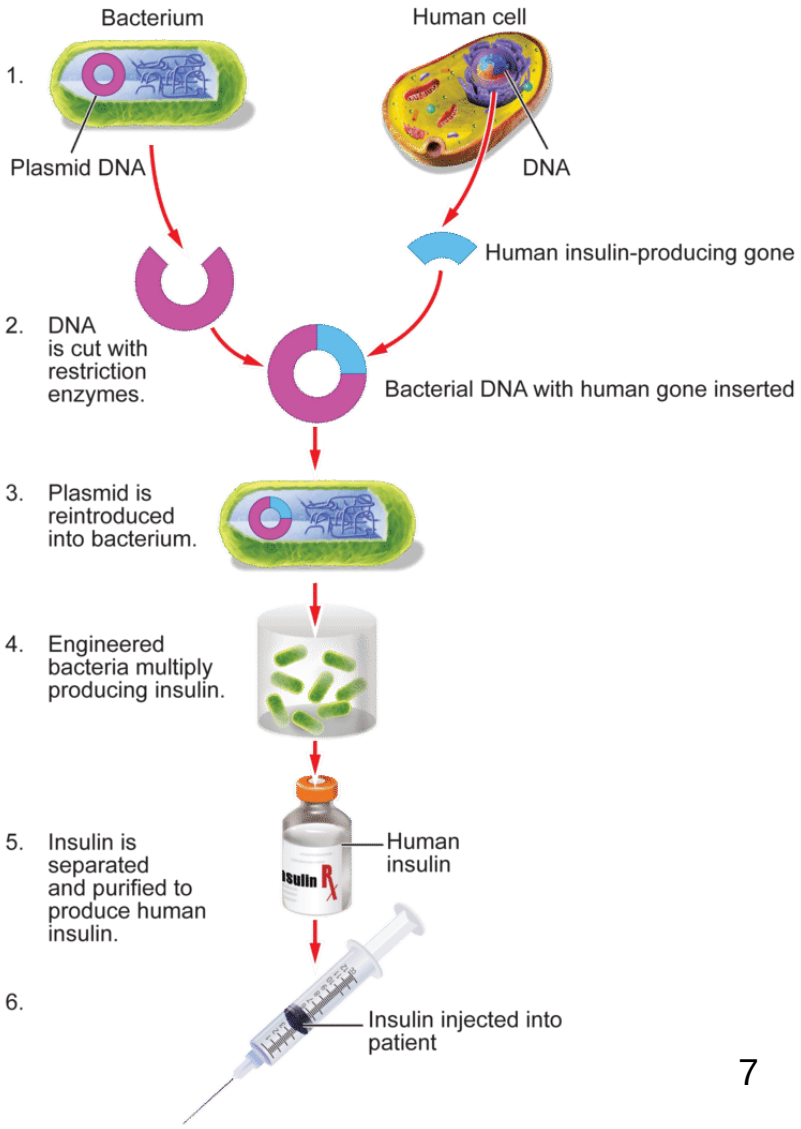


GLUTAMINE (Q)
Gln
CAA, CAG

Note: This chart only shows those amino acids for which the human genetic code directly codes for. Selenocysteine is often referred to as the 21st amino acid, but is encoded in a special manner. In some cases, distinguishing between asparagine/aspartic acid and glutamine/glutamic acid is difficult. In these cases, the codes asx (B) and glx (Z) are respectively used.

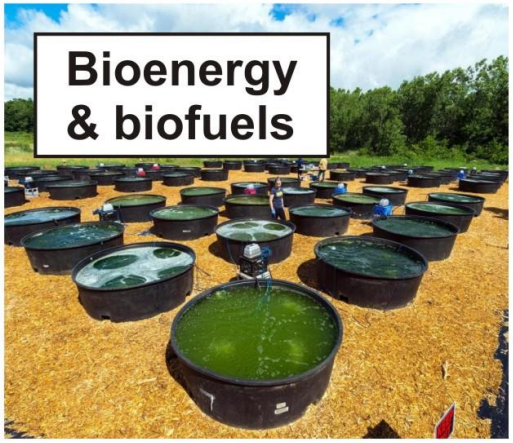
What is genetic engineering?

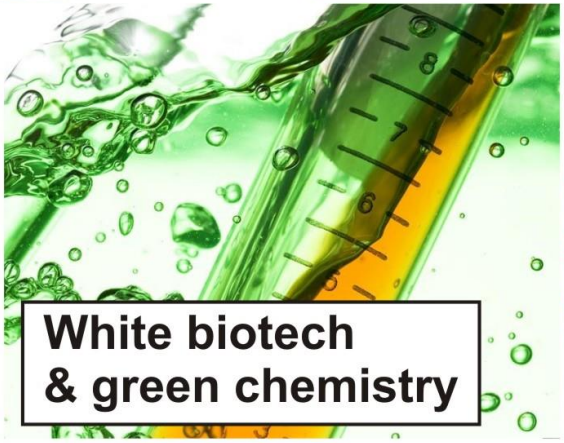
- Genetic engineering is the action to modify the genetic information present in a living cell
- Adding, substituting or removing a genetic information in a given biological system necessarily implies to physically introduce a new information into the target cell. DNA is the physical support of genetic information.
- Therefore, genetic engineering relies on the generation of artificial DNA molecules, containing the information of interest, so it can be transferred to the target cells
- In molecular biotechnologies, genetic engineering represents a process of taking a gene from one species and putting it into another species



Applications of molecular biotechnologies

Why should we bother to think about molecular biotechnologies?





Molecular biology in rational (computational) protein design

Computational protein design

1. Gene synthesis & cloning

- *De novo* DNA synthesis
- Gene assembly
- Site-directed mutagenesis

2. Protein expression & purification

- Expression host
- Affinity and solubility tags
- Buffer screening

3. Biophysical & biochemical characterization

- CD spectroscopy
- Differential scanning fluorimetry
- Dynamic light scattering
- SAXS

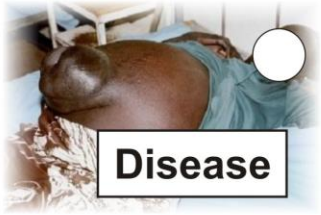
4. 3D structure determination

- NMR spectroscopy
- Cryo-EM
- X-ray crystallography

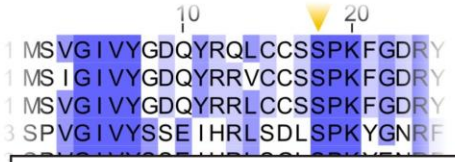
5. Biological assays

- Biocatalysis
- Binding assays
- Protein-protein interaction
- Photoactivation

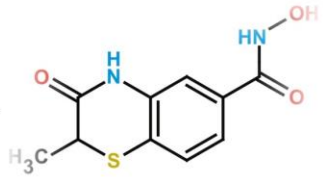
The key role of protein production in pharma industry



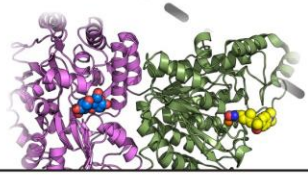
Disease



Identify target
(Genomics, biological assays)

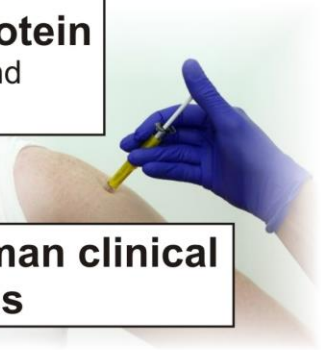


J1038



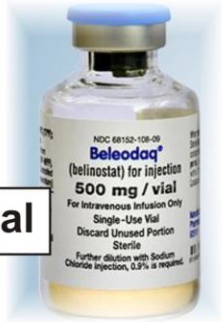
Isolate protein involved in disease
(Molecular cloning, recombinant expression, biochemical and structural analyses)

Find a drug effective against disease protein
(virtual screening, HTS and library screening etc.)



Human clinical trials

Preclinical testing
(Animal studies, scale-up, formulation)



FDA approval

Protein production is
a significant bottleneck in early
phase drug discovery

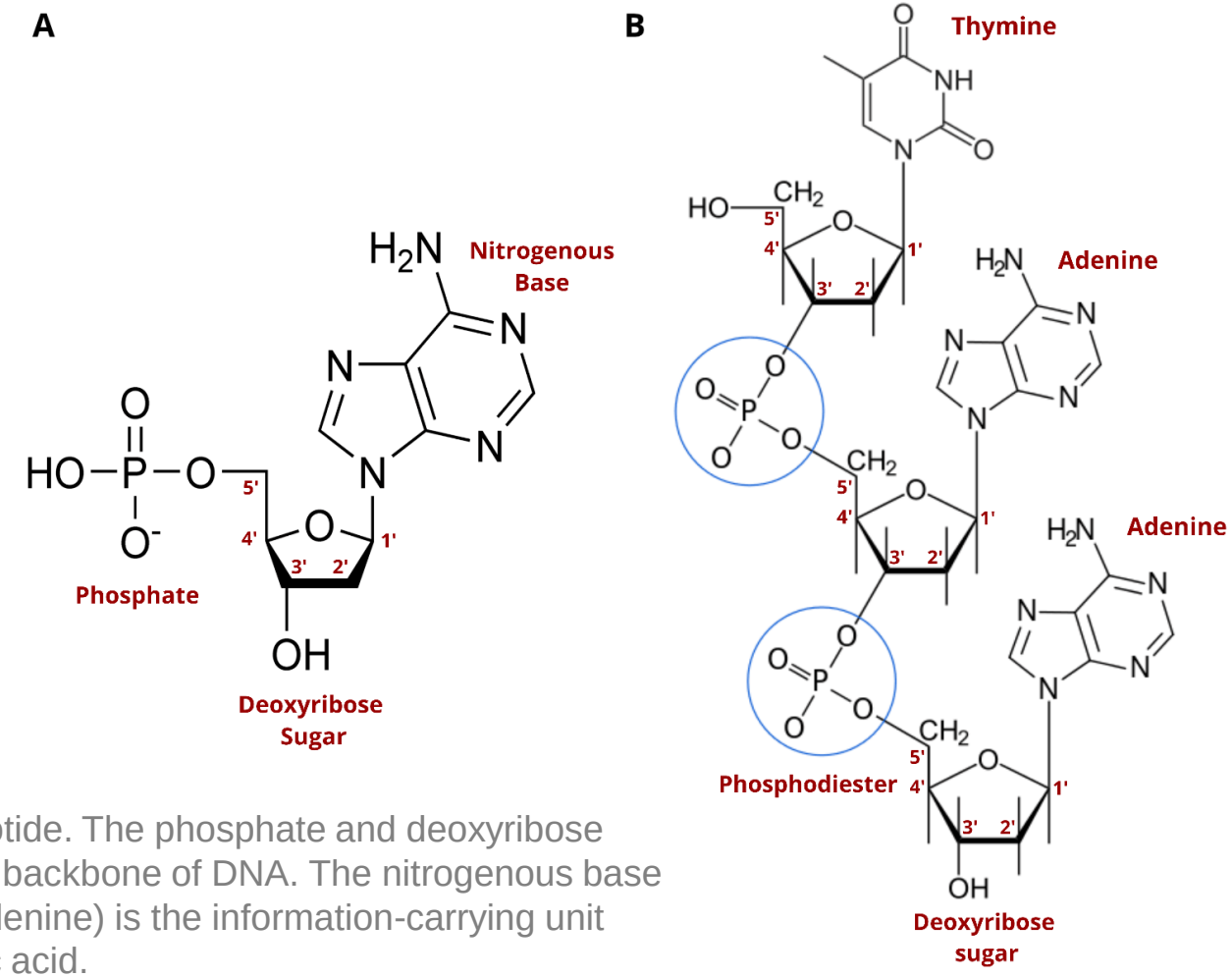
Recombinant protein production workflow

- Gene synthesis and molecular cloning
- Protein expression
- Protein purification
- Protein characterization
- (Protein structure determination)

DNA synthesis & molecular cloning

Concepts
Methods
Applications

Structure of DNA



A) A single nucleotide. The phosphate and deoxyribose sugar form the backbone of DNA. The nitrogenous base (in this case adenine) is the information-carrying unit of each nucleic acid.

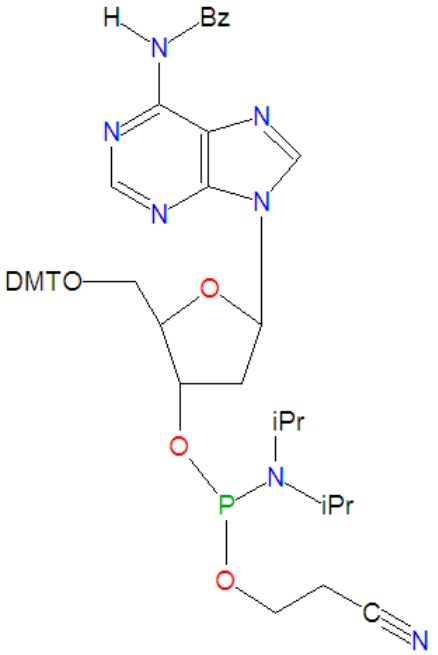
B) The structure of single-stranded DNA. In nature, enzymes form phosphodiester bonds (blue circles) that link the 5th position and 3rd position of adjacent deoxyribose sugars. Due to the modular nature of nucleotides, this chain can grow indefinitely.

Chemical synthesis of DNA

- Making DNA chemically rather than biologically was one of the first new technologies to be applied by the biotechnology industry. The ability to make short synthetic stretches of DNA is crucial to using DNA replication in laboratory techniques. DNA polymerase cannot synthesize DNA without a free 3′-OH end to elongate. Therefore, to use DNA polymerase *in vitro*, the researcher must supply a short primer. Such primers are used to sequence DNA, to amplify DNA with PCR, to introduce DNA mutations, and even to find genes in library screening.

- Technically, oligonucleotides are any piece of DNA approx. 20 nucleotides in length, but today, oligonucleotide denotes a short piece of DNA (approx. 125 nucleotides) that is chemically synthesized.

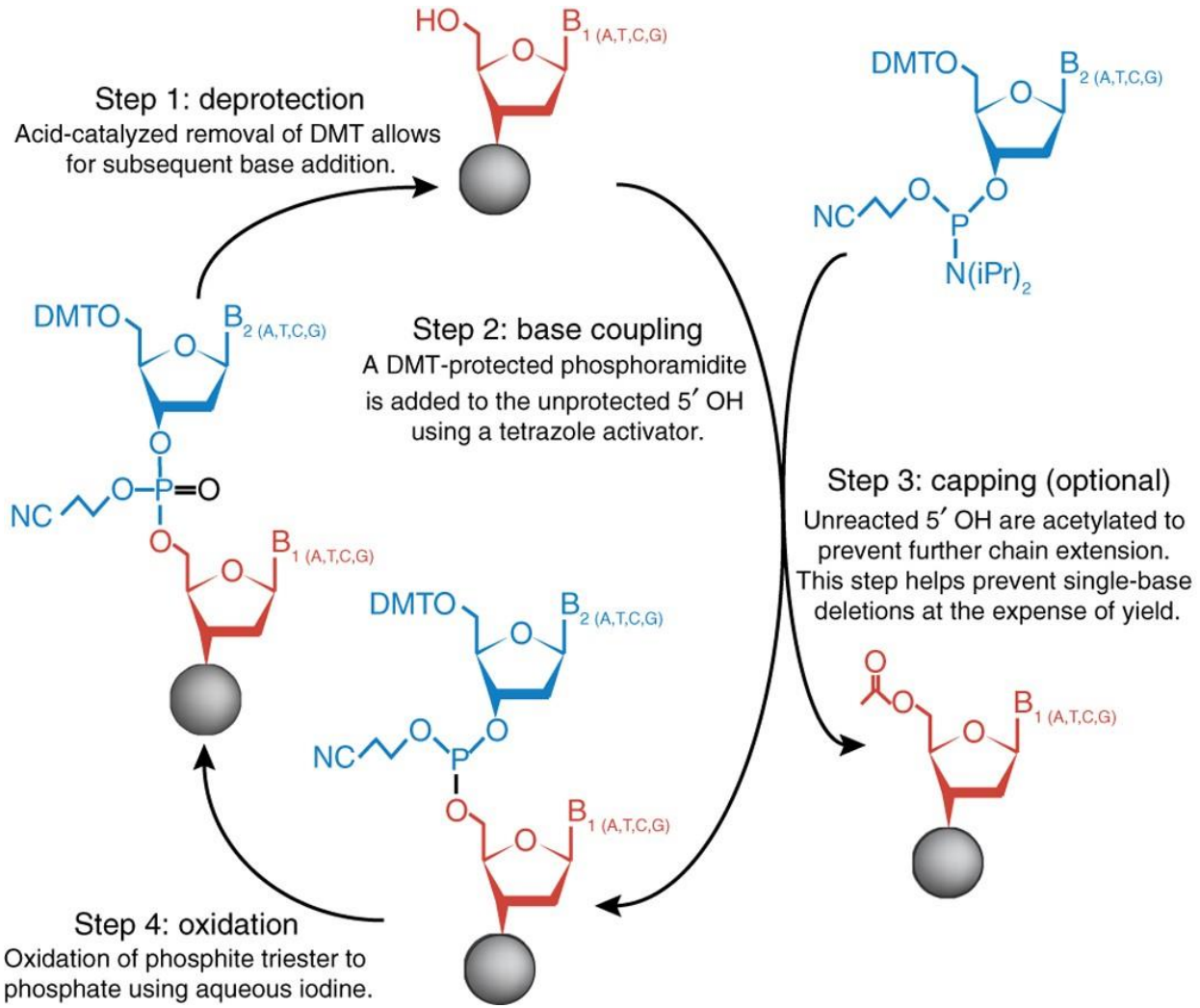
- Unlike *in vivo* DNA synthesis, artificial (chemical) synthesis is done in the **3′ to 5′ direction**.



dA-CE-phosphoramidite

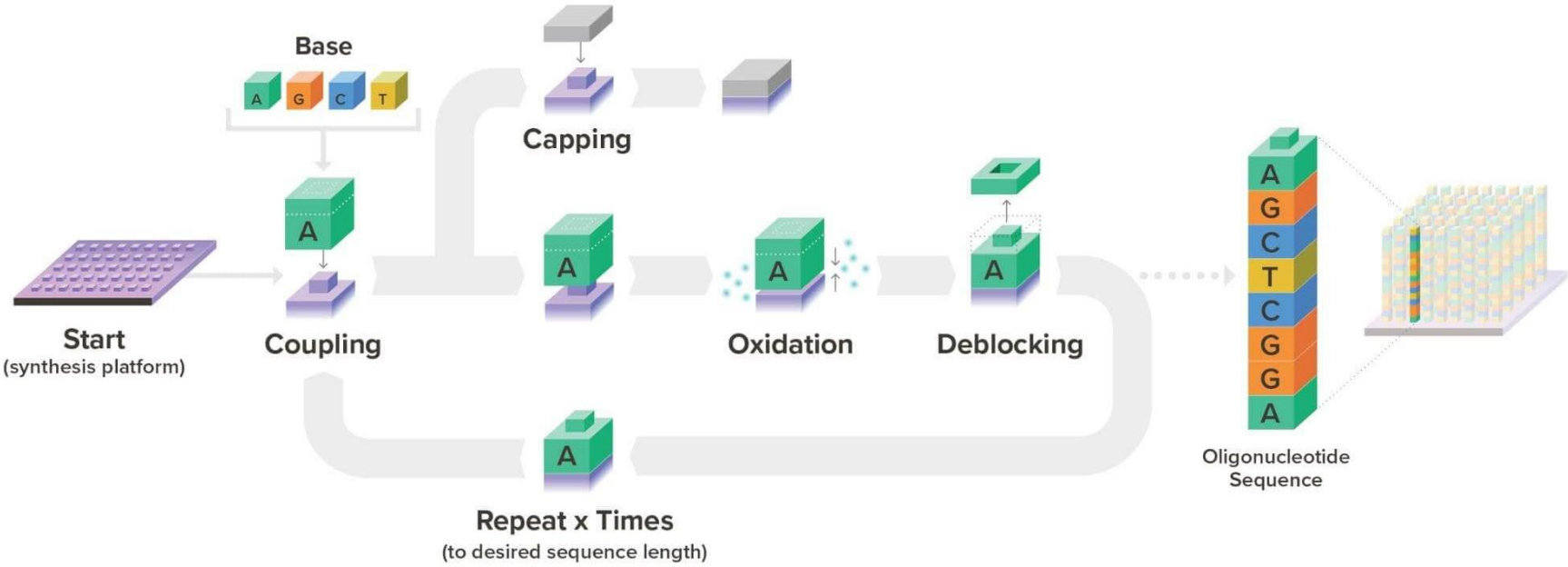
5'-dimethoxytrityl-N-benzoyl-2'-deoxyadenosine-3'-[(2-cyanoethyl)-(N,N-diisopropy)]-phosphoramidite

Overview of the phosphoramidite approach



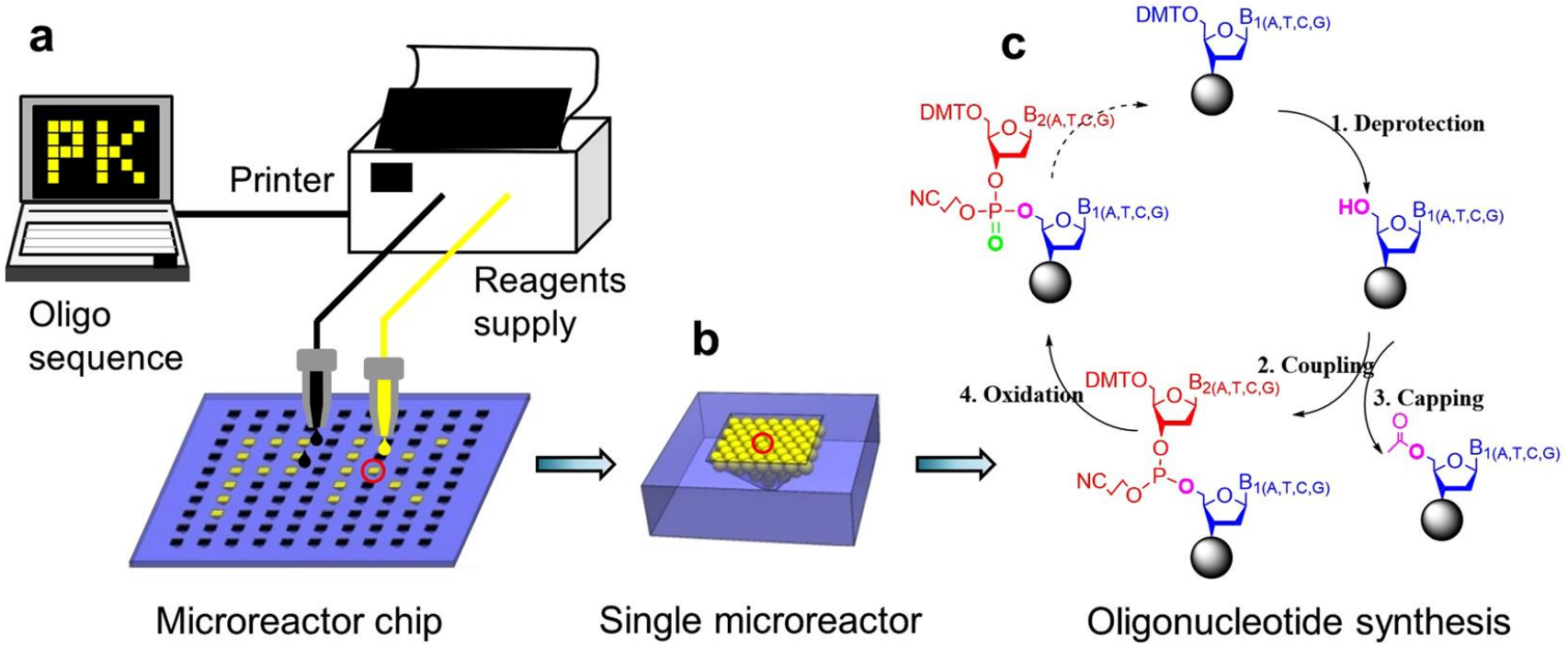
This 4-step cycle repeats until the oligo receives its final nucleotide.

Automated and miniaturized oligonucleotide synthesis



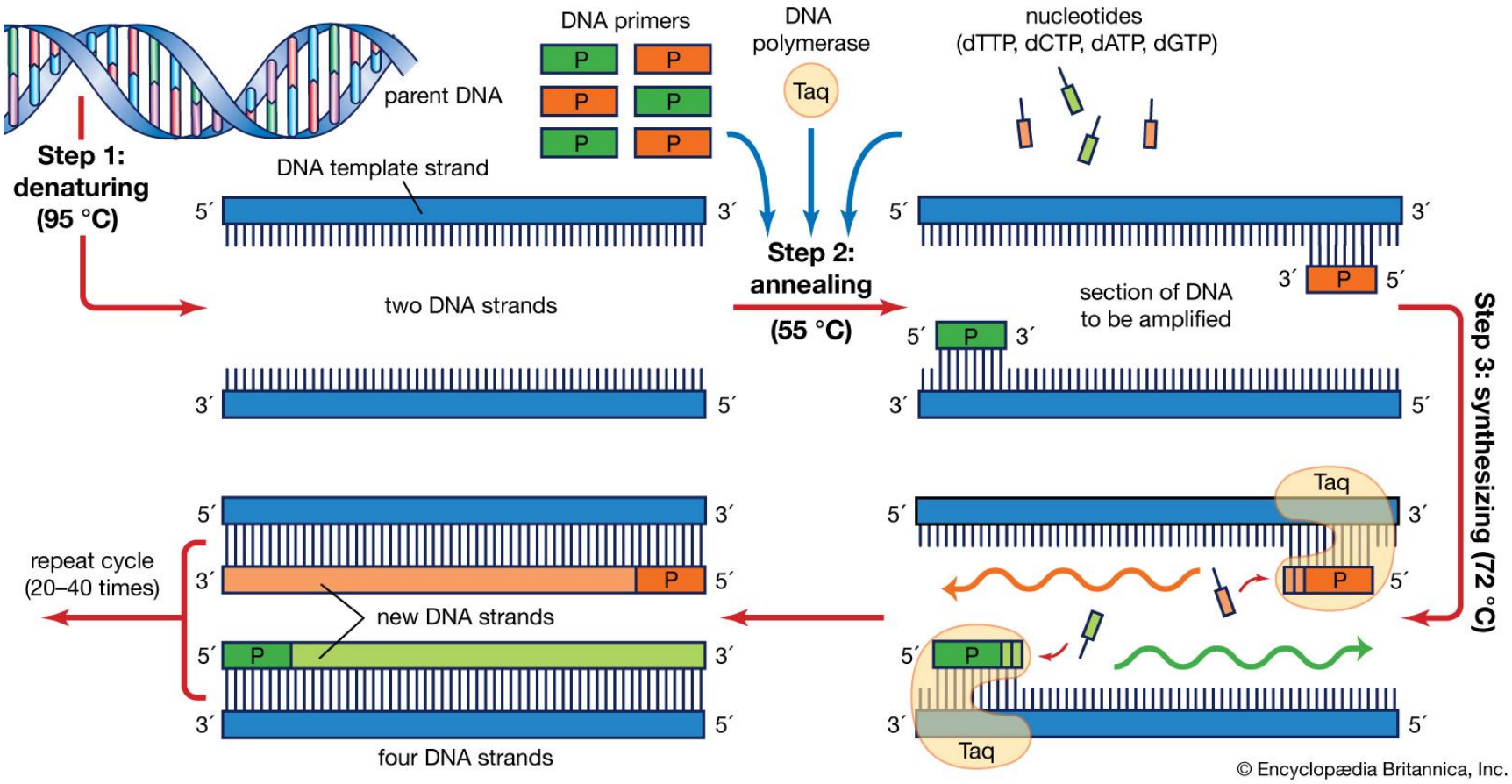
Start-Coupling: the first phosphoramidite in the chain is attached to the solid surface with a catalyzed condensation reaction (think of this as linking pinkies when trying to hold hands). Oxidation: the phosphite triester is unstable, so it is converted to a phosphate to improve sequence integrity (think of this as grabbing hold of the hand tightly). Deblocking: the 5' protecting group is removed in acidic conditions. Coupling: the next phosphoramidite in the chain is coupled to the available -OH on the previous deblocked molecule in a catalyzed reaction. Capping: as the coupling is not 100% efficient, sometimes the coupling fails. Therefore uncoupled sequences could create errors in the synthesized molecule. To stop this, an unreactive group is added blocking further extension. Repetition: the oxidation -> coupling cycle can be repeated to extend the oligonucleotide molecule in a desired sequence.

The schematic view of the oligonucleotide synthesis



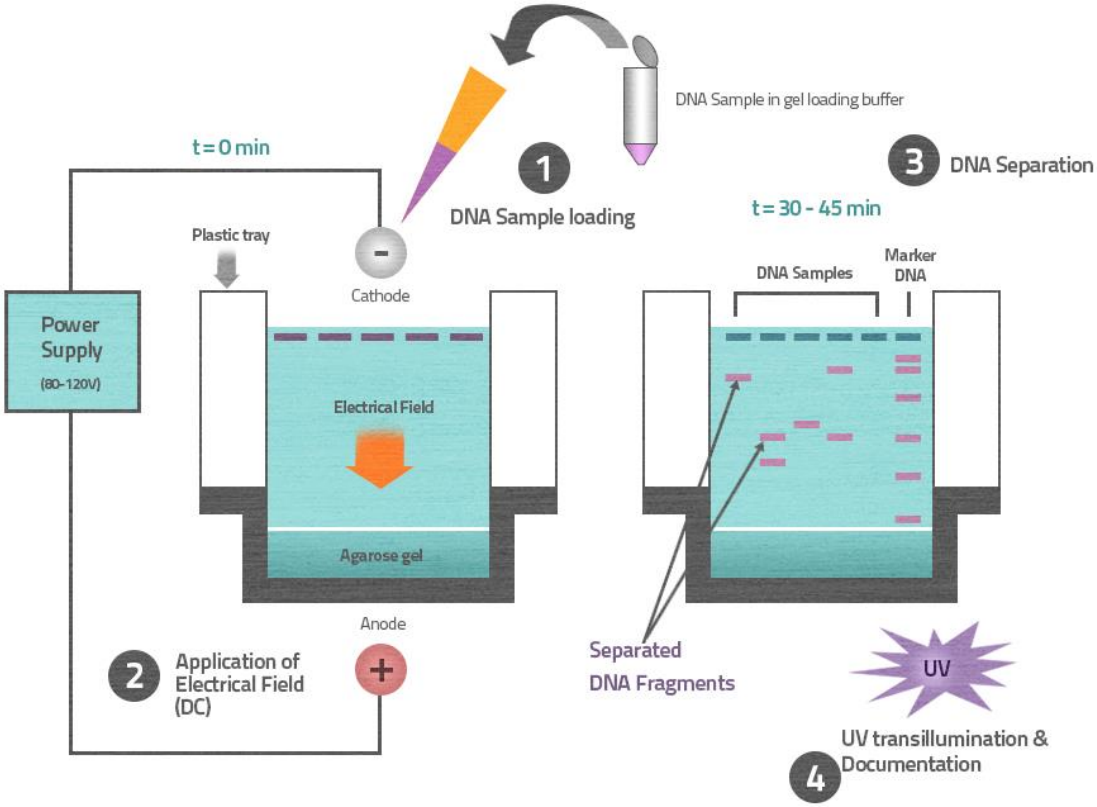
(**a**) The synthesis processes involve designing target sequence, delivering chemical reagents through the inkjet printer, and oligonucleotide synthesis in the microreactor chip. (**b**) The single microreactor is filled with silica beads which enhance the surface area for the following synthesis. The beads are inherently fixed in the microreactor using sintering process. (**c**) The oligonucleotide synthesis on the silica beads follows the four-step with phosphoramidite strategy: deprotection, coupling, capping and oxidation.

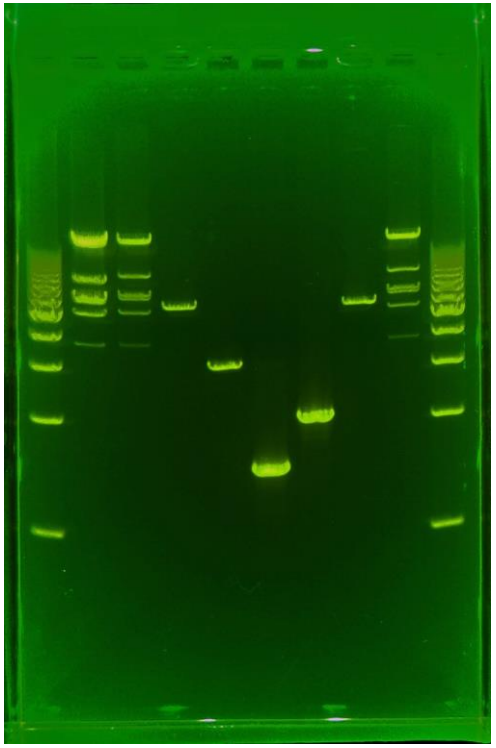
Enzymatic DNA synthesis: polymerase chain reaction (PCR)



Amplification of up to 20 kbp DNA fragment from pre-existing template (genomic loci, cDNA library, cloned fragment etc.)

Gel electrophoresis



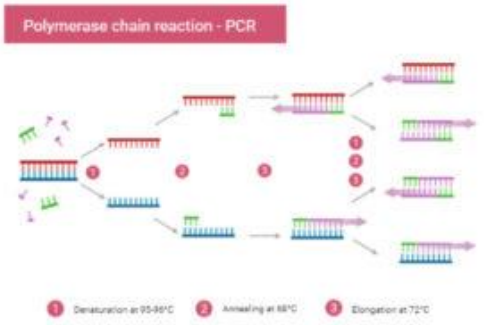


- Gel electrophoresis is a method for separation and analysis of macromolecules (DNA) and their fragments, based on their size and charge.
- It is used in molecular biology to separate a mixed population of DNA fragments by length, to estimate the size of DNA fragments.

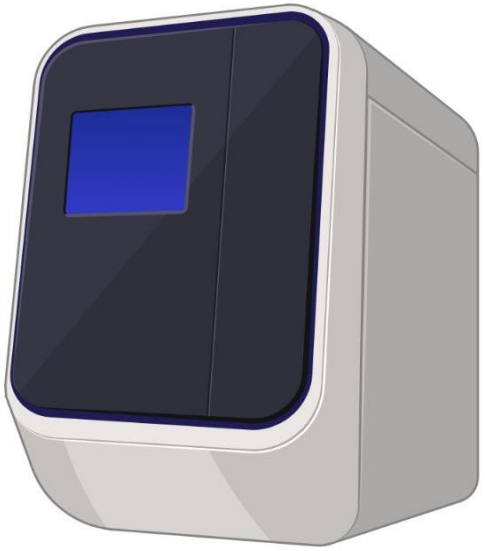
Different PCR protocols

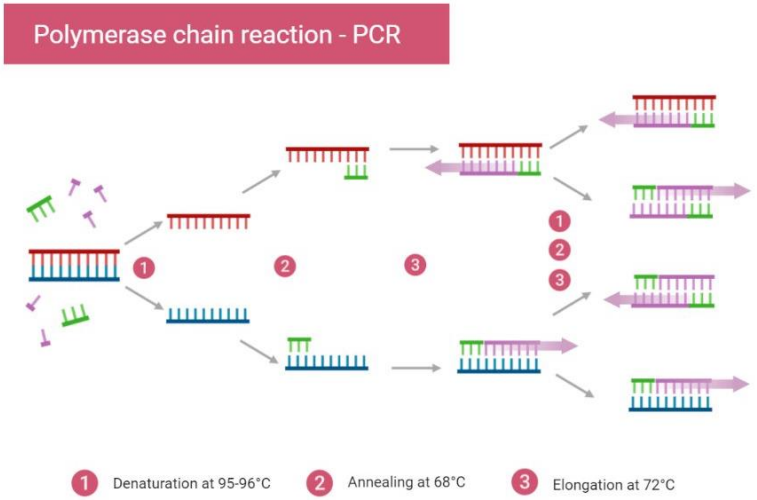
Types of PCR with definition and uses

1. AFLP PCR
2. Allele-specific PCR
3. Alu PCR
4. Assembly PCR
5. Asymmetric PCR
6. COLD PCR
7. Colony PCR
8. Conventional PCR
9. Digital PCR (dPCR)
10. Fast-cycling PCR
11. High-fidelity PCR
12. Hot-start PCR
13. In situ PCR
14. Intersequence-specific (ISSR) PCR
15. Inverse PCR
16. LATE (linear after the exponential) PCR
17. Ligation-mediated PCR
18. Long-range PCR
19. Methylation-specific PCR (MSP)
20. Miniprimer PCR
21. Multiplex-PCR
22. Nanoparticle-Assisted PCR (nanoPCR)
23. Nested PCR
24. Overlap extension PCR
25. Real-Time PCR (quantitative PCR or qPCR)
26. Repetitive sequence-based PCR
27. Reverse-Transcriptase (RT-PCR)
28. Reverse-Transcriptase Real-Time PCR (RT-qPCR)
29. RNase H-dependent PCR (rhPCR)
30. Single cell PCR
31. Single Specific Primer-PCR (SSP-PCR)
32. Solid phase PCR
33. Suicide PCR
34. Thermal asymmetric interlaced PCR (TAIL-PCR)
35. Touch down (TD) PCR
36. Variable Number of Tandem Repeats (VNTR) PCR



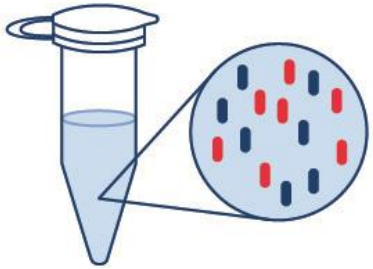
Real-time PCR / quantitative PCR (qPCR)





- It is a technique used to monitor the progress of a PCR reaction in real-time.
- At the same time, a relatively small amount of PCR product (DNA, cDNA or RNA) can be quantified.
- The process is monitored in “real-time”. The reaction is placed into a real-time PCR machine that watches the reaction occur with a camera or detector.
- To link the amplification of DNA to the generation of fluorescence which can simply be detected with a camera during each PCR cycle.
- Hence, as the number of gene copies increases during the reaction, so does the fluorescence, indicating the progress of the reaction.

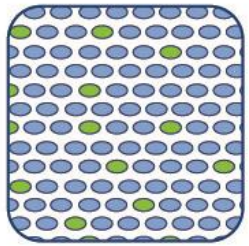
Digital PCR (dPCR)



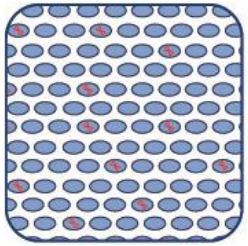
Sample dilution and PCR reaction mix setup

Blue – Target

Red - Background (gDNA, cDNA; primers/probes; master mix)



PCR reaction partitioning into thousands of individual reactions



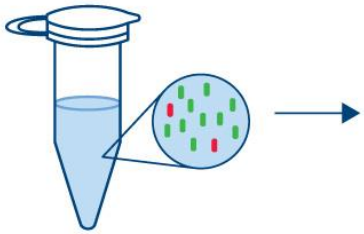
End-point PCR amplification of partitions

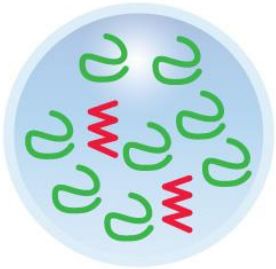
Readout and absolute quantification

- Digital PCR is a highly precise approach to sensitive and reproducible nucleic acid detection and quantification.
- Measurements are performed by dividing the sample into partitions, such that there are either zero or one or more target molecules present in any individual reaction.
- Each partition is analyzed after end-point PCR cycling for the presence (positive reaction) or absence (negative reaction) of a fluorescent signal, and the absolute number of molecules present in the sample is calculated. It does not rely on a standard curve for sample target quantification.
- Eliminating the reliance on standard curves reduces error and improves precision.

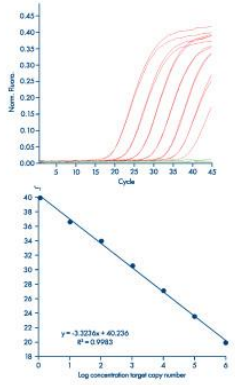
Comparison between real-time qPCR and digital PCR

Real-time qPCR



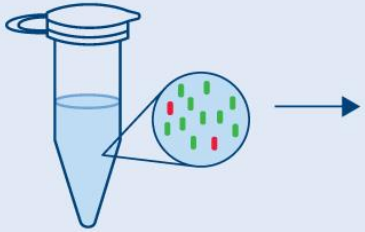


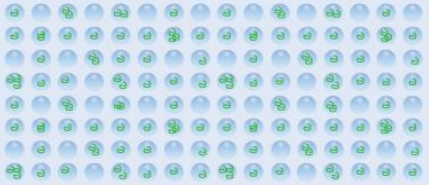
Bulk reaction analysis



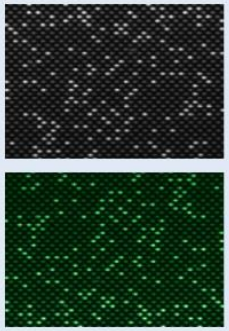
Relative quantification: C_q

Digital PCR



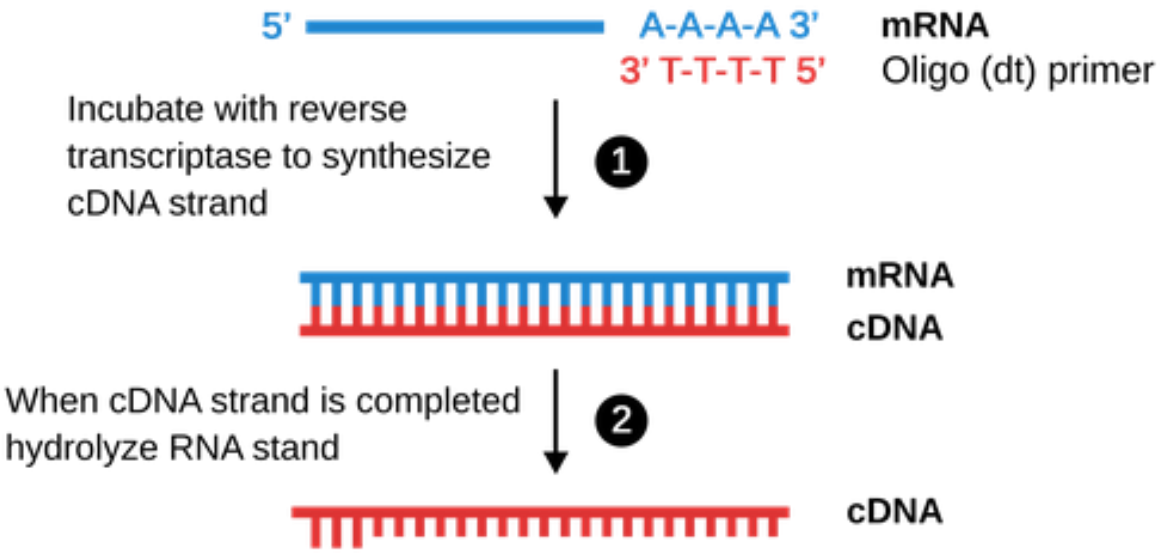


Random distribution
of molecules into partitions



Absolute quantification: Copies/µl

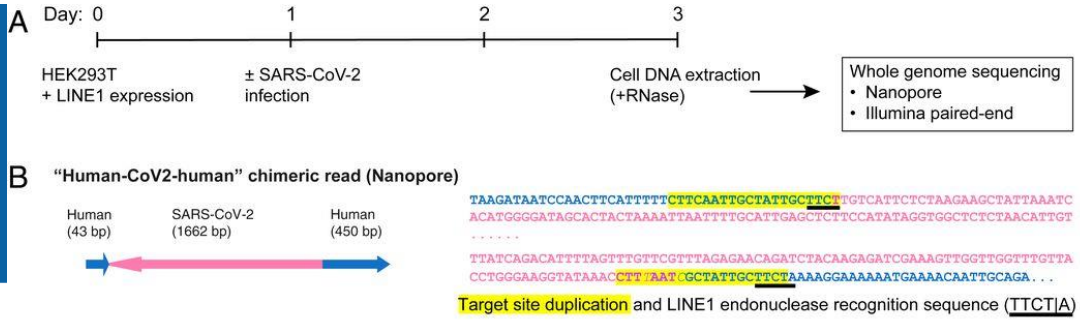
Reverse transcription polymerase chain reaction (RT-PCR)



- Conversion of RNA into cDNA using reverse transcriptase
- Amplification of cDNA using PCR
- cDNA is DNA that is synthesized from messenger RNA molecules. cDNA synthesis is catalyzed by an enzyme called **reverse transcriptase**, which uses RNA as a template for DNA synthesis. Reverse transcriptase was initially discovered and isolated from a retrovirus. These viruses contain an RNA genome; therefore the viruses need to produce a cDNA copy of their genome to be compatible with the host cell's molecular machinery.

PNAS
Proceedings of the National Academy of Sciences of the United States of America

Home | Articles | Front Matter | News | Podcasts

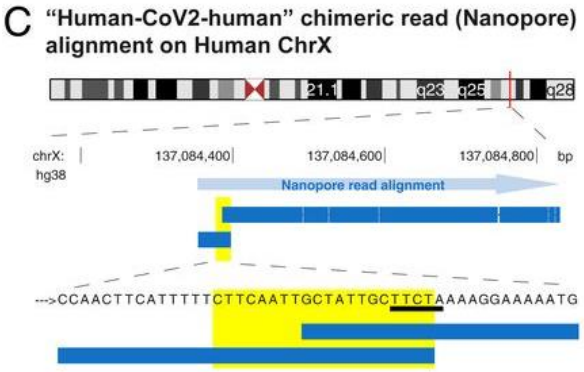


RESEARCH ARTICLE

Reverse-transcribed SARS-CoV-2 RNA can integrate into the genome of cultured human cells and can be expressed in patient-derived tissues

Liguo Zhang, Alexsia Richards, M. Inmaculada Barrasa, Stephen H. Hughes, Richard A. ...
+ See all authors and affiliations

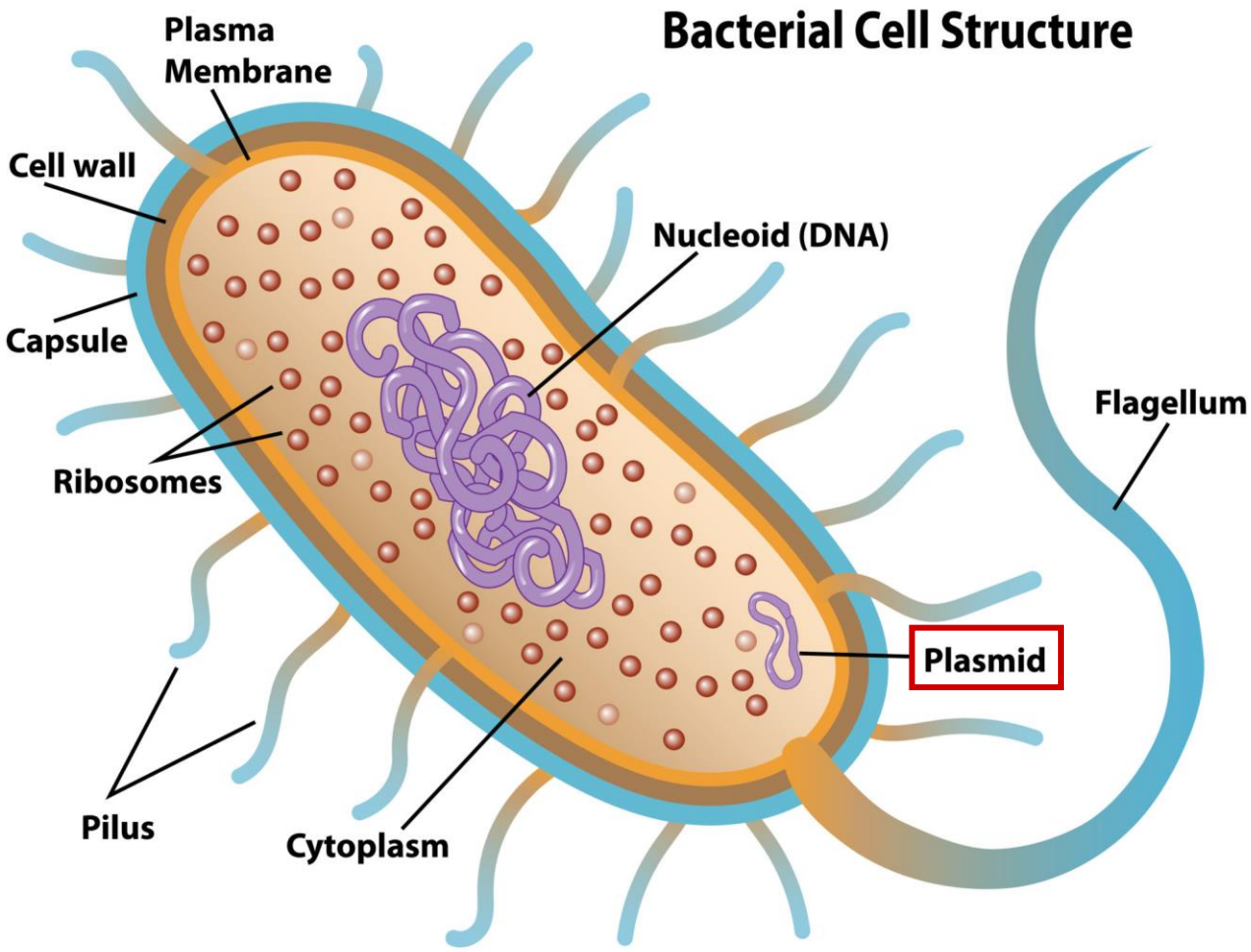
PNAS May 25, 2021 118 (21) e2105968118; https://doi.org/10.1073/pnas.2105968118



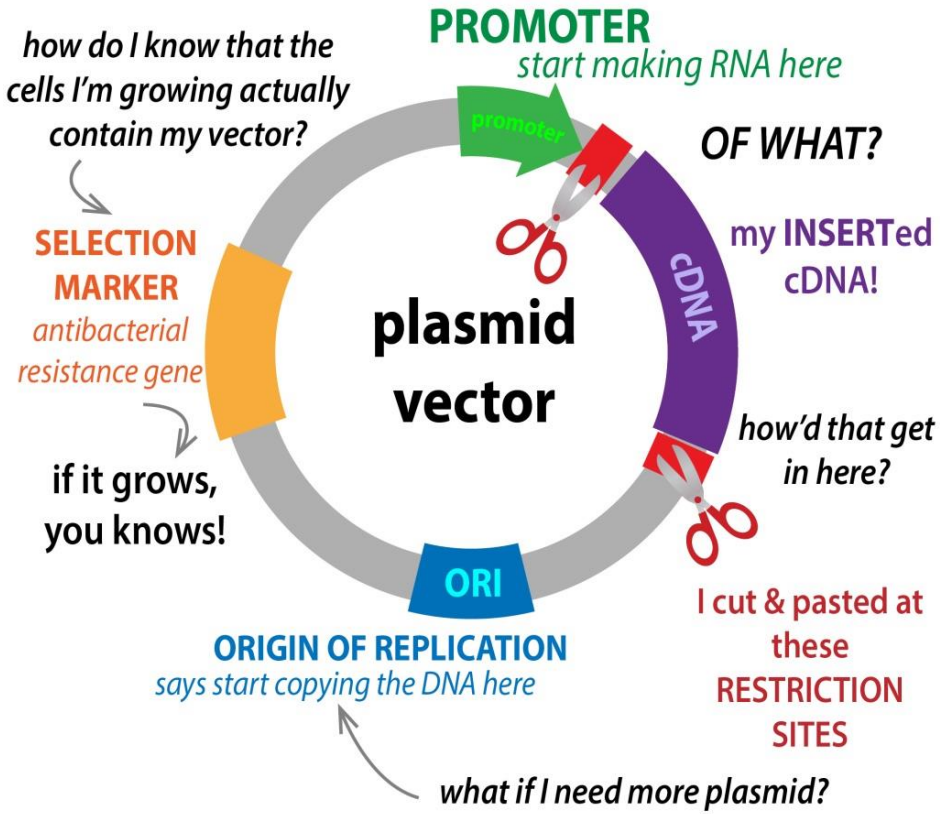
Liguo Zhang et al. PNAS 2021;118:21:e2105968118

An unresolved issue of SARS-CoV-2 disease is that patients often remain positive for viral RNA as detected by PCR many weeks after the initial infection in the absence of evidence for viral replication. We show here that SARS-CoV-2 RNA can be reverse-transcribed and integrated into the genome of the infected cell and be expressed as chimeric transcripts fusing viral with cellular sequences. Importantly, such chimeric transcripts are detected in patient-derived tissues. Our data suggest that, in some patient tissues, the majority of all viral transcripts are derived from integrated sequences. Our data provide an insight into the consequence of SARS-CoV-2 infections that may help to explain why patients can continue to produce viral RNA after recovery.

Plasmids: essential tools for genetic engineering



What is a plasmid?



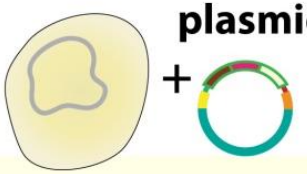
- PLASMIDS are "extrachromosomal" (not part of the chromosomes), circular pieces of DNA.
- Similarly to chromosomes, they are double-stranded, which means they can easily be "unzipped" and copied (replicated).
- Plasmids use the host's machinery (DNA polymerase), but they don't have to wait for the host to divide to copy themselves → lots of copies of themselves.
- When the cell does divide, these copies will get split between the daughter cells, so they'll inherit the plasmid as well.
- The plasmid can act as a VECTOR – a vehicle for taking genes we want to deliver into cells.

Always sequence your plasmid to double-check that the gene is correctly inserted!

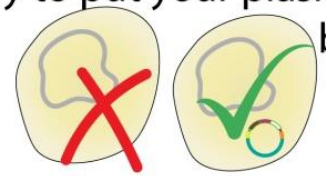
Antibiotics as selectable markers

host cell DNA doesn't have the resistance gene

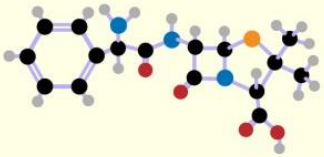
plasmids do



try to put your plasmid into bacteria



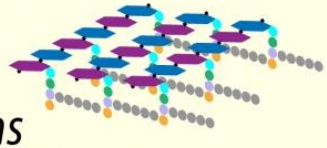
use antibiotics to select for those that actually took it



AMPICILLIN

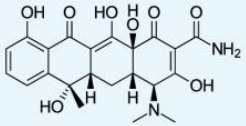
and other Beta-lactams

target cell wall synthesis



TETRACYCLINE

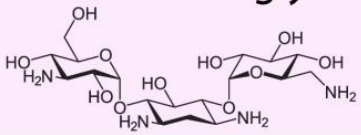
and other tetracyclines



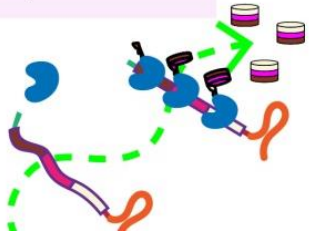
&

KANAMYCIN

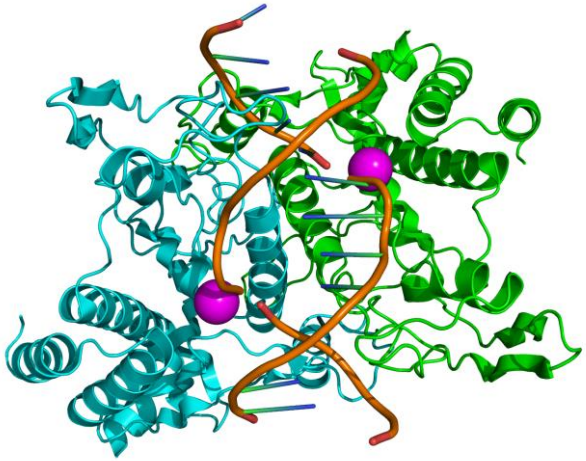
and other aminoglycosides

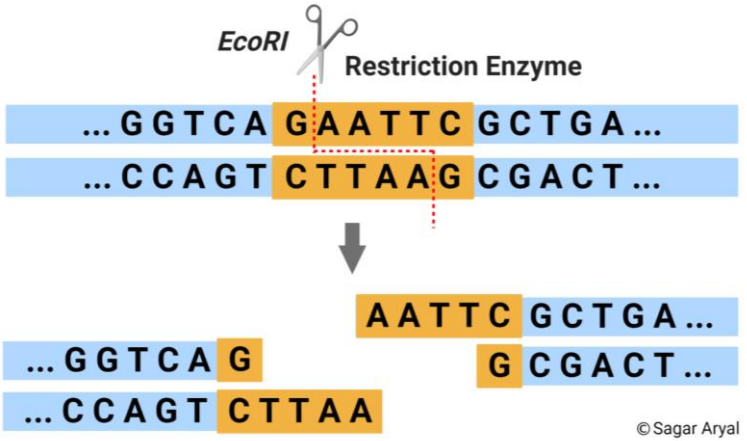


both target protein synthesis (translation) *but in different ways*



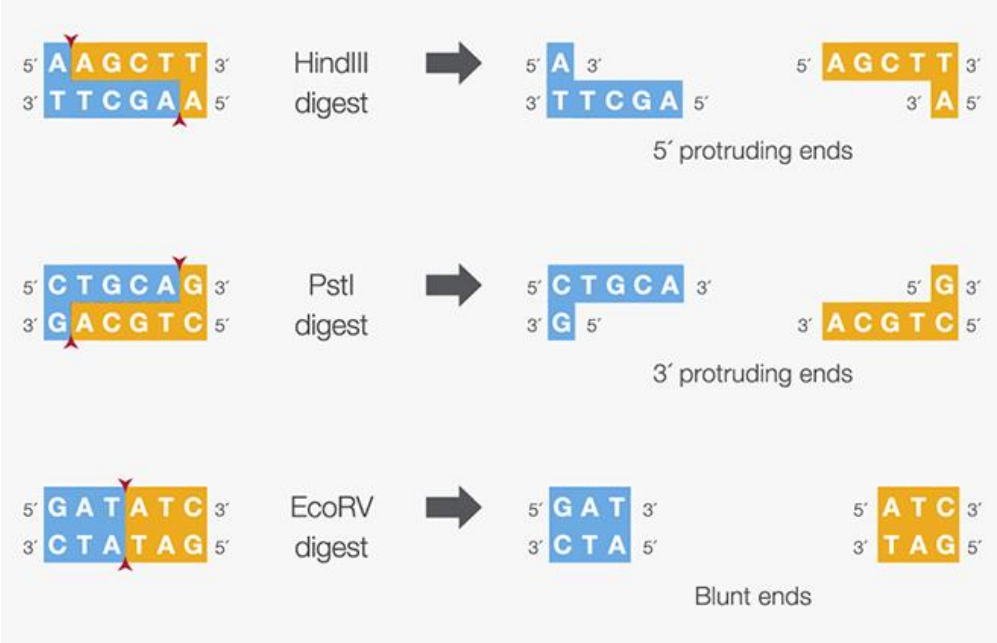
Restriction endonucleases





- Restriction enzyme, also called restriction endonuclease, is a protein produced by bacteria that cleaves DNA at specific sites along the molecule.
- Restriction endonucleases cut the DNA double helix in very precise ways. It cleaves DNA into fragments at or near specific recognition sites within the molecule known as restriction sites.
- They have the capacity to recognize specific base sequences on DNA and then to cut each strand at a given place. Hence, they are also called as ‘molecular scissors’.

Restriction endonucleases (restriction enzymes)



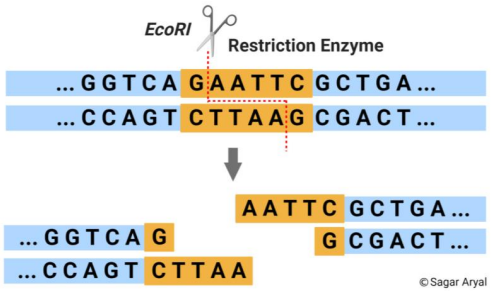
Type I enzymes cleave at sites remote from a recognition site; require both ATP and S-adenosyl-L-methionine to function; multifunctional protein with both restriction and methylase activities.

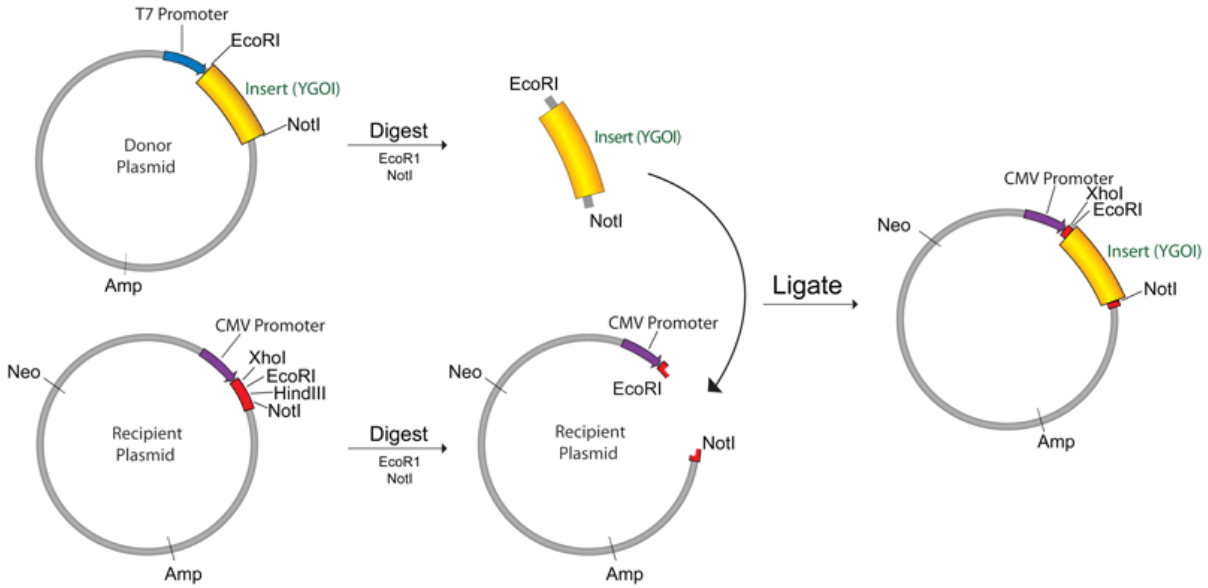
Type II enzymes cleave within or at short specific distances from a recognition site; most require magnesium; single function (restriction) enzymes independent of methylase.

Type III enzymes cleave at sites a short distance from a recognition site; require ATP (but do not hydrolyze it); S-adenosyl-L-methionine stimulates the reaction but is not required; it exists as part of a complex with a modification methylase.

Type IV enzymes target modified DNA, e.g. methylated, hydroxymethylated and glucosyl-hydroxymethylated DNA.

Restriction cloning

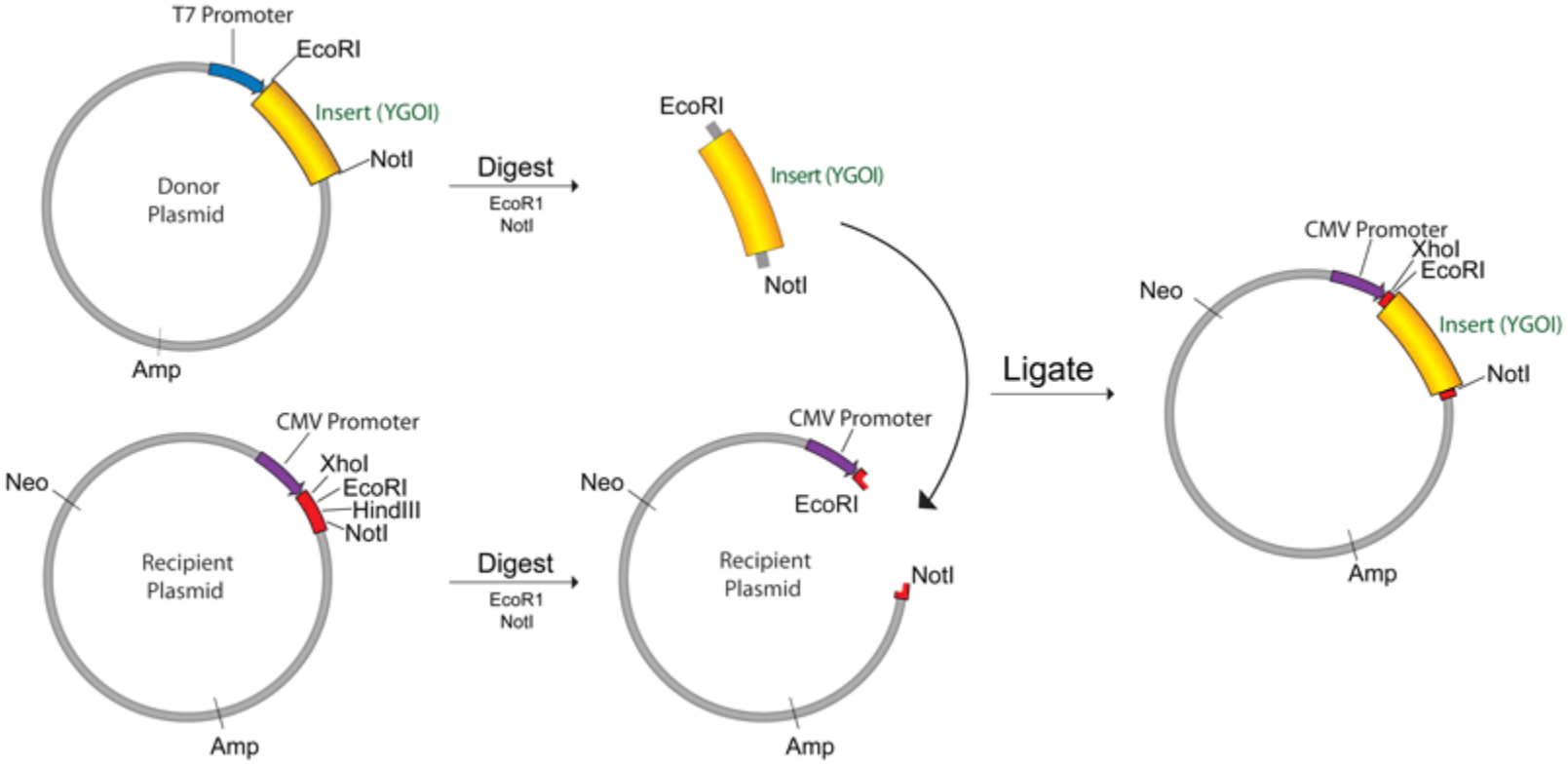




In vitro

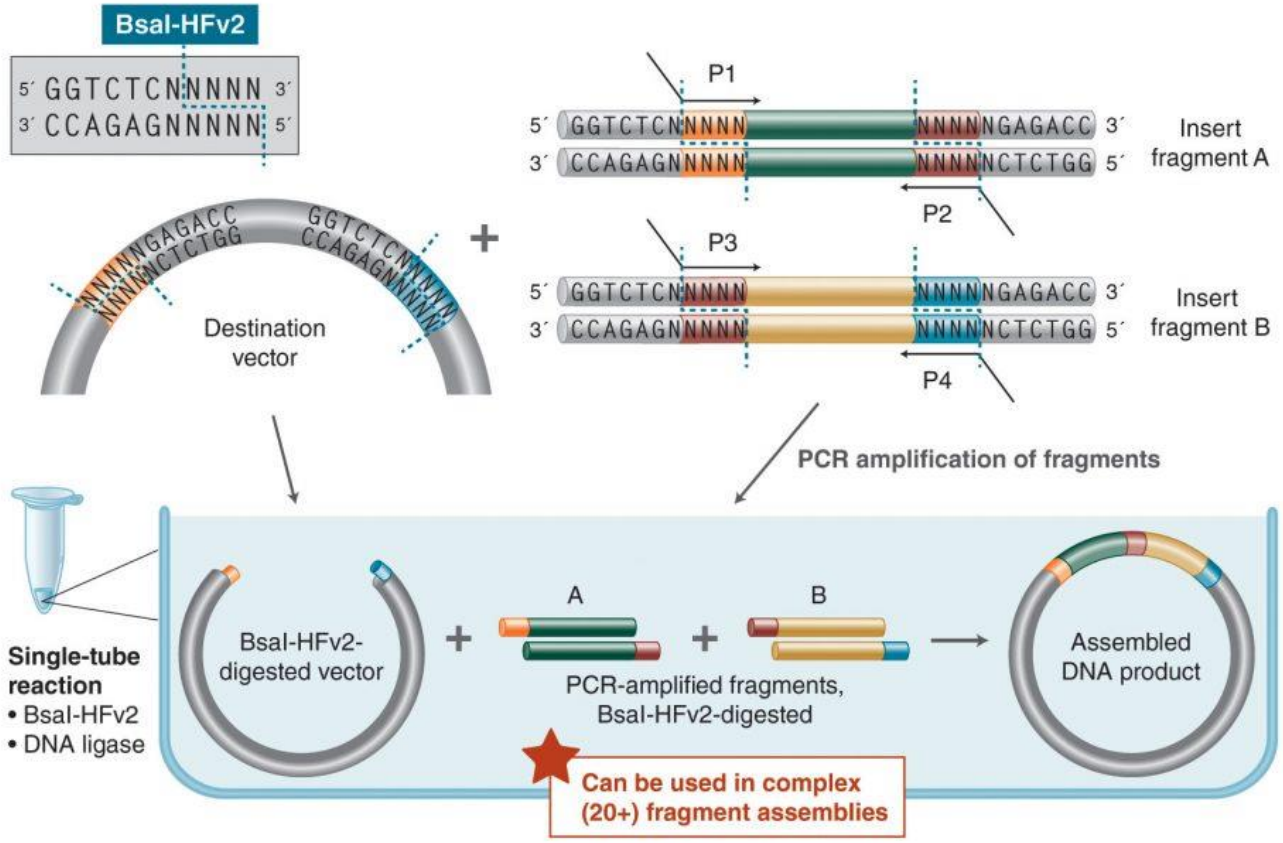
The production of exact copies of a particular gene or DNA sequence using genetic engineering techniques is called gene cloning.

PCR-based cloning



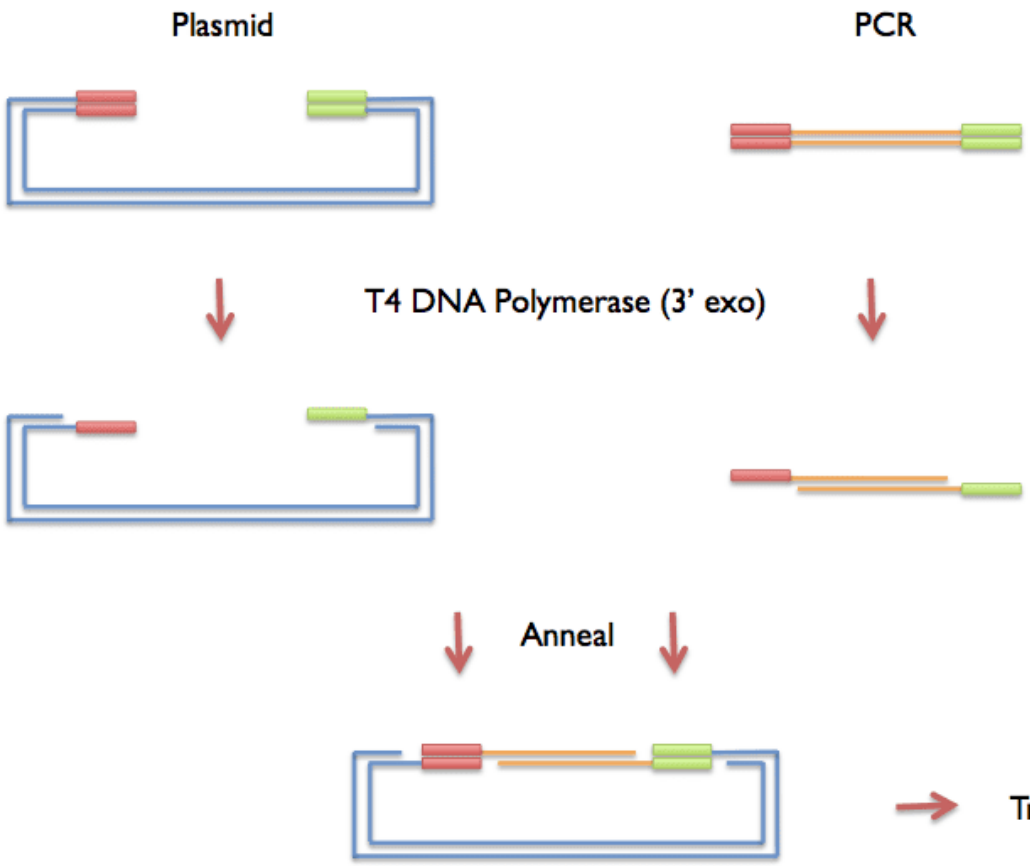
- DNA fragment (gene) of interest can be flanked by any restriction site

Golden gate DNA assembly



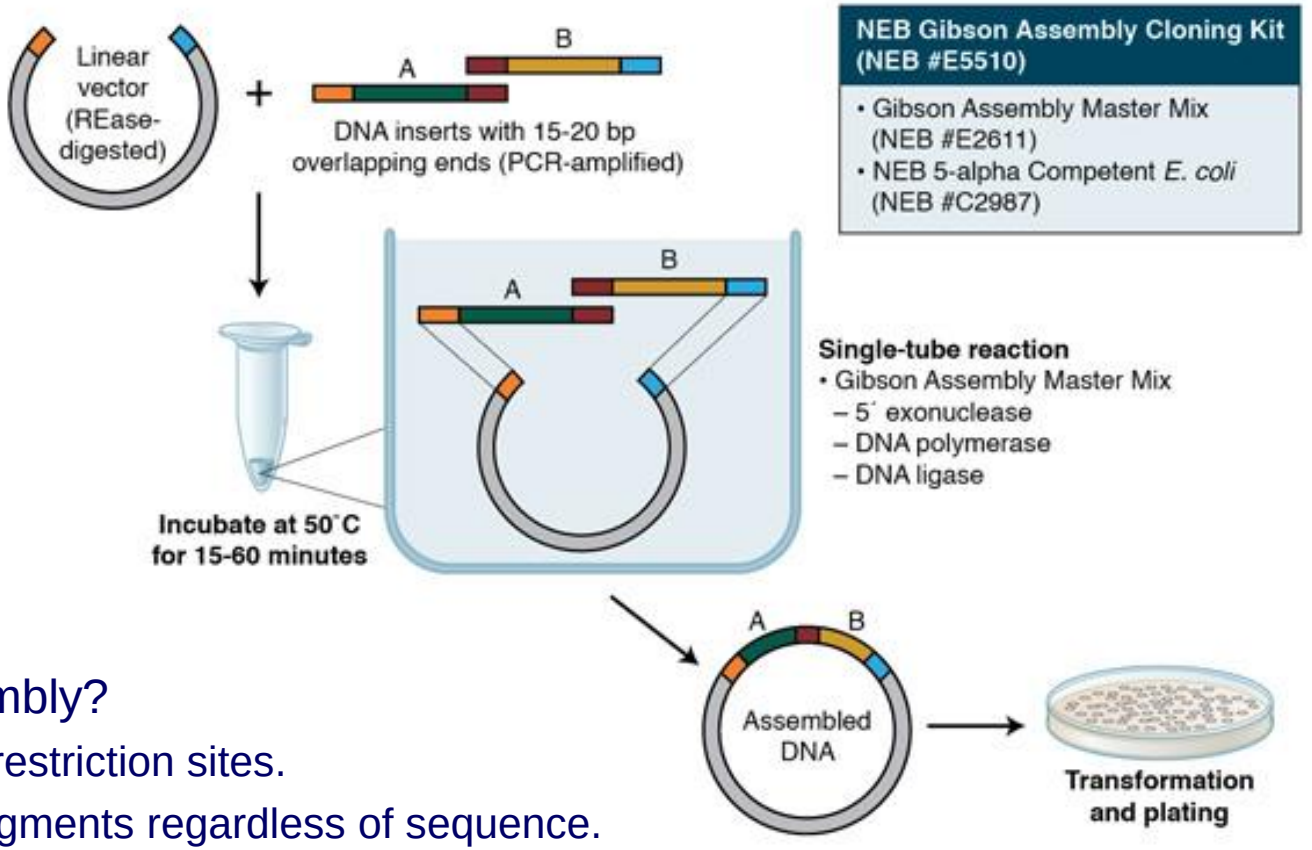
Simultaneous and directional assembly of multiple DNA fragments into a single piece using Type IIs restriction enzymes and T4 DNA ligase.

Sequence and ligation independent cloning (SLIC)



- SLIC, or sequence and ligase independent cloning, does not utilize restriction enzymes or ligase.
- A DNA sequence fragment to be cloned into a destination vector is PCR amplified with oligos whose 5' termini contain about 25 bp of sequence homology to the ends of the destination vector, linearized either by restriction digest or PCR amplification.

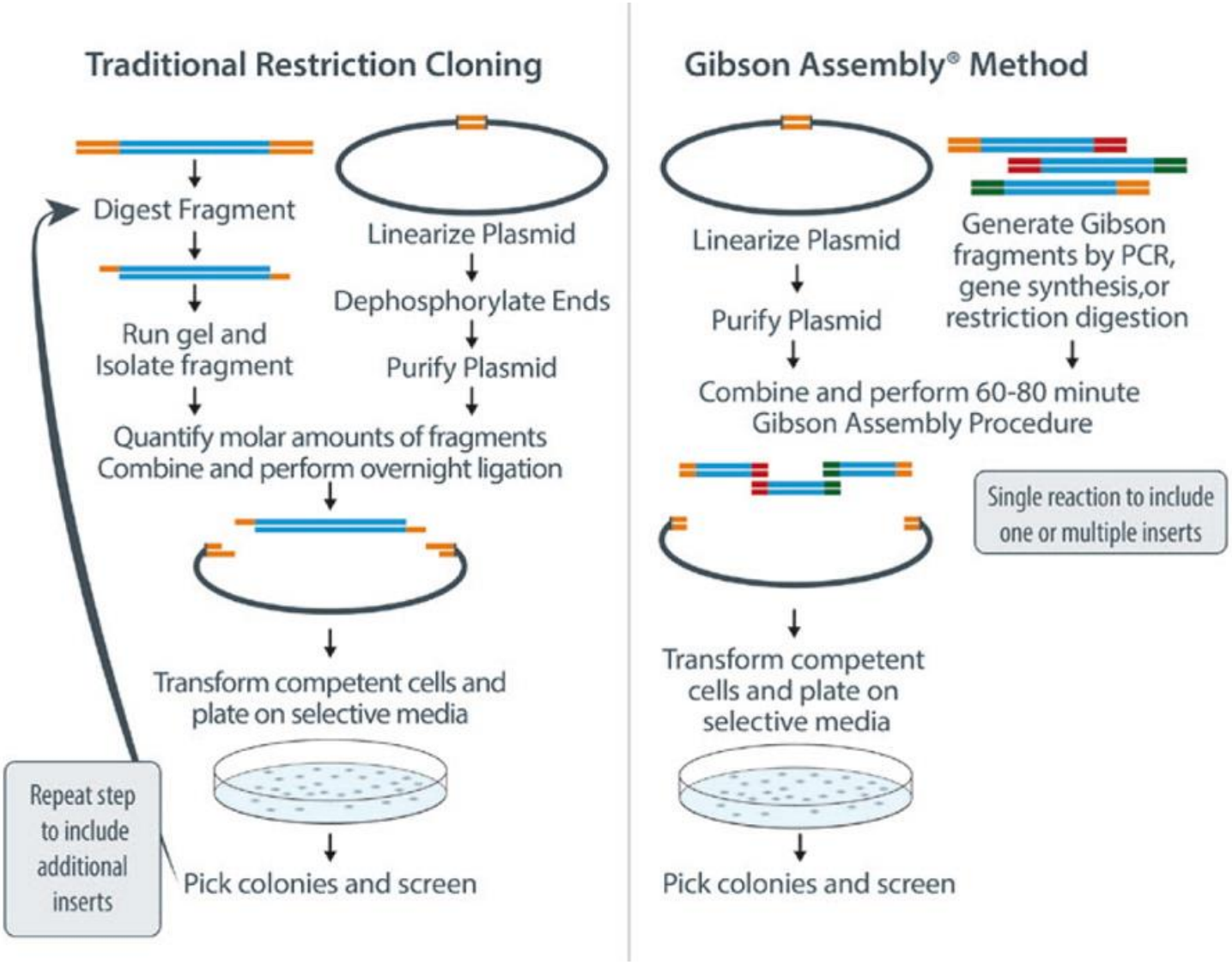
Gibson DNA assembly



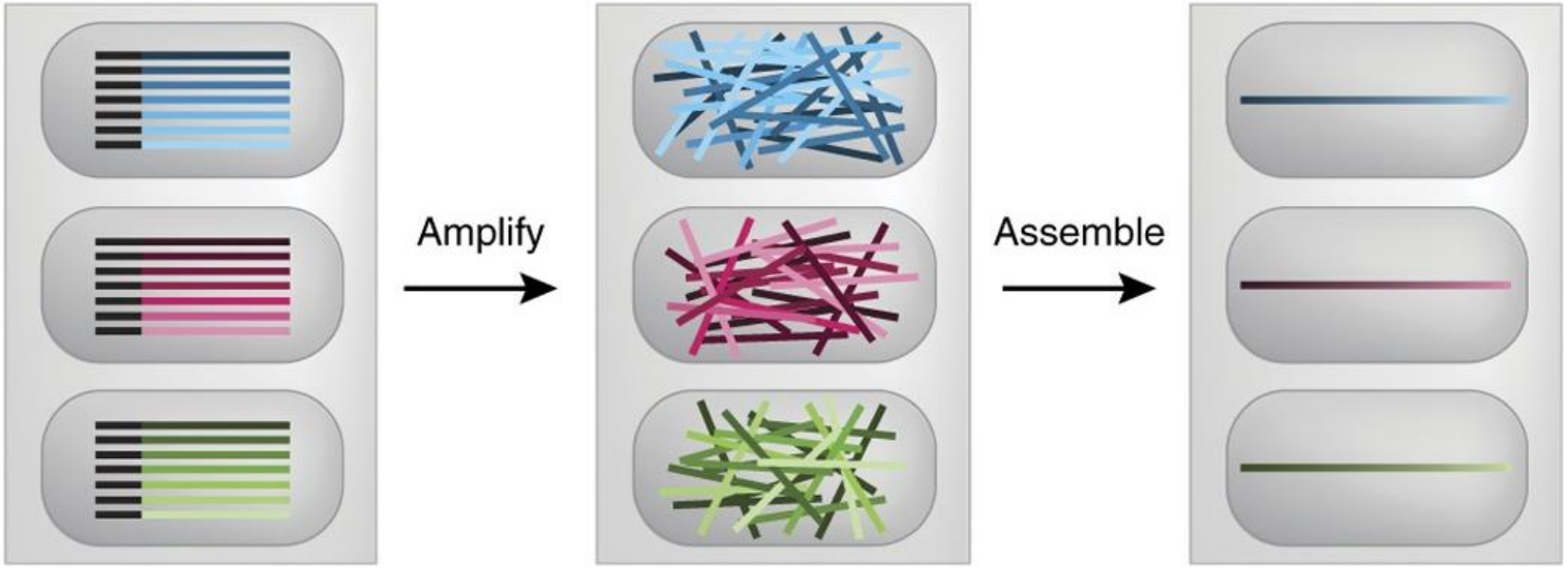
Why Gibson Assembly?

- No need for specific restriction sites.
- Join almost any 2 fragments regardless of sequence.
- No scar between joined fragments.
- Fewer steps. One tube reaction.
- Can combine many DNA fragments at once.

Traditional restriction cloning versus Gibson assembly

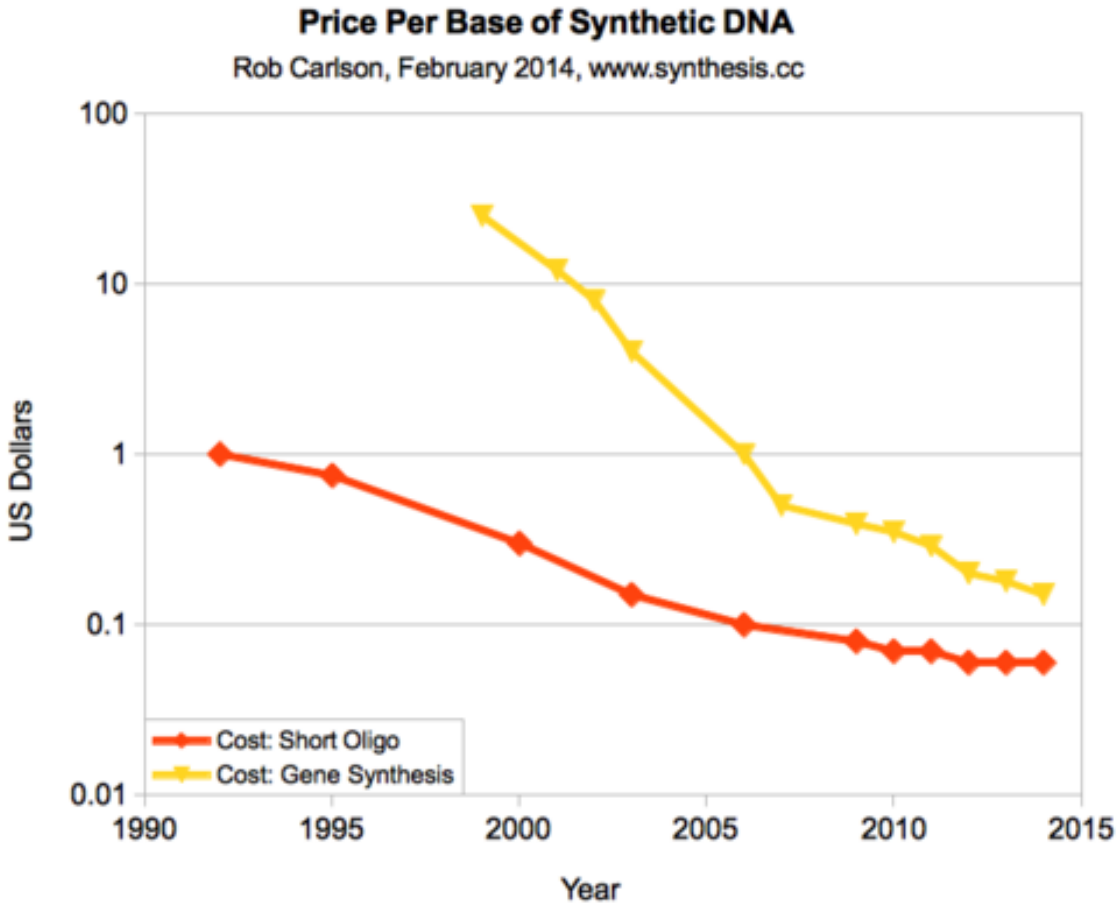


Large-scale *de novo* DNA synthesis



- Protein engineering
- Engineered metabolic pathways
- Synthetic biology
- Whole-genome syntheses
- DNA nanotechnology (DNA computers)

DNA synthesis prices

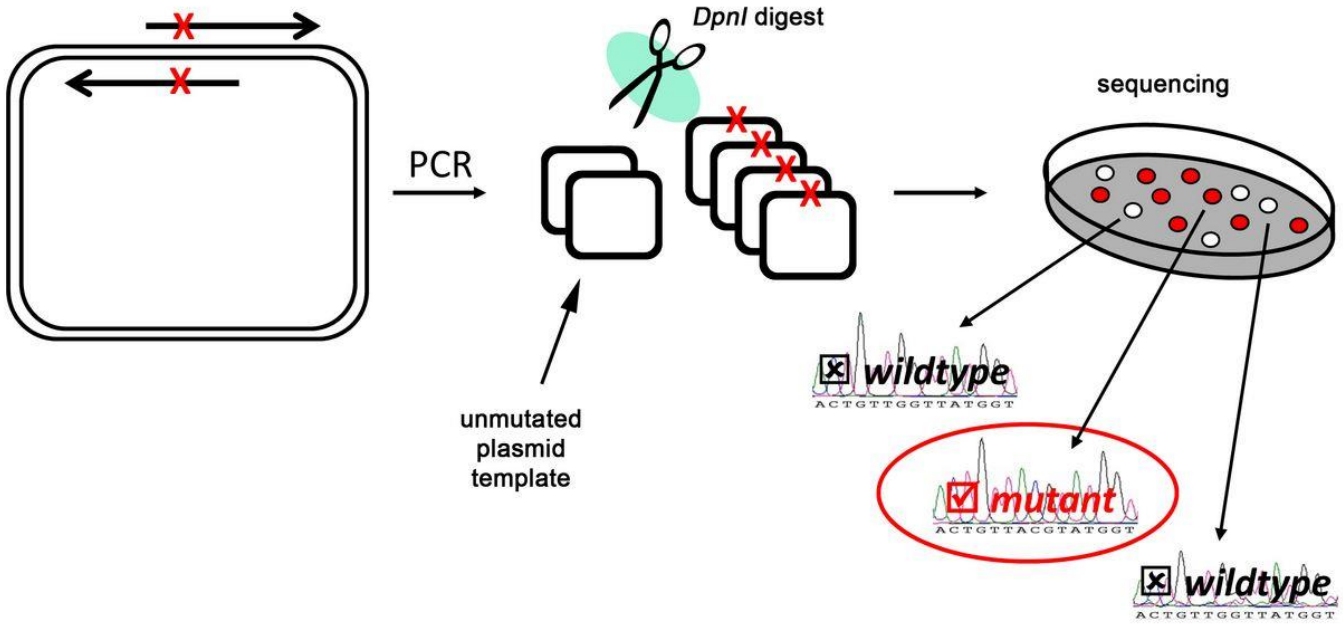


DNA mutagenesis

Concepts
Methods
Applications

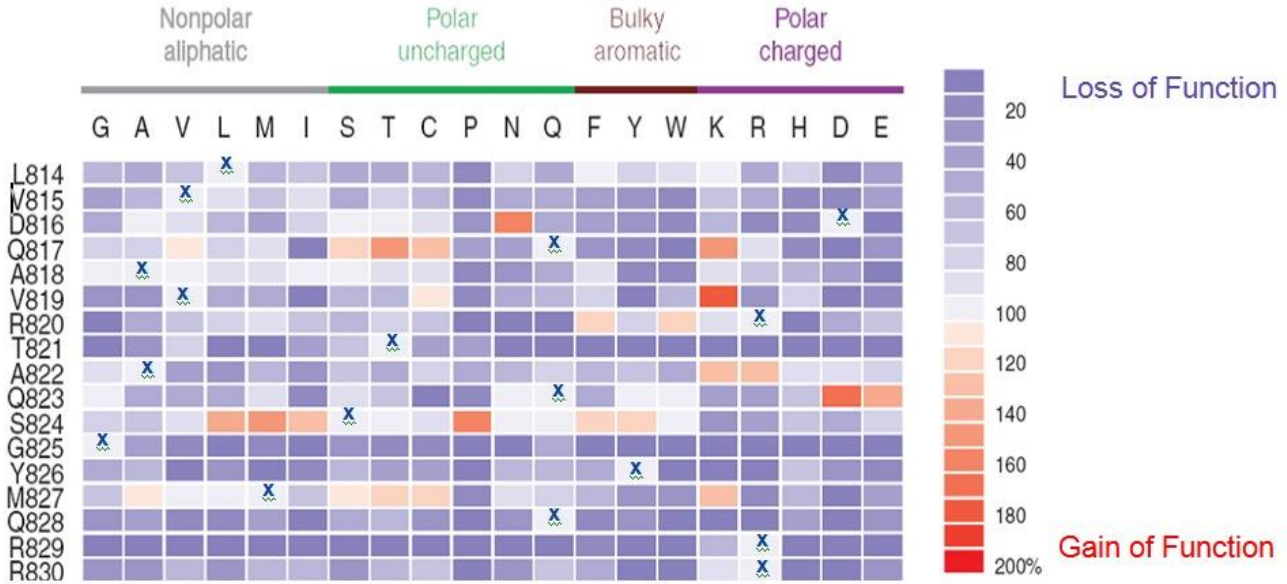
Site-directed mutagenesis

- Site-directed mutagenesis is used to generate mutations that may produce a rationally designed protein that has improved or special properties (i.e. protein engineering).
- The basic procedure requires the synthesis of a short DNA primer. This synthetic primer contains the desired mutation and is complementary to the template DNA around the mutation site so it can hybridize with the DNA in the gene of interest. The mutation may be a single base change (a point mutation), multiple base changes, deletion, or insertion. The single-strand primer is then extended using a DNA polymerase, which copies the rest of the gene. The copied gene thus contains the mutated site, and is then introduced into a host cell in a vector and cloned. Finally, mutants are selected by DNA sequencing to check that they contain the desired mutation.

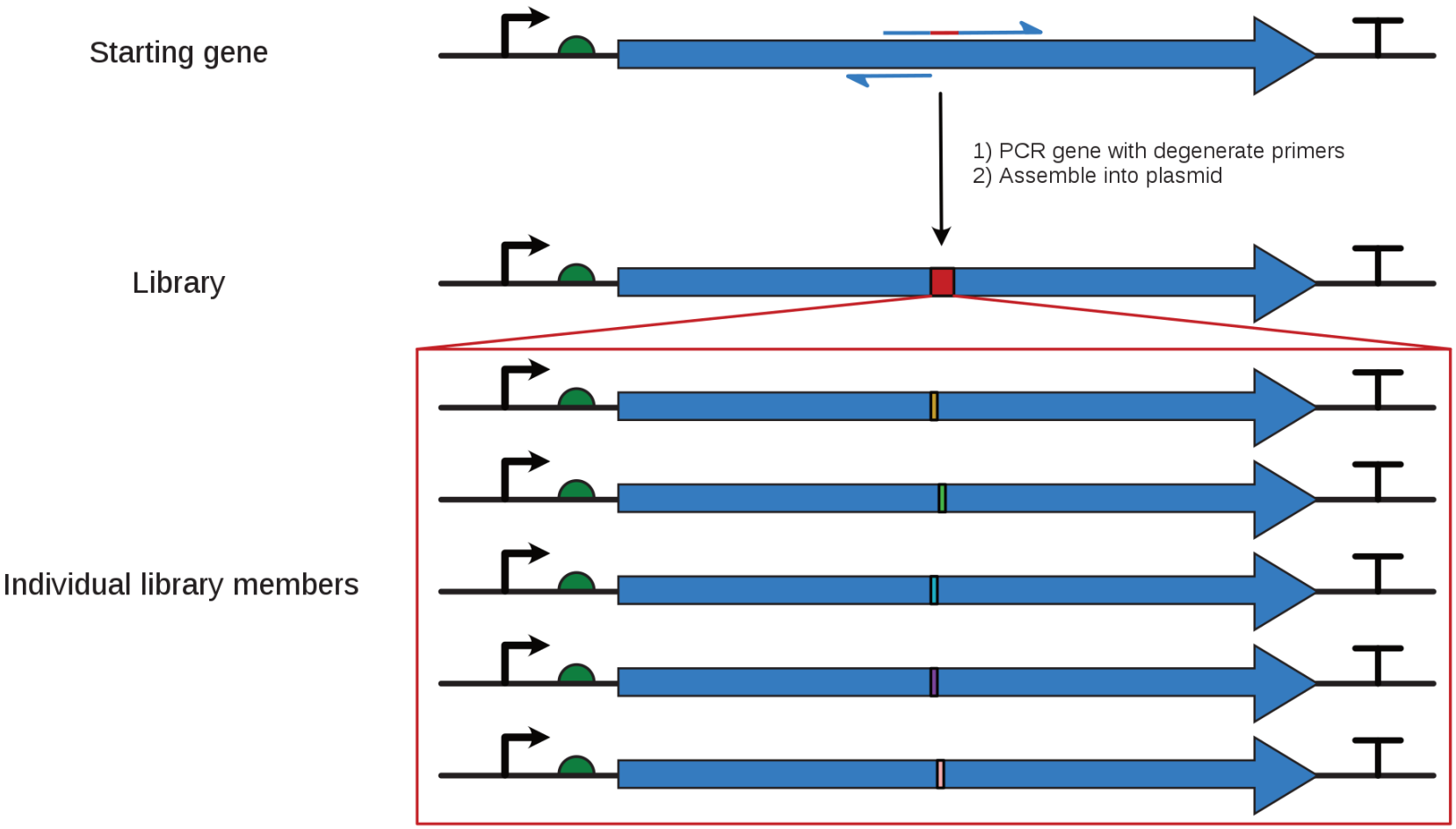


Site-saturation mutagenesis (SSM)

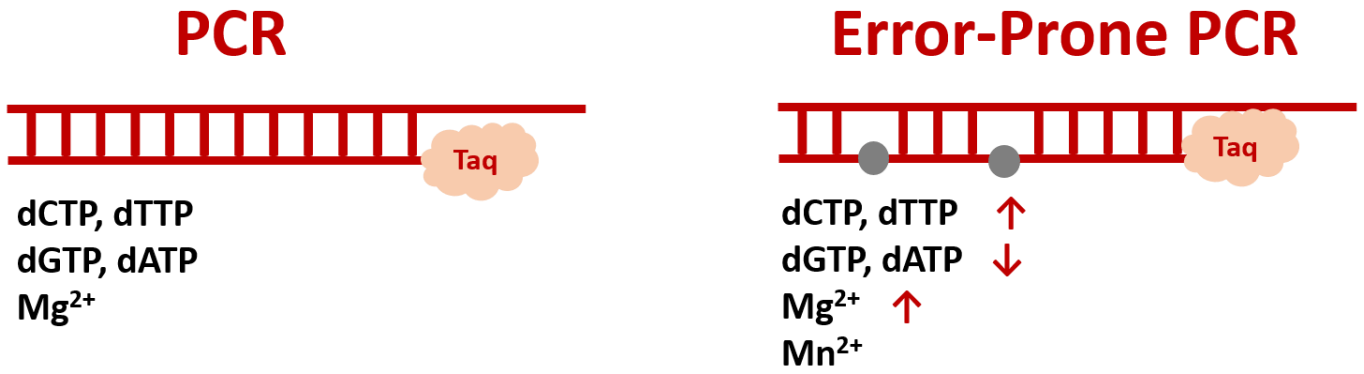
- Site saturation mutagenesis is used to substitute targeted residues to any other naturally occurring amino acid
- The core of a SSM experiment lies in the codon degeneracy or randomness. A completely randomized codon (NNN, where N=A, C, G or T) results in a library size of 64 different sequences encoding all 20 amino acids and 3 stop codons
- When an experiment targets multiple codons, the library size can be considerably higher, making it difficult to perform a complete screening (e.g. targeting three NNN codons has 262,144 unique codon configurations



Site-saturation mutagenesis (SSM)



Error-prone PCR (epPCR)

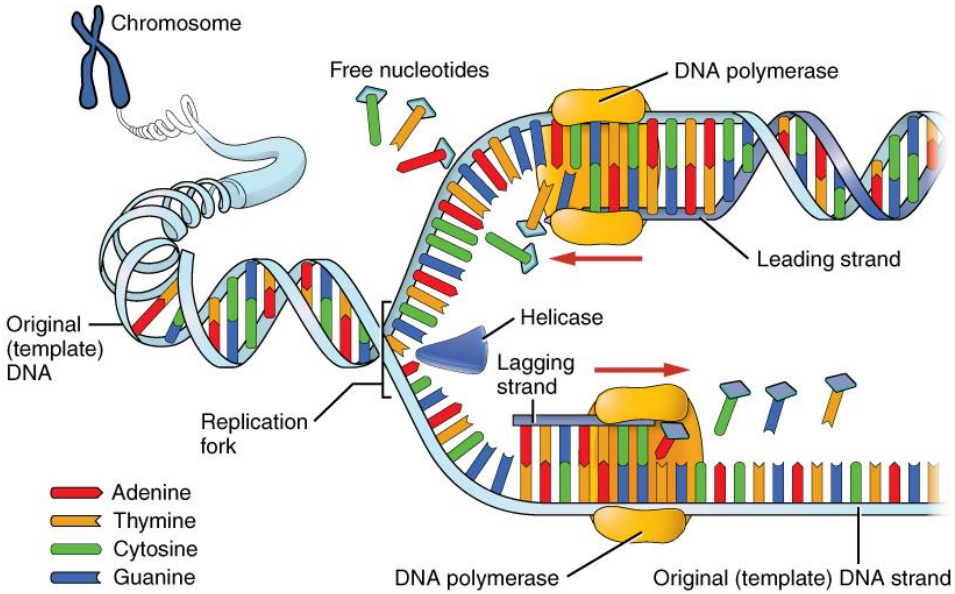


- The error rate of *Taq* DNA polymerase is 0.001-0.002 % per nucleotide per replication cycle under standard conditions which is sufficient to create mutant libraries of large genes but not for small genes
- Error-prone PCR (epPCR) takes advantage of the inherently low fidelity of *Taq* DNA Polymerase, which may be further decreased by the addition of Mn^{2+}, increasing the Mg^{2+} concentration, and using unequal dNTP concentrations.
- The rate of mutagenesis achieved by error-prone PCR is in the range of 0.6-2.0 %

https://lifescience.canvaxbiotech.com/product/pickmutant-error-prone-pcr-kit/

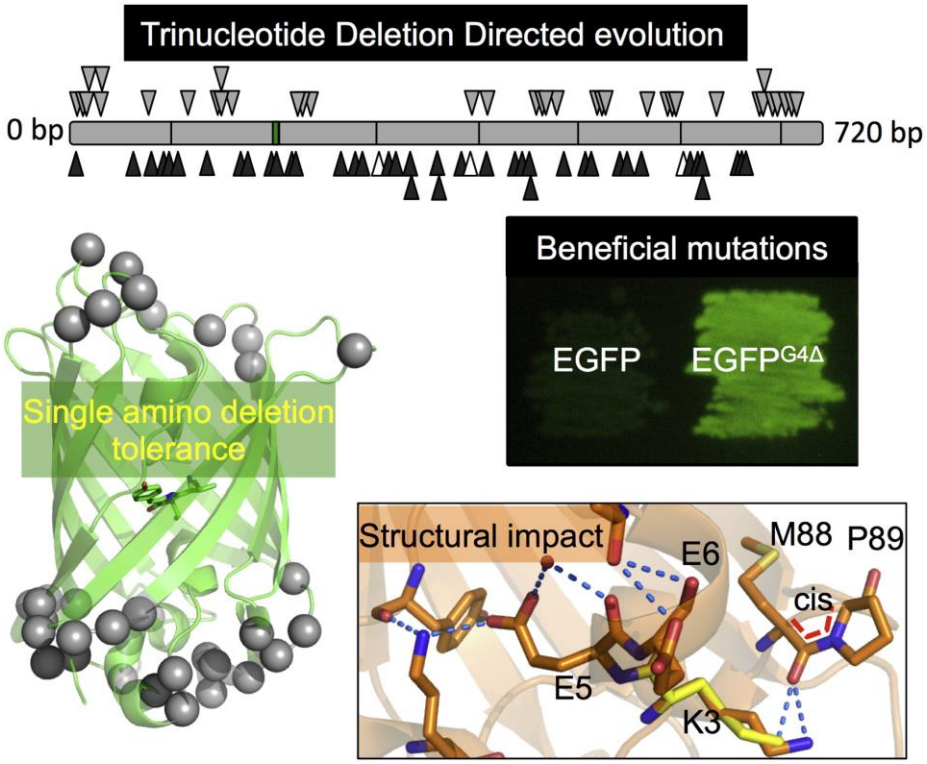
Mutator strains

- Mutator strains of *E*. *coli* are deficient in one or more of DNA repair genes, leading to single base **substitutions** at a rate of approximately 1 mutation per 1000 base pairs
- Generation of mutant libraries
- Process is simple



Insertion and deletion (InDel) mutagenesis

- Gain or lost of one or more nucleotides produces frameshift mutations (triplet reading frame)
- Triplet InDel mutagenesis may trigger protein backbone changes essential for evolvability
- Insertion and deletion mutations can enhance proteins through **structural rearrangements** not possible by substitution mutations alone

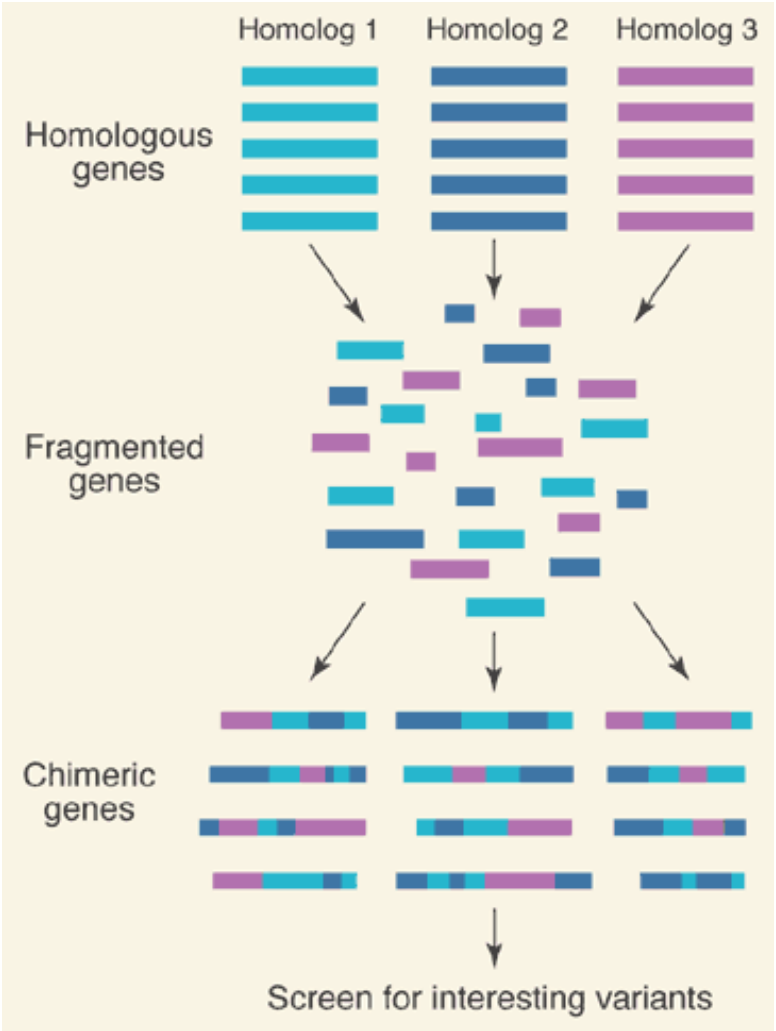


Using directed evolution, green fluorescent protein (GFP) was observed to tolerate residue deletions, particularly within short and long loops, helical elements, and at the termini of strands. A variant with G4 removed from a helix (EGFPG4Δ) conferred significantly higher cellular fluorescence.

Arpino et al., Structure 22: 889-898 (2014)

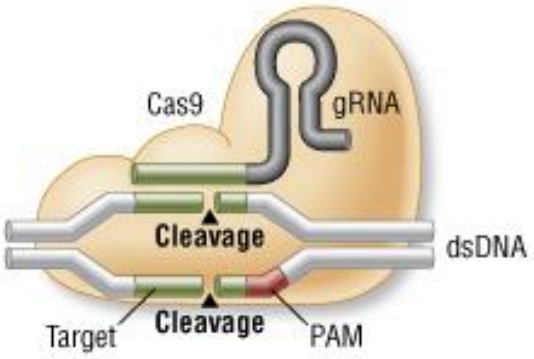
DNA shuffling

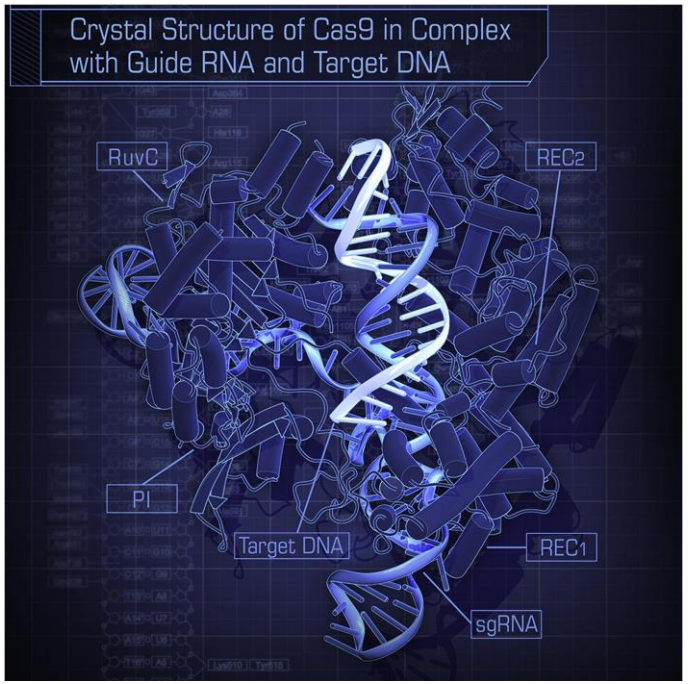
- DNA shuffling is a method for *in vitro* recombination of homologous genes
- The genes to be recombined are randomly fragmented by DNaseI, and fragments of the desired size are purified from an agarose gel
- These fragments are then reassembled using cycles of denaturation, annealing, and extension by a polymerase
- Recombination occurs when fragments from different parental templates anneal at a region of high sequence identity
- Following this reassembly reaction, PCR amplification with primers is used to generate full-length chimeras suitable for cloning into an expression vector
- Moving from DNA shuffling to whole genome shuffling is known as **GENOME SHUFFLING**



The CRISPR/Cas9 system

- The **Cas9** (CRISPR associated protein 9) is a protein which plays a vital role in the immunological defense of bacteria against DNA viruses, and which is used in genetic engineering. Its main function is to cut DNA and therefore it can alter a cell's genome
- Structurally, Cas9 is an **RNA-guided DNA endonuclease enzyme** associated with CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) adaptive immunity system in *Streptococcus pyogenes*
- Cas9 performs this by unwinding foreign DNA and checking for sites complementary to the 20 bp spacer region of the guide RNA
- If the DNA substrate is complementary to the guide RNA, Cas9 cleaves the invading DNA

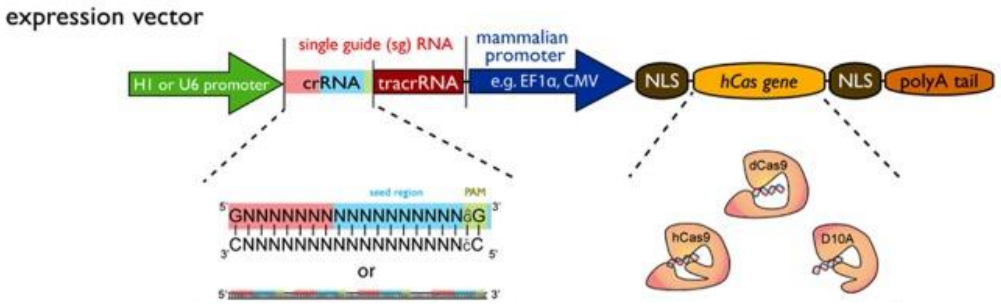


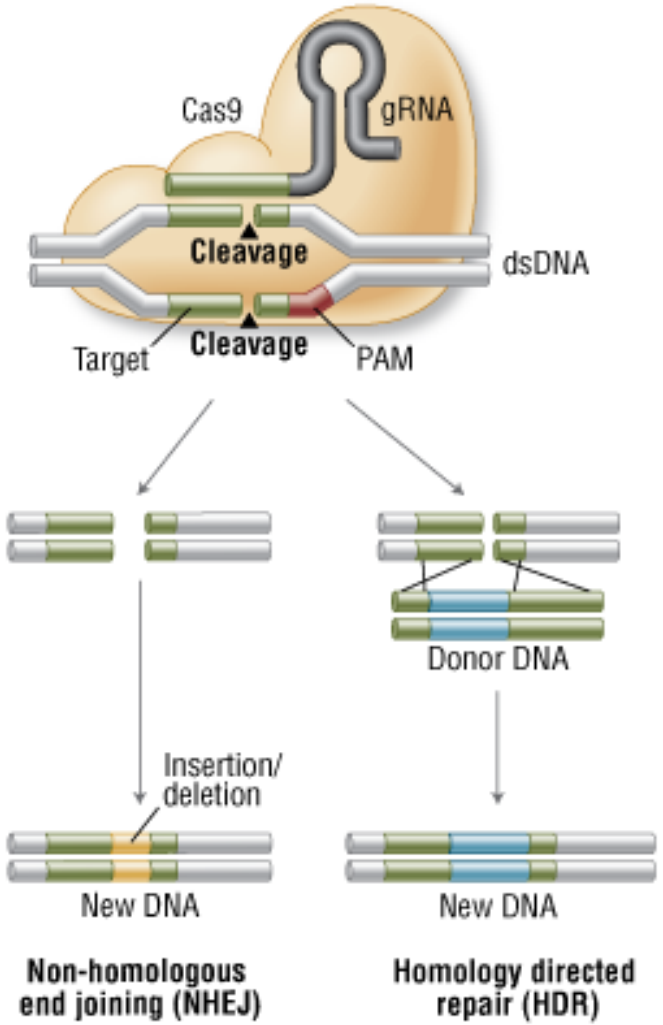


Crystal structure of *S. pyogenes* Cas9 in complex with sgRNA and its target DNA at 2.5 Å resolution, Nishimazu *et al*., Cell 156: 935–49 (2014)

The CRISPR/Cas9 system: key elements

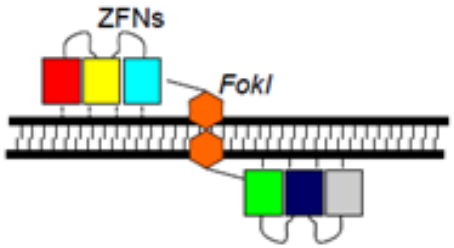
- Cas9 nuclease specifically cleaves double-stranded DNA activating double-strand break repair machinery
- In the absence of a homologous repair template non-homologous end joining can result in indels disrupting the target sequence
- Alternatively, precise mutations and knock-ins can be made by providing a homologous repair template and exploiting the homology directed repair pathway



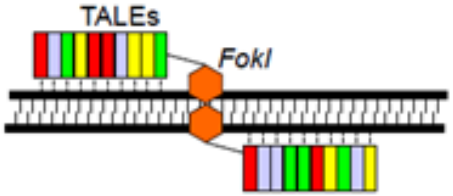


Advantages of CRISPR/Cas9-mediated mutagenesis

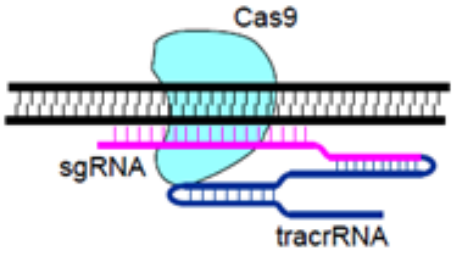
- The CRISPR/Cas9 system requires only the redesign of the crRNA to change target specificity
- This contrasts with other genome editing tools, including zinc finger and TALENs, where redesign of the protein-DNA interface is required
- Furthermore, CRISPR/Cas9 enables rapid genome-wide interrogation of gene function by generating large gRNA libraries for genomic screening



Zinc-finger nucleases

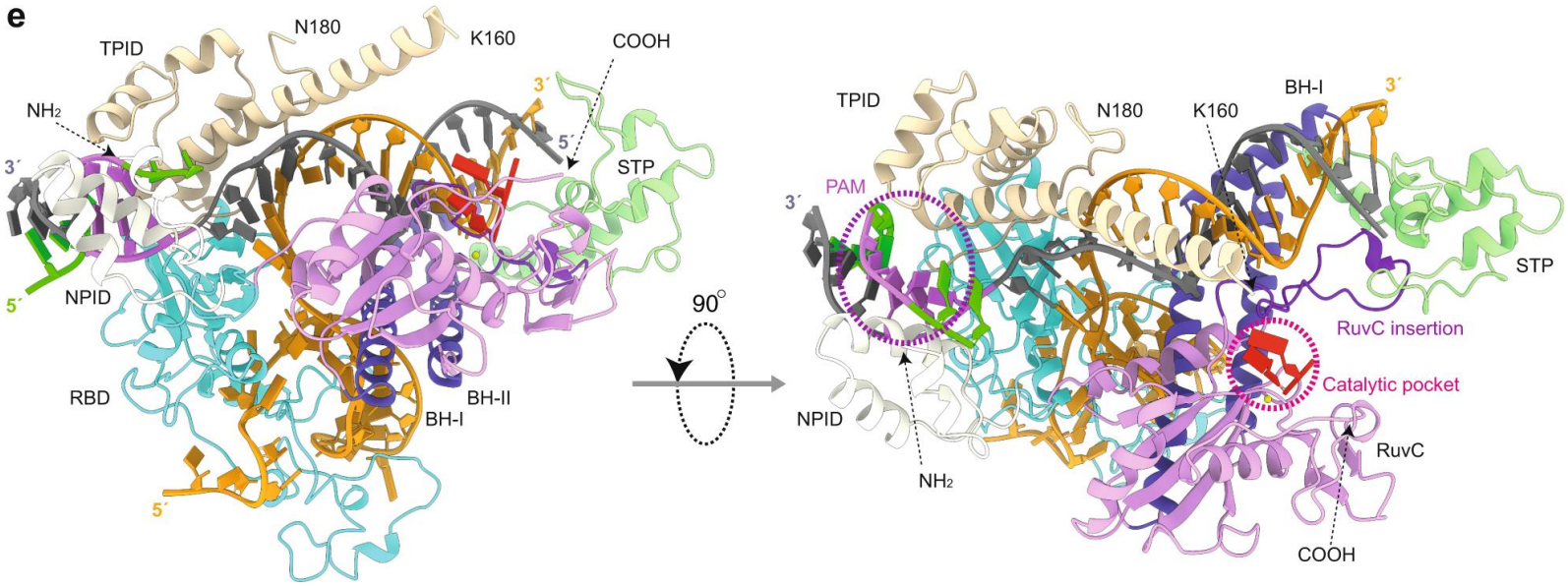


TALENs



Cas9

The mini-RNA-guided endonuclease CRISPR-Cas12j3



- CRISPR-Cas12j is a recently identified family of miniaturized RNA-guided endonucleases from phages.
- These ribonucleoproteins provide a compact scaffold gathering all key activities of a genome editing tool.
- A site-directed mutagenesis analysis supports the DNA cutting mechanism, providing new avenues to redesign CRISPR-Cas12j nucleases for genome editing.

nature communications

Explore content | About the journal | Publish with us

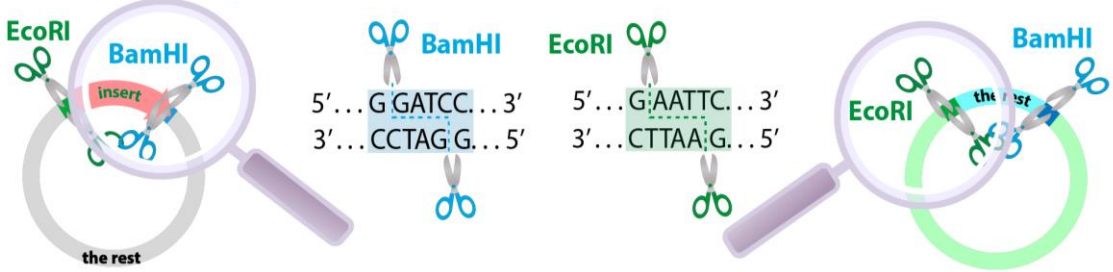
nature > nature communications > articles > article

Article | Open Access | Published: 22 July 2021

Molecular carpentry

- DNA polymerases
- DNases
- Exonucleases
- Restriction endonucleases
- Kinases
- Phosphatases
- Ligases
- Etc.

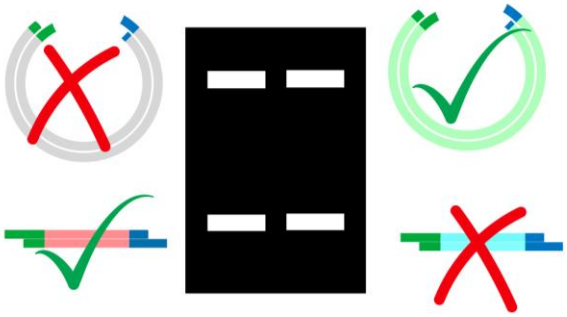
cut insert (what you want to put in) and **vector** (home you want to put it in) with the same 2 restriction enzymes



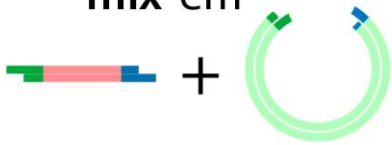
This generates DNA pieces with complementary "sticky ends" you can mix & match (once you separate them)

purify the pieces you want

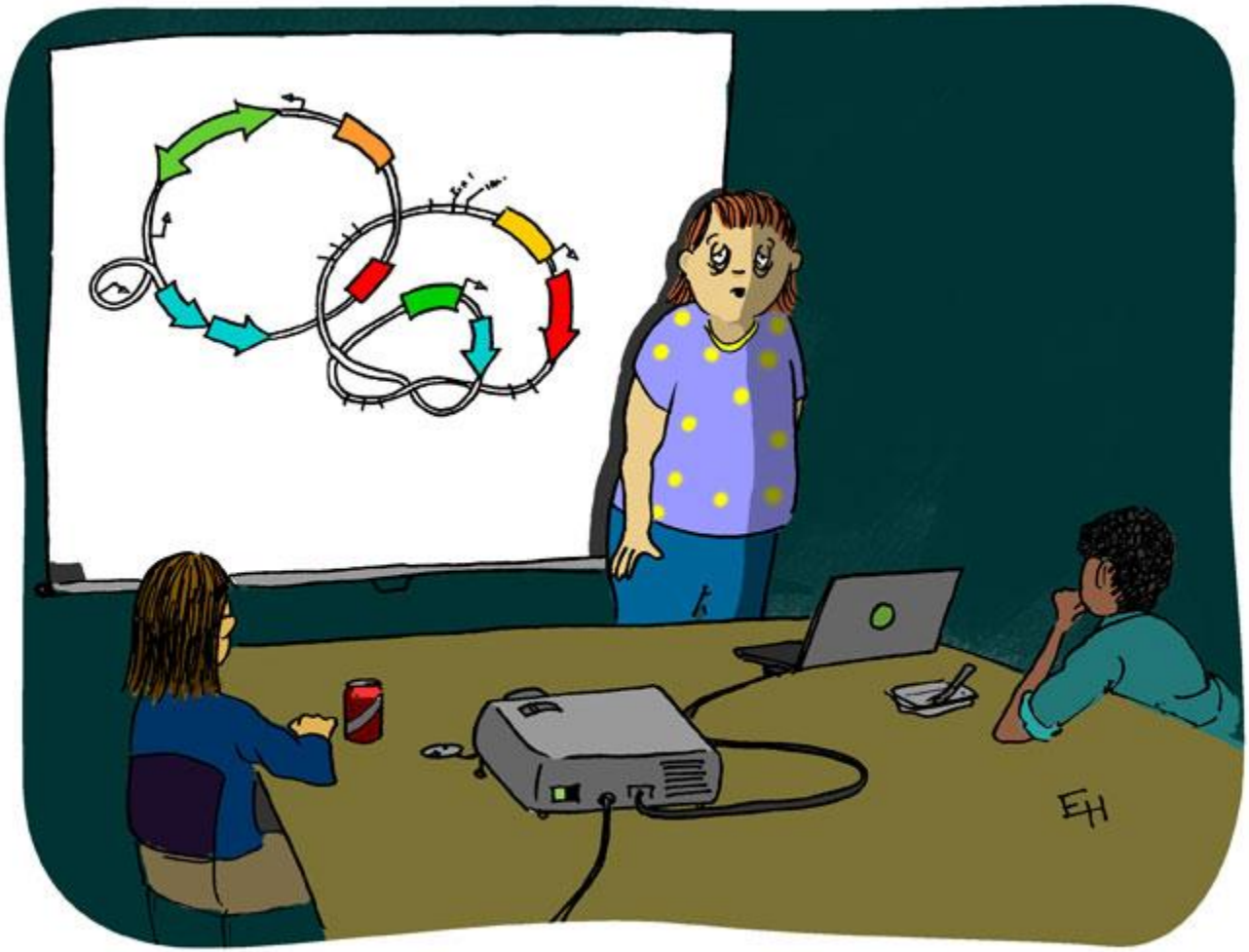
run an agarose gel to separate by size, then cut out & purify those you want



mix 'em



stitch 'em



I wish I could report otherwise,
but the cloning is not going very well.

Molecular biology in protein technologies

Computational protein design

1. Gene synthesis & cloning

- *De novo* DNA synthesis
- Gene assembly
- Site-directed mutagenesis

2. Protein expression & purification

- Expression host
- Affinity and solubility tags
- Buffer screening

3. Biophysical & biochemical characterization

- CD spectroscopy
- Differential scanning fluorimetry
- Dynamic light scattering
- SAXS

4. 3D structure determination

- NMR spectroscopy
- Cryo-EM
- X-ray crystallography

5. Biological assays

- Biocatalysis
- Binding assays
- Protein-protein interaction
- Photoactivation

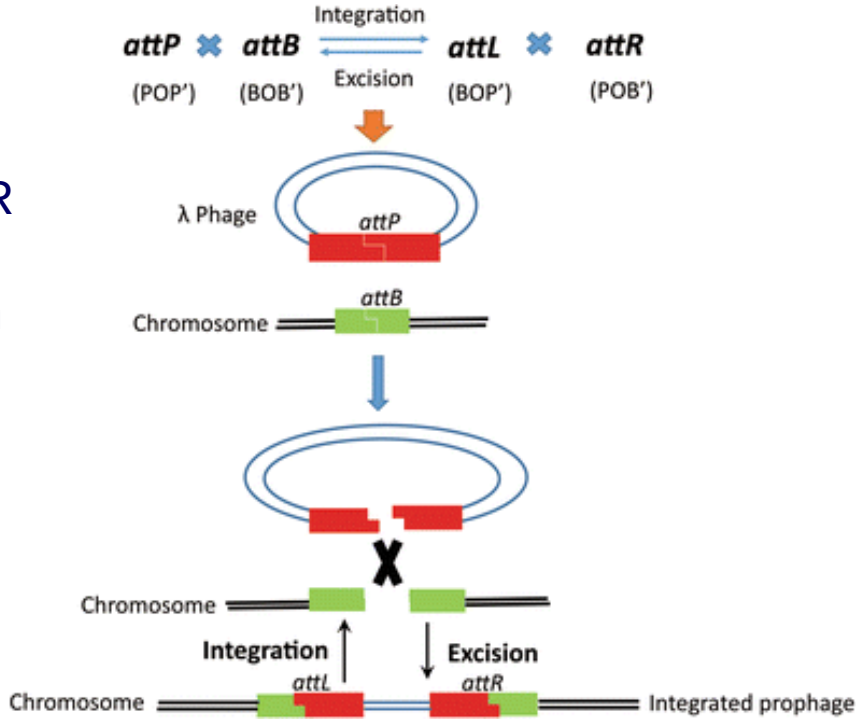
Questions

Supplementary materials

Gateway cloning

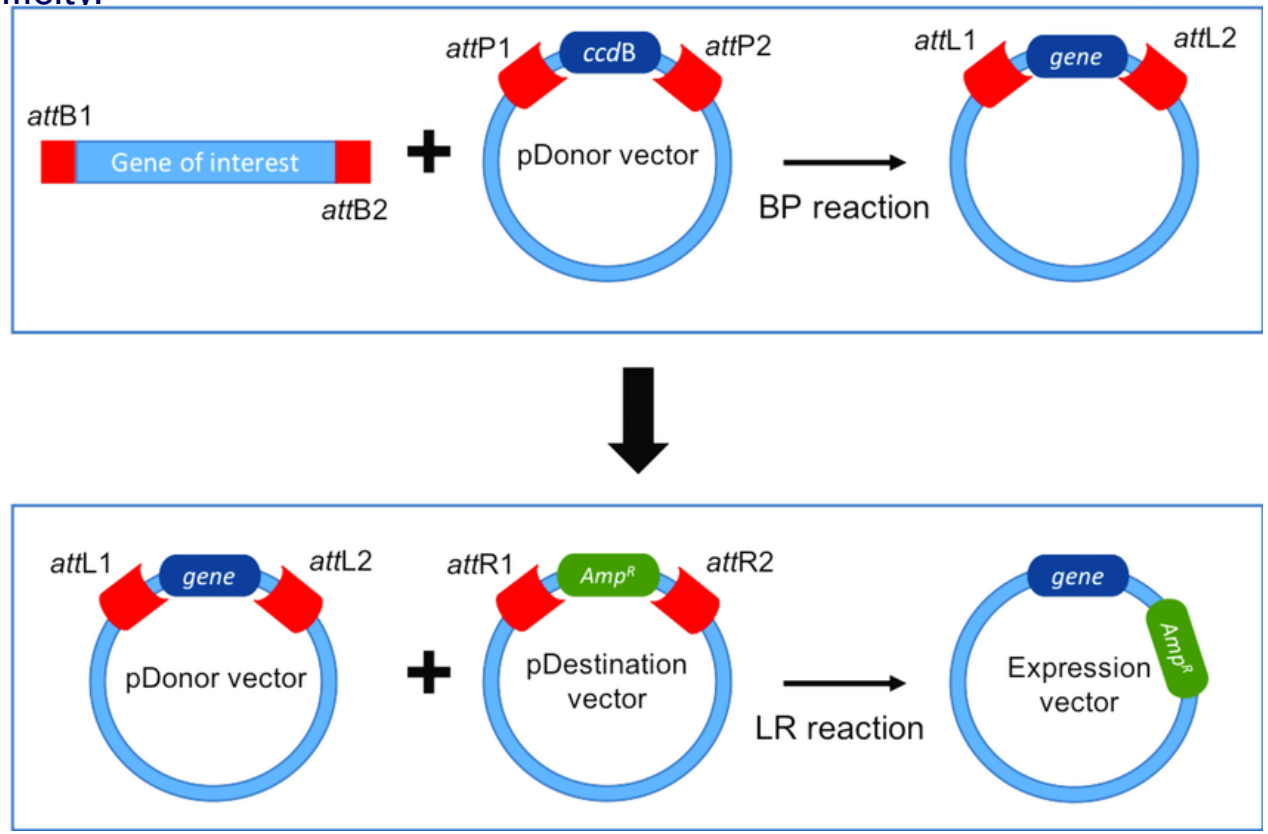
The GATEWAY Cloning Technology is based on the site-specific recombination system used by phage λ to integrate its DNA in the *E. coli* chromosome. Both organisms have specific recombination sites called *att*P in phage λ site and *att*B in *E. coli*. The integration process (lysogeny) is catalyzed by 2 enzymes: the phage λ encoded protein Int (Integrase) and the *E. coli* protein IHF (Integration Host Factor). Upon integration, the recombination between *att*B (25 nt) and *att*P (243 nt) sites generate *att*L (100 nt) and *att*R (168 nt) sites that flank the integrated phage l DNA.

The process is reversible and the excision is again catalyzed Int and IHF in combination with the phage λ protein Xis. The *att*L and *att*R sites surrounding the inserted phage DNA recombine site-specifically during the excision event to reform the *att*P site in phage λ and the *att*B site in the *E. coli* chromosome.

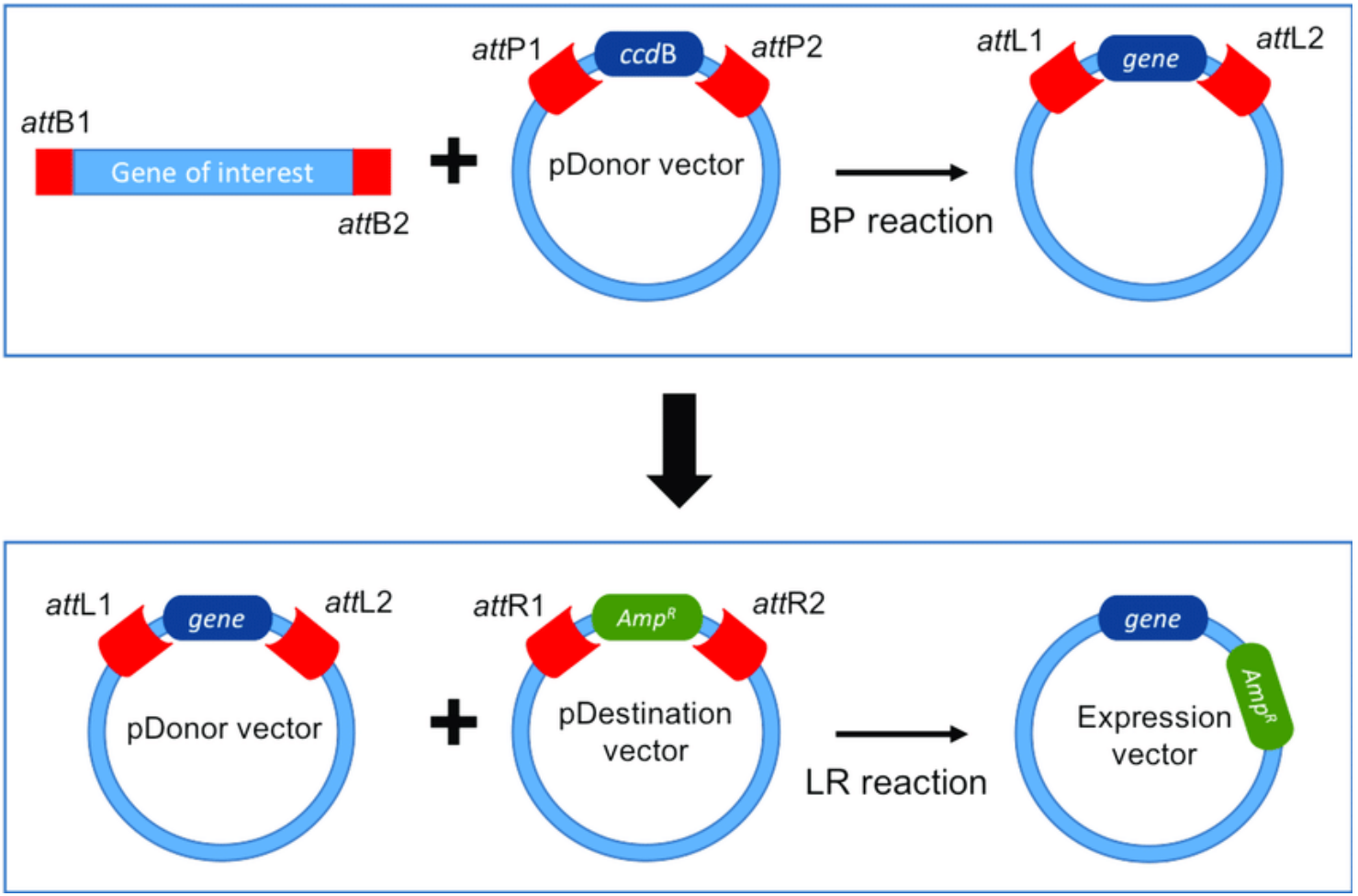


Gateway cloning

The GATEWAY reactions are *in vitro* versions of the integration and excision reactions. To make the reactions directional two slightly different and specific site were developed, *att*1 and *att*2 for each recombination site. These sites react very specifically with each other. For instance in the BP Reaction *att*B1 only reacts with *att*P1 resulting in *att*L1 and *att*R1, and *att*B2 only with *att*P2 giving *att*L2 and *att*R2. The reverse reaction (LR Reaction) shows the same specificity.



Gateway cloning



- Lambda bacteriophage site specific integration system

SLIC Sequence and Ligation Independent Cloning

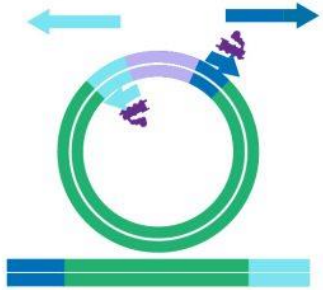
start with **PCR** to make copies of insert piece & vector piece with overlapping regions on the ends

something with **insert** (such as gene) you want

vector you want to put it in, such as a plasmid

design your PCR primers to have bits of the flanking vector sequences on their ends

note: anything can be in the purple part or gray part - we're not copying those parts so they don't matter



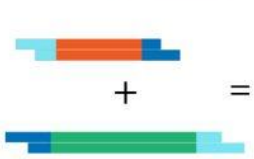
add **DpnI** to degrade leftover parent plasmids

it only cuts methylated DNA - parent plasmids are methylated but PCR products aren't

use **T4 polymerase** to chew the ends back (a little) so they'll stick

this exposes complementary overhangs

mix pieces together



no need to pre-ligate

& **stick in bacteria** to do the rest

homologous recombination machinery will fill in the gaps

PCR

Polymerase Chain Reaction

copying DNA the (tiny) test tube way!

design **primers** (1 per strand) - short DNA sequences bookending the region of DNA you want copied

each cycle

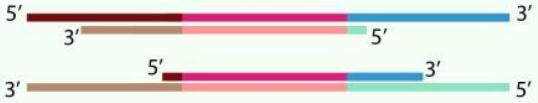
melt - raise temp to separate strands

then **anneal** - lower temp so primers can bind

and **extend** - let the DNA Pol add matching bases

1st cycle

end where Pol runs out of steam or time or falls off



now when you melt, these newly-made strands can get used as templates

2nd cycle

end where 1st primer started because template stops there & Pol falls off

do it again and again and again

all future cycles

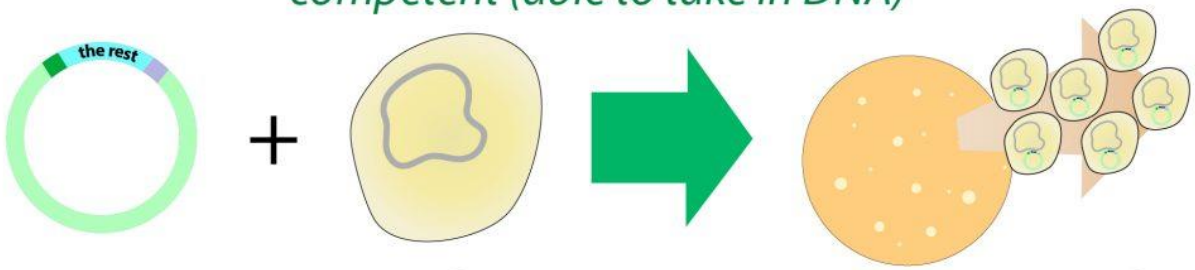
amplify pieces bookended by the primer start sites

Cloning Controls

antibiotic resistance gene in the vector allows for antibiotic selection - grow with that antibiotic so only bacteria with the vector can grow

uncut vector

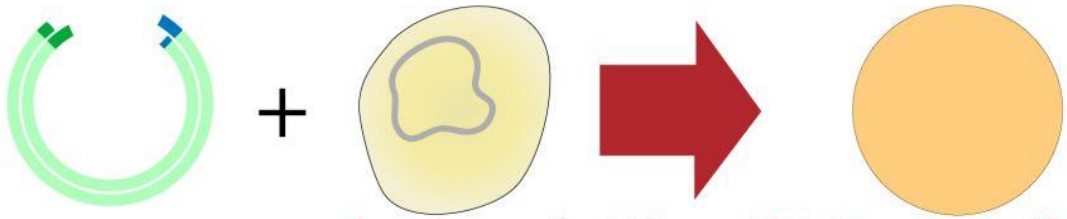
***positive control** that transformation worked & cells are competent (able to take in DNA)*



each colony's made up of genetically-identical bacteria (hopefully with your insert)

cut vector

negative control that all vector got cut so you didn't have any "parent" vectors & or self-circularized vector

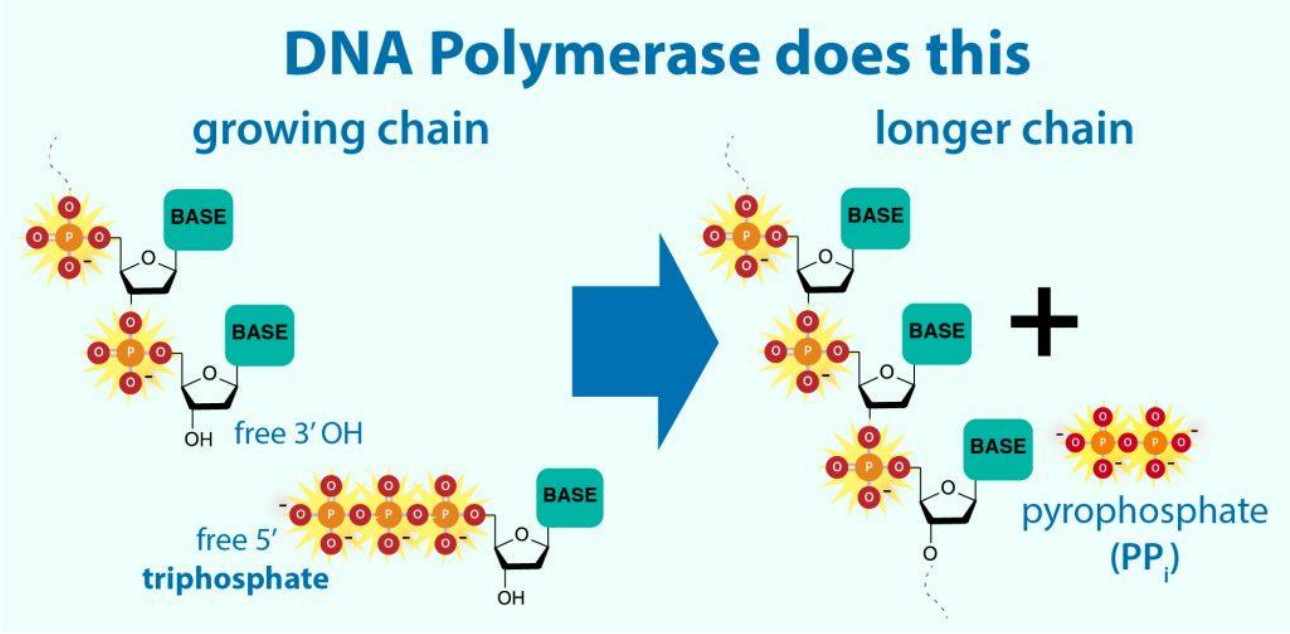


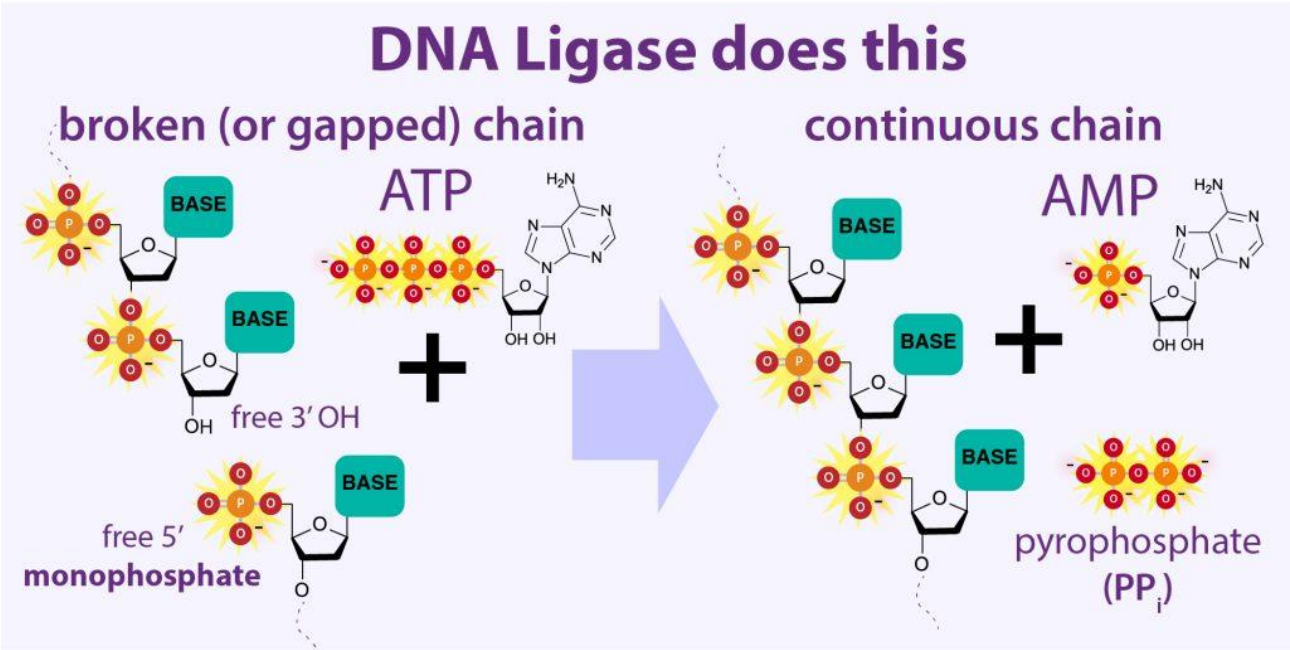
also shows that the antibiotics were effective

Gateway cloning

The GATEWAY Cloning Technology is based on the site-specific recombination system used by phage λ to integrate its DNA in the *E. coli* chromosome. Both organisms have specific recombination sites called *att*P in phage λ site and *att*B in *E. coli*. The integration process (lysogeny) is catalyzed by 2 enzymes: the phage λ encoded protein Int (Integrase) and the *E. coli* protein IHF (Integration Host Factor). Upon integration, the recombination between *att*B (25 nt) and *att*P (243 nt) sites generate *att*L (100 nt) and *att*R (168 nt) sites that flank the integrated phage l DNA.

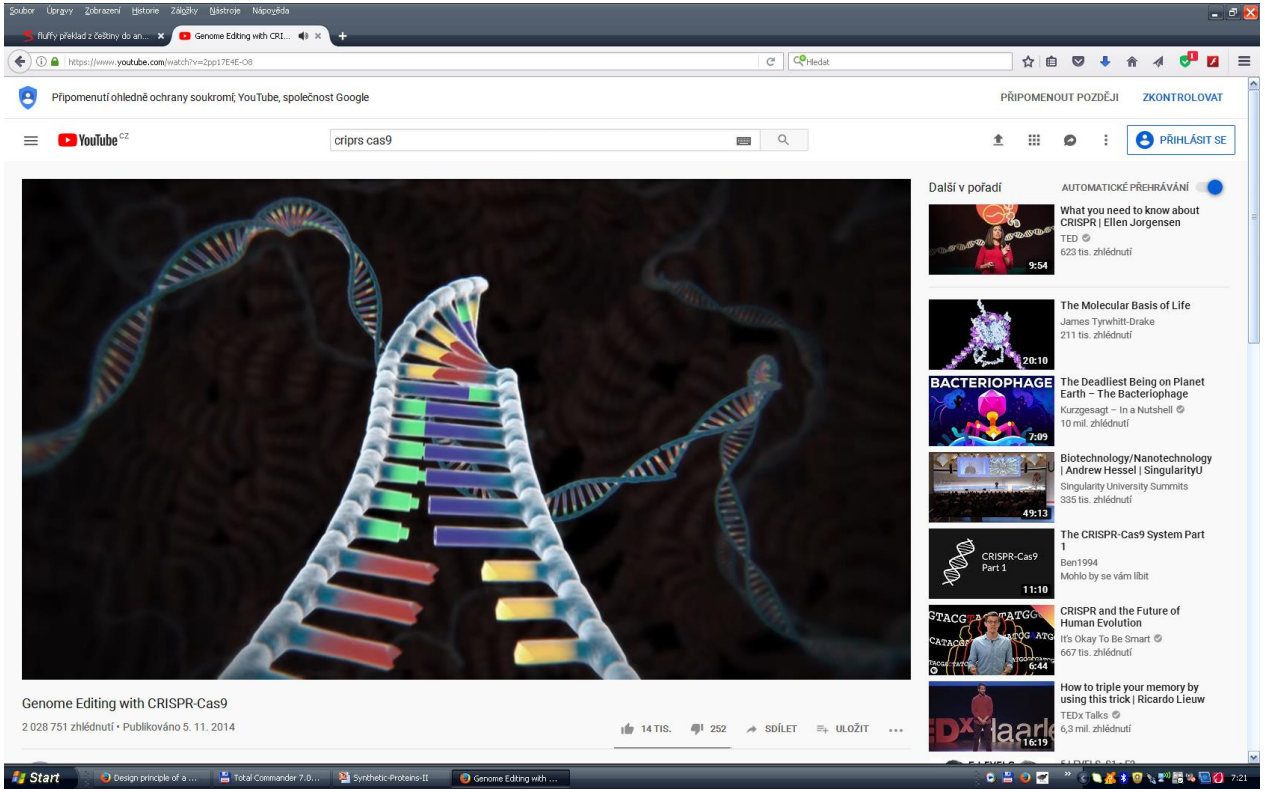
The process is reversible and the excision is again catalyzed Int and IHF in combination with the phage λ protein Xis. The *att*L and *att*R sites surrounding the inserted phage DNA recombine site-specifically during the excision event to reform the *att*P site in phage λ and the *att*B site in the *E. coli* chromosome.


DNA Polymerase does this
growing chain
longer chain
BASE
BASE
free 3′ OH
OH
free 5′
triphosphate
BASE
BASE
BASE
pyrophosphate
(PPi)

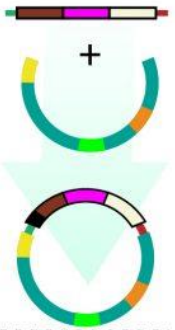

DNA Ligase does this
broken (or gapped) chain
continuous chain
ATP
AMP
BASE
BASE
free 3′ OH
OH
free 5′
monophosphate
BASE
BASE
BASE
pyrophosphate
(PPi)

The CRISPR/Cas9 system on YouTube

https://www.youtube.com/watch?v=bXnWIk8FgKc
https://www.youtube.com/watch?v=OjNrbPMXyMA
https://www.youtube.com/watch?v=0dRT7slyGhs
https://www.youtube.com/watch?v=2pp17E4E-O8



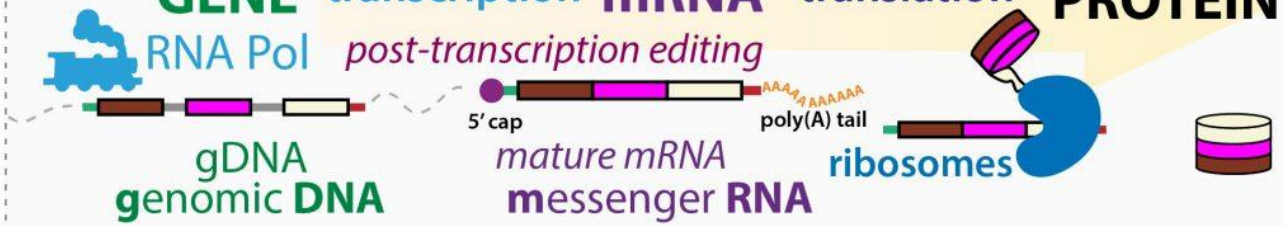
recombinant DNA/protein



molecular cloning take a segment of DNA from its home & stick it into a piece of DNA that's easier to work with

since you've "recombined" DNA, we call this **recombinant DNA**

when we do this with DNA containing protein instructions, the protein made from it is referred to as **recombinant protein**



cDNA complementary DNA — DNA version of the edited copy (mRNA) we put into a vector plasmid for protein expression

a common place to put the DNA is a circular piece of DNA called a **plasmid**

a plasmid can act as a **vector** ("vehicle") for getting that DNA into bacterial cells

note: we can use different vectors for different cell types/organisms

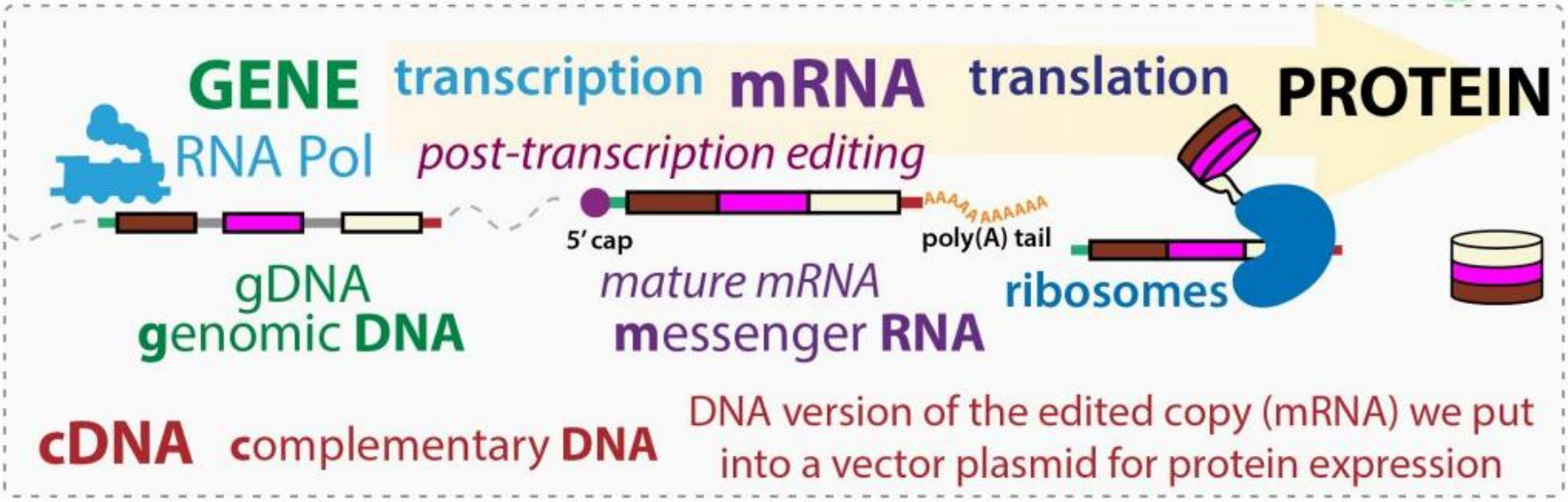
vector plasmids have some key features...

promoter - allows for transcription & subsequent translation

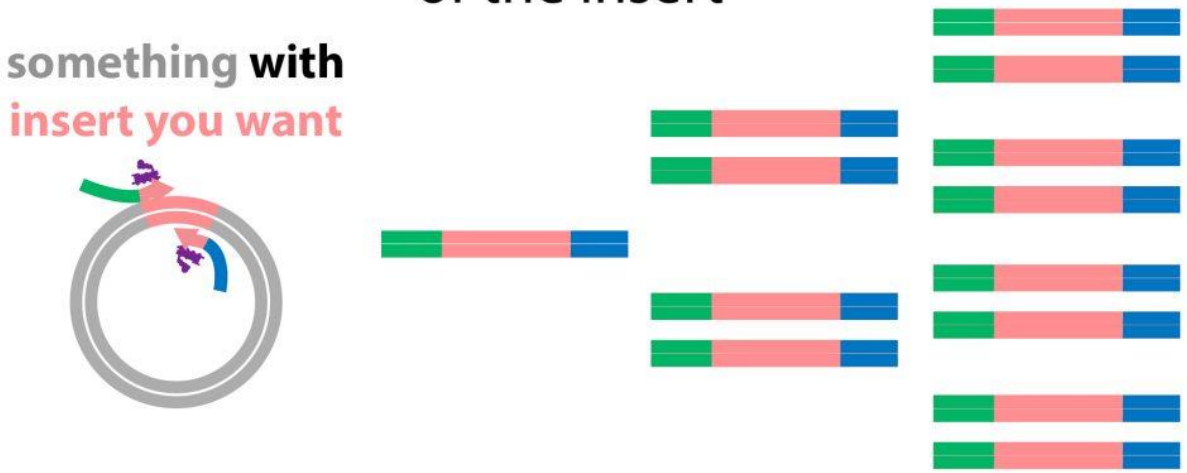
selection marker such as antibiotic resistance gene - allows for selection for cells with your plasmid

DNA Pol **ORI** - allows for copying of plasmid

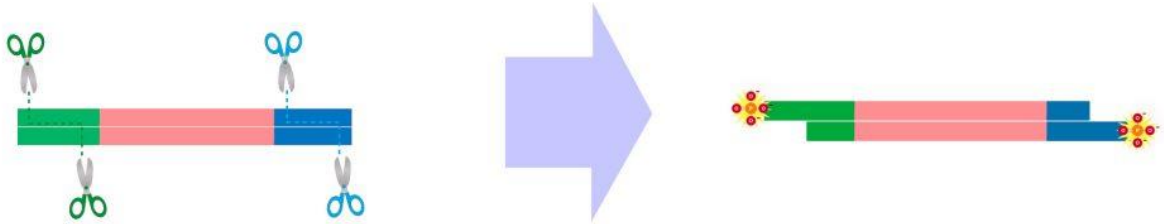
since we control what DNA we put in, we can make changes to it to "custom-make" proteins, add tags for easier purification, etc.


GENE
transcription
mRNA
translation
PROTEIN
RNA Pol
post-transcription editing
5' cap
AAAAAAAAAA
poly(A) tail
gDNA
genomic DNA
mature mRNA
messenger RNA
ribosomes
cDNA
complementary DNA
DNA version of the edited copy (mRNA) we put into a vector plasmid for protein expression

you can use use PCR to make lots of copies of the insert



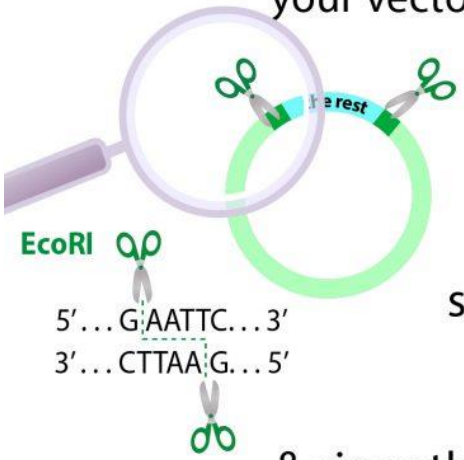
and you can use the primers to add on cut sites you want

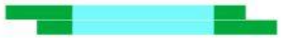


this cutting leaves you with phosphorylated ends but the primers are usually synthesized without phosphates - this only comes into play if your vector is dephosphorylated

dephosphorylation/phopshorylation ***may*** be needed

if you're using a single REase, you'll need to dephosphorylate your vector to prevent self-circularization





this now has sticky ends for itself

so when you add ligase, it can self-ligate

& since the antibiotic resistance gene is in there, it can survive - but it doesn't have your gene

you can prevent self-circularization by dephosphorylating the ends

+ phosphatase

now ligase can't seal it

but you need your insert to be phosphorylated!

you'll still have a couple of nicks, which the bacteria can fix

Dr. Martin Marek
Loschmidt Laboratories
Faculty of Science, MUNI
Kamenice 5, bld. A13, room 332
martin.marek@recetox.muni.cz

LOSCHMIDT
LABORATORIES