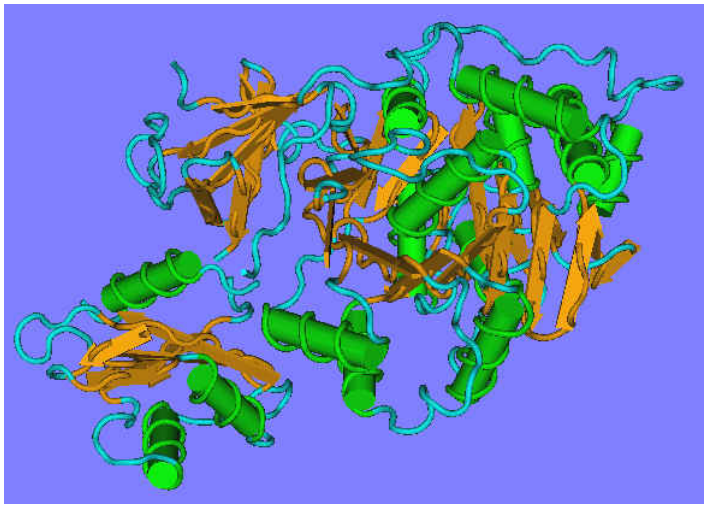
# HTS in Bio Catalysis

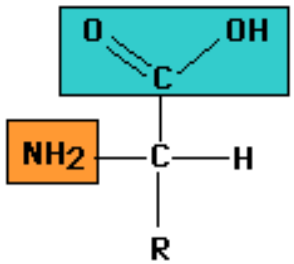
**Uwe Dingerdissen**

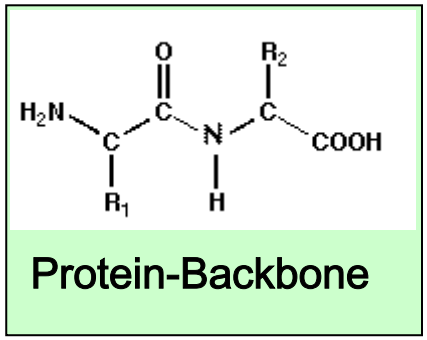
**Leibniz Institute for Catalysis**
**(former ACA- Berlin)**

# The Genetic Code - Alphabet of Life

## Phenotyp Level: Protein

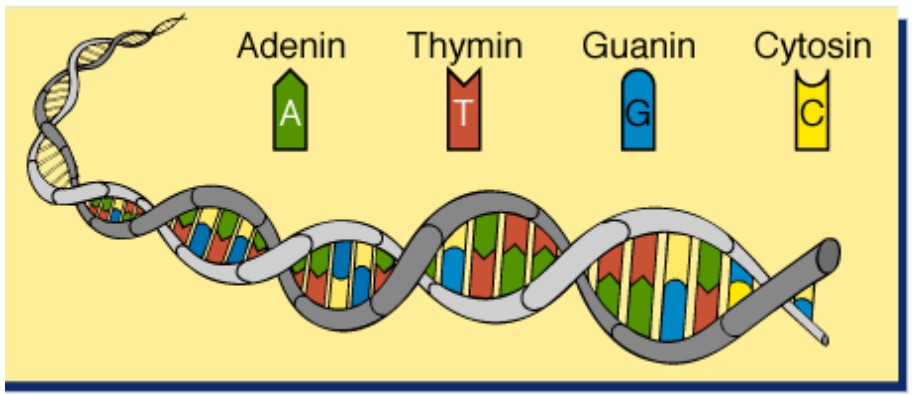




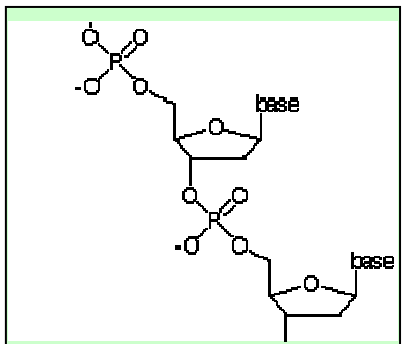

Protein-Backbone

## Genotype Level: DNA



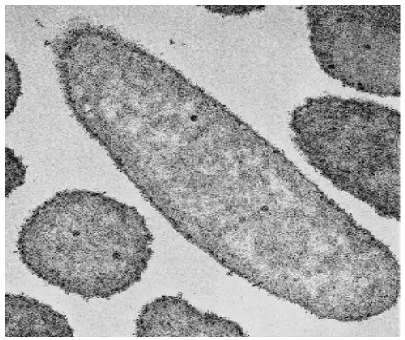

Nucleotides

DNA-Helix

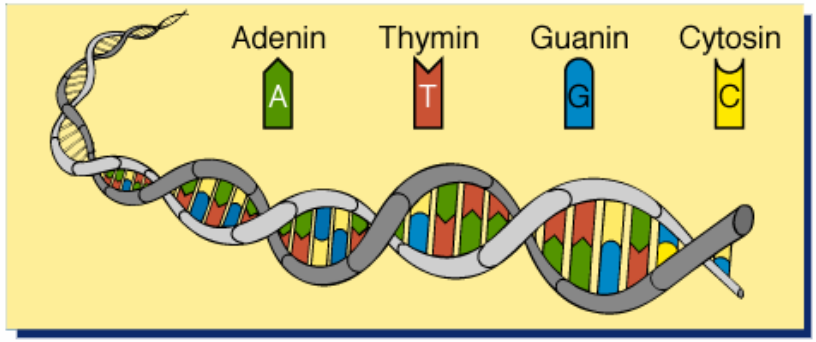



DNA-Backbone

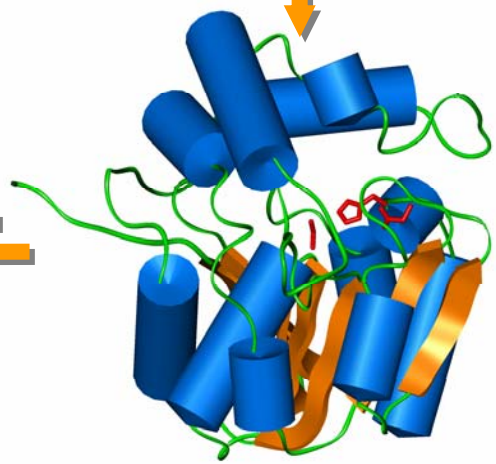
# Biocatalysis:



Procaryotic Organism




Gene

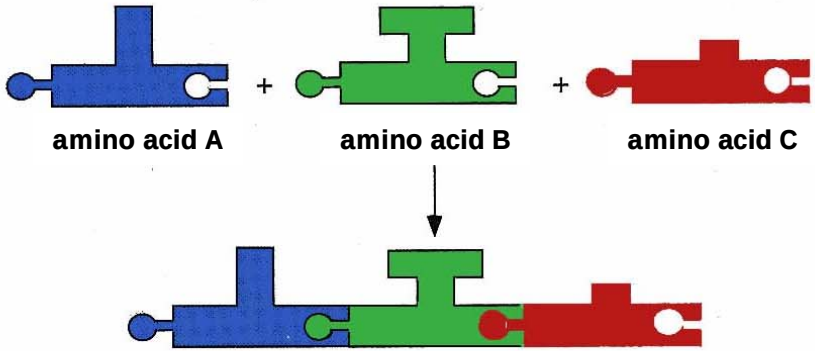


Enzyme

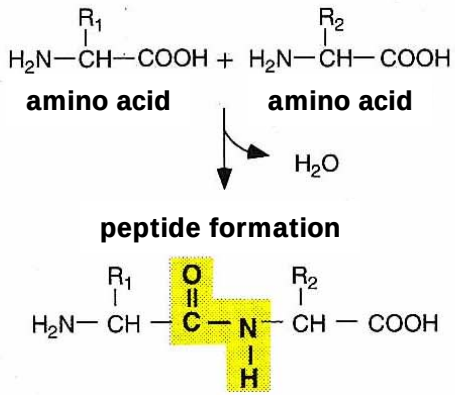
Process

# Basics: Amino Acids LEGO-Chemistry of Amino Acids

A

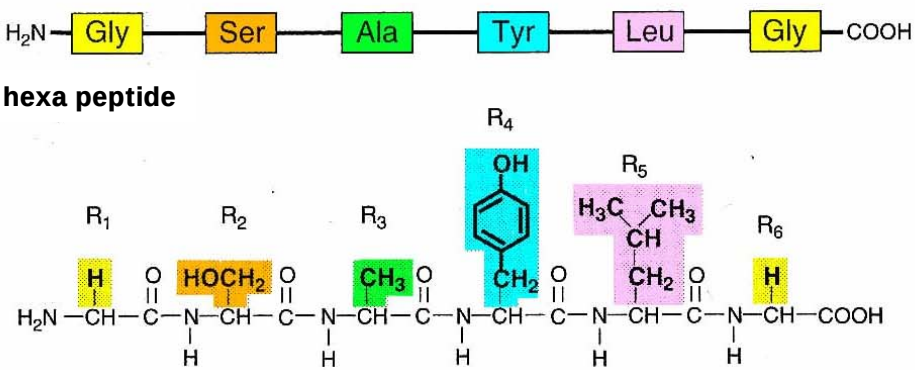



B




**peptide formation = condensation**

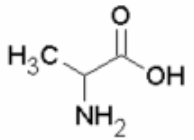
C

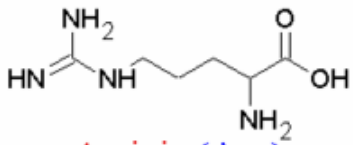


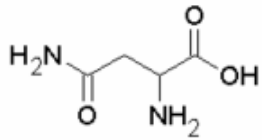

**small chain length = peptide**

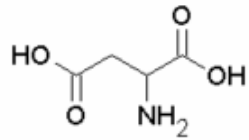
**chain length >100 amino acids= proteins**

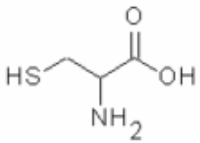
# Basics: Amino Acids
## how many natural amino acids exists ?

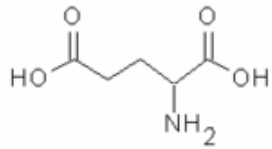

Alanin **(Ala)**

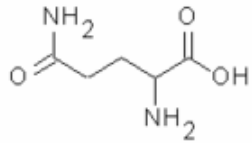

Arginin **(Arg)**

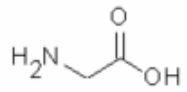

Asparagin **(Asn)**

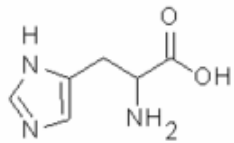

Asparaginsäure **(Asp)**

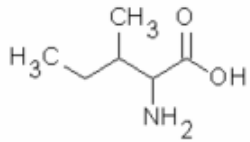

Cysteine **(Cys)**

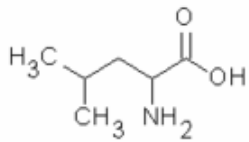

Glutaminsäure **(Glu)**

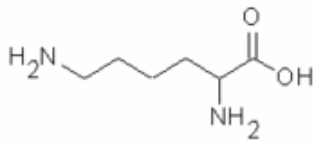

Glutamin **(Gln)**

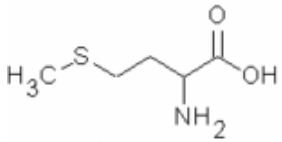

Glycin **(Gly)**

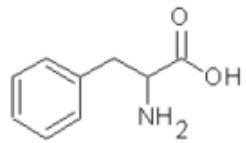

Histidin **(His)**

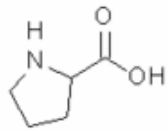

Isoleucin **(Ile)**

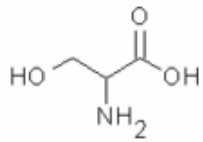

Leucin **(Leu)**

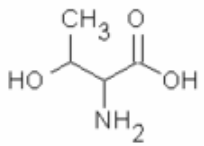

Lysin **(Lys)**

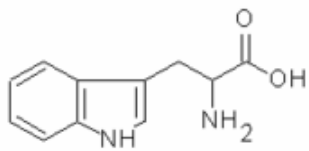

Methionin **(Met)**

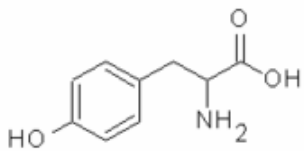

Phenylalanin **(Phe)**

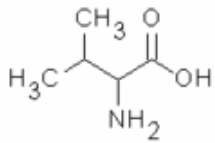

Prolin **(Pro)**


Serin **(Ser)**


Threonin **(Thr)**


Tryptophan **(Trp)**


Tyrosin **(Tyr)**


Valin **(Val)**

# Basics: Amino Acids
## how many natural amino acids exists ?

| building blocks of the proteins | | | |
|---|---|---|---|
| **amino acid** | **abbreviation** | **amino acid** | **abbreviation** |
| Glycine | Gly, G | Methionine | Met, M |
| Alanine | Ala, A | Tryptophane | Trp, W |
| Valine | Val, V | Tyrosine | Tyr, Y |
| Leucine | Leu, L | Asparagine | Asn, N |
| Isoleucine | Ile, I | Glutamine | Gln, Q |
| Phenylalanine | Phe, F | Aspartic Acide | Asp, D |
| Proline | Pro, P | Glutamic Acide | Glu, E |
| Serine | Ser, S | Lysine | Lys, K |
| Threonine | Thr, T | Arginine | Arg, R |
| Cysteine | Cys, C | Histidine | His, H |

# Basics: DNA, RNA
# J.D. Watson, F. H. C. Crick, Nature 1953

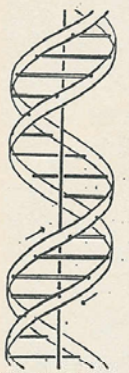
MOLECULAR STRUCTURE OF NUCLEIC ACIDS

**A Structure for Deoxyribose Nucleic Acid**

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.



This figure is purely diagrammatic. The two ribbons symbolise the two phosphate–sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxyribofuranose residues with 3′,5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made a... 1 to pyrimidine position ... pyrimidine position 6.

If it is assumed that th... structure in the most pl... (that is, with the keto r... figurations) it is found th... bases can bond together. ... (purine) with thymine (p... (purine) with cytosine (pyr...

In other words, if an ade... a pair, on either chain, th... the other member must b... guanine and cytosine. Th... single chain does not appe... way. However, if only spe... formed, it follows that if ... one chain is given, then t... chain is automatically dete...

It has been found experi... of the amounts of adenine ... of guanine to cytosine, are ... for deoxyribose nucleic aci...

It is probably impossibl... with a ribose sugar in pla... the extra oxygen atom wo... der Waals contact.

The previously published ... ribose nucleic acid are insu... of our structure. So far as ... compatible with the experi... be regarded as unproved ... against more exact results. ... in the following communica... of the details of the results ... devised our structure, whic... entirely on published exper... chemical arguments.

It has not escaped our ... pairing we have postulated ... possible copying mechanism.

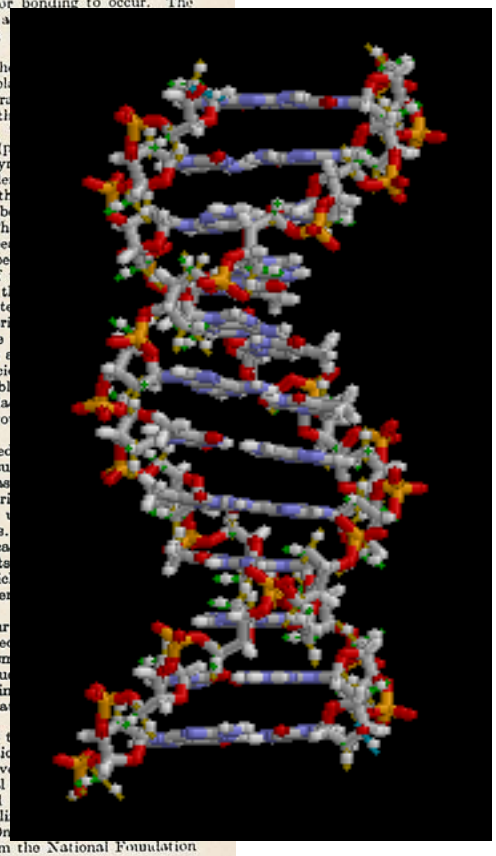
Full details of the stru... ditions assumed in buildi... of co-ordinates for the a... elsewhere.

We are much indebted t... constant advice and criti... atomic distances. We hav... a knowledge of the general ... experimental results and ... Wilkins, Dr. R. E. Frankli... King's College, London. Or... aided by a fellowship from the National Foundation for Infantile Paralysis.

J. D. WATSON
F. H. C. CRICK

Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems,
Cavendish Laboratory, Cambridge.
April 2.

[1] Pauling, L., and Corey, R. B., *Nature*, 171, 346 (1953); *Proc. U.S. Nat. Acad. Sci.*, 39, 84 (1953).
[2] Furberg, S., *Acta Chem. Scand.*, 6, 634 (1952).
[3] Chargaff, E., for references see Zamenhof, S., Brawerman, G., and Chargaff, E., *Biochim. et Biophys. Acta*, 9, 402 (1952).
[4] Wyatt, G. R., *J. Gen. Physiol.*, 36, 201 (1952).
[5] Astbury, W. T., Symp. Soc. Exp. Biol. 1, Nucleic Acid, 66 (Camb. Univ. Press, 1947).
[6] Wilkins, M. H. F., and Randall, J. T., *Biochim. et Biophys. Acta*, 10, 192 (1953).



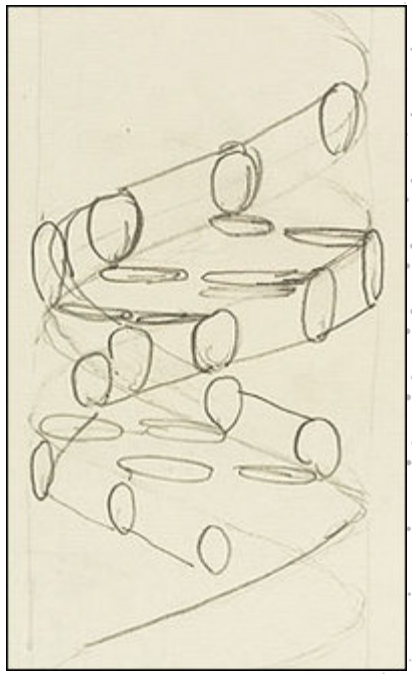
**first discovered in 1869**
**genetic inheritance demonstrated in 1943**
**Structure determination in 1953**

**Input:**

**publication length = 1 page**

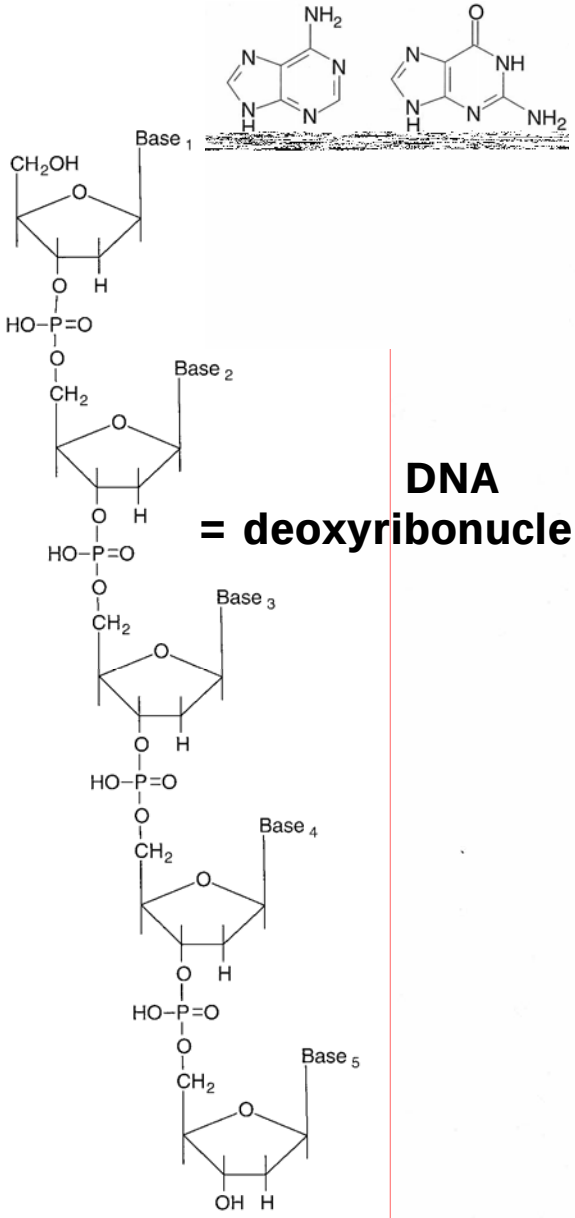
**Output:**

**Nobel Prize Medicine 1962**

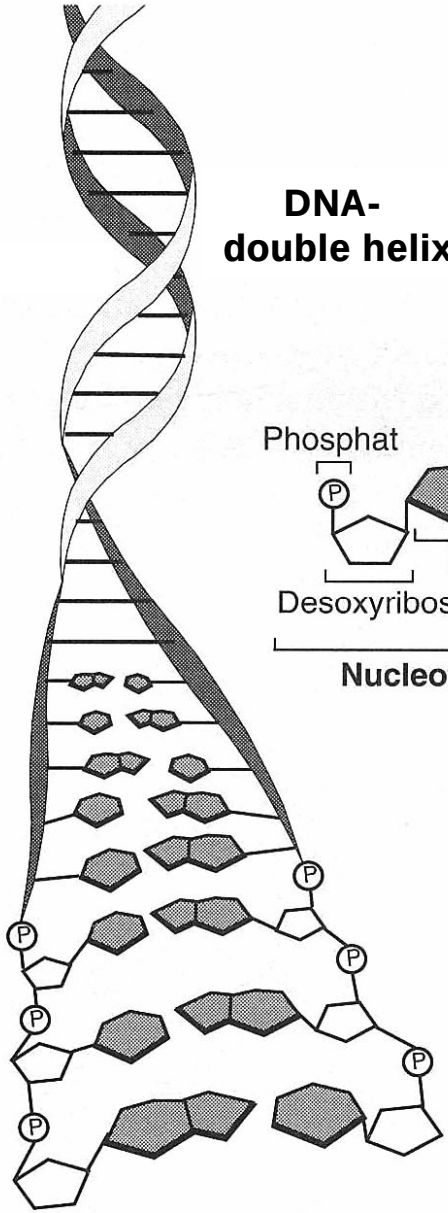


Francis Crick's first sketch of the deoxyribonucleic acid double-helix pattern

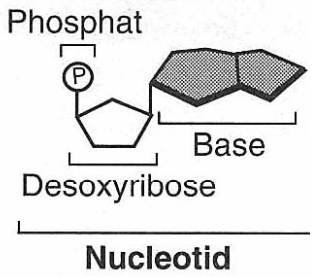
# Basics: DNA, RNA
# why do we have two different ones ?



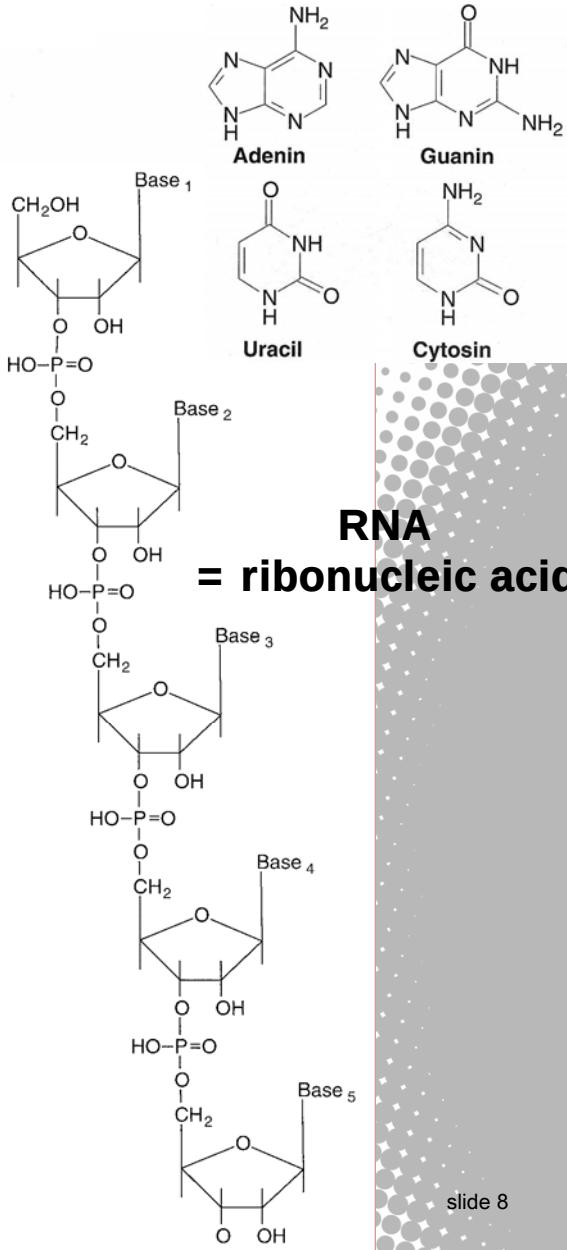

**DNA**
**= deoxyribonucleic acid**



**DNA-**
**double helix**




**Nucleotid**

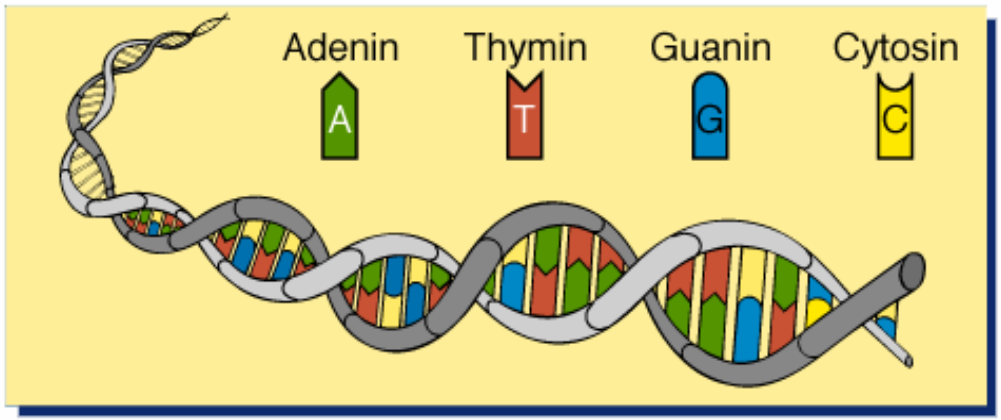



**RNA**
**= ribonucleic acid**

# Basics: The language of the genetic code
# How many combinations are possible ?

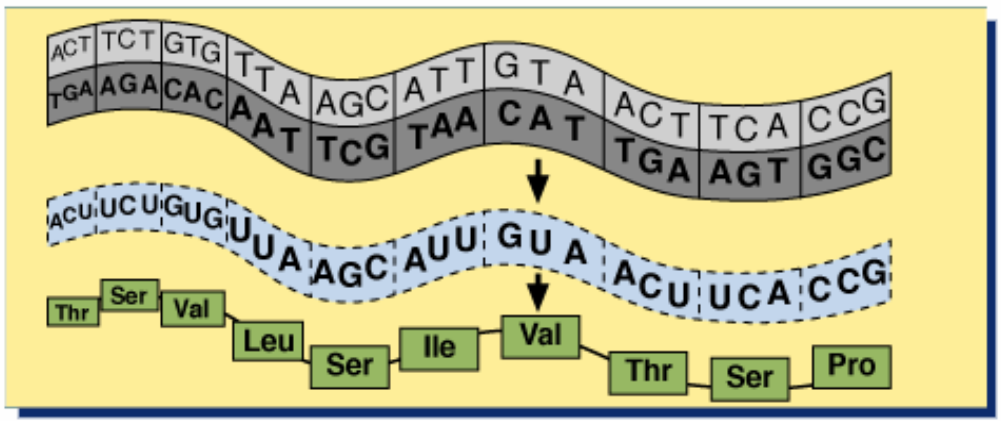
| RNA | | 2nd base | | | |
|---|---|---|---|---|---|
| | | **U** (uracile) | **C** (cytosine) | **A** (adenine) | **G** (guanidine) |
| **1st base** | **U** | UUU (Phe/F)Phenylalanine<br>UUC (Phe/F)Phenyla*nine*<br>UUA (Leu/L)Leucine<br>UUG (Leu/L)Leucine, *Start* | UCU (Ser/S)Serine<br>UCC (Ser/S)Serine<br>UCA (Ser/S)Serine<br>UCG (Ser/S)Serine | UAU (Tyr/Y)Tyrosine<br>UAC (Tyr/Y)Tyrosine<br>UAA Ochre (*Stop*)<br>UAG Amber (*Stop*) | UGU (Cys/C)Cysteine<br>UGC (Cys/C)Cysteine<br>UGA Opal (*Stop*)<br>UGG (Trp/W)Tryptophan |
| | **C** | CUU (Leu/L)Leucine<br>CUC (Leu/L)Leucine<br>CUA (Leu/L)Leucine<br>CUG (Leu/L)Leucine, *Start* | CCU (Pro/P)Proline<br>CCC (Pro/P)Proline<br>CCA (Pro/P)Proline<br>CCG (Pro/P)Proline | CAU (His/H)Histidine<br>CAC (His/H)Histidine<br>CAA (Gln/Q)Glutamine<br>CAG (Gln/Q)Glutamine | CGU (Arg/R)Arginine<br>CGC (Arg/R)Arginine<br>CGA (Arg/R)Arginine<br>CGG (Arg/R)Arginine |
| | **A** | AUU (Ile/I)Isoleucine, *Start*<br>AUC (Ile/I)Isoleucine<br>AUA (Ile/I)Isoleucine<br>AUG (Met/M)Methionine, *Start* | ACU (Thr/T)Threonine<br>ACC (Thr/T)Threonine<br>ACA (Thr/T)Threonine<br>ACG (Thr/T)Threonine | AAU (Asn/N)Asparagine<br>AAC (Asn/N)Asparagine<br>AAA (Lys/K)Lysine<br>AAG (Lys/K)Lysine | AGU (Ser/S)Serine<br>AGC (Ser/S)Serine<br>AGA (Arg/R)Arginine<br>AGG (Arg/R)Arginine |
| | **G** | GUU (Val/V)Valine<br>GUC (Val/V)Valine<br>GUA (Val/V)Valine<br>GUG (Val/V)Valine, *Start* | GCU (Ala/A)Alanine<br>GCC (Ala/A)Alanine<br>GCA (Ala/A)Alanine<br>GCG (Ala/A)Alanine | GAU (Asp/D)Aspartic acid<br>GAC (Asp/D)Aspartic acid<br>GAA (Glu/E)Glutamic acid<br>GAG (Glu/E)Glutamic acid | GGU (Gly/G)Glycine<br>GGC (Gly/G)Glycine<br>GGA (Gly/G)Glycine<br>GGG (Gly/G)Glycine |

# The Genetic Code - Alphabet of Life




Nucleotides

codogenic strand
**DNA-Helix**
non-codogenic strand

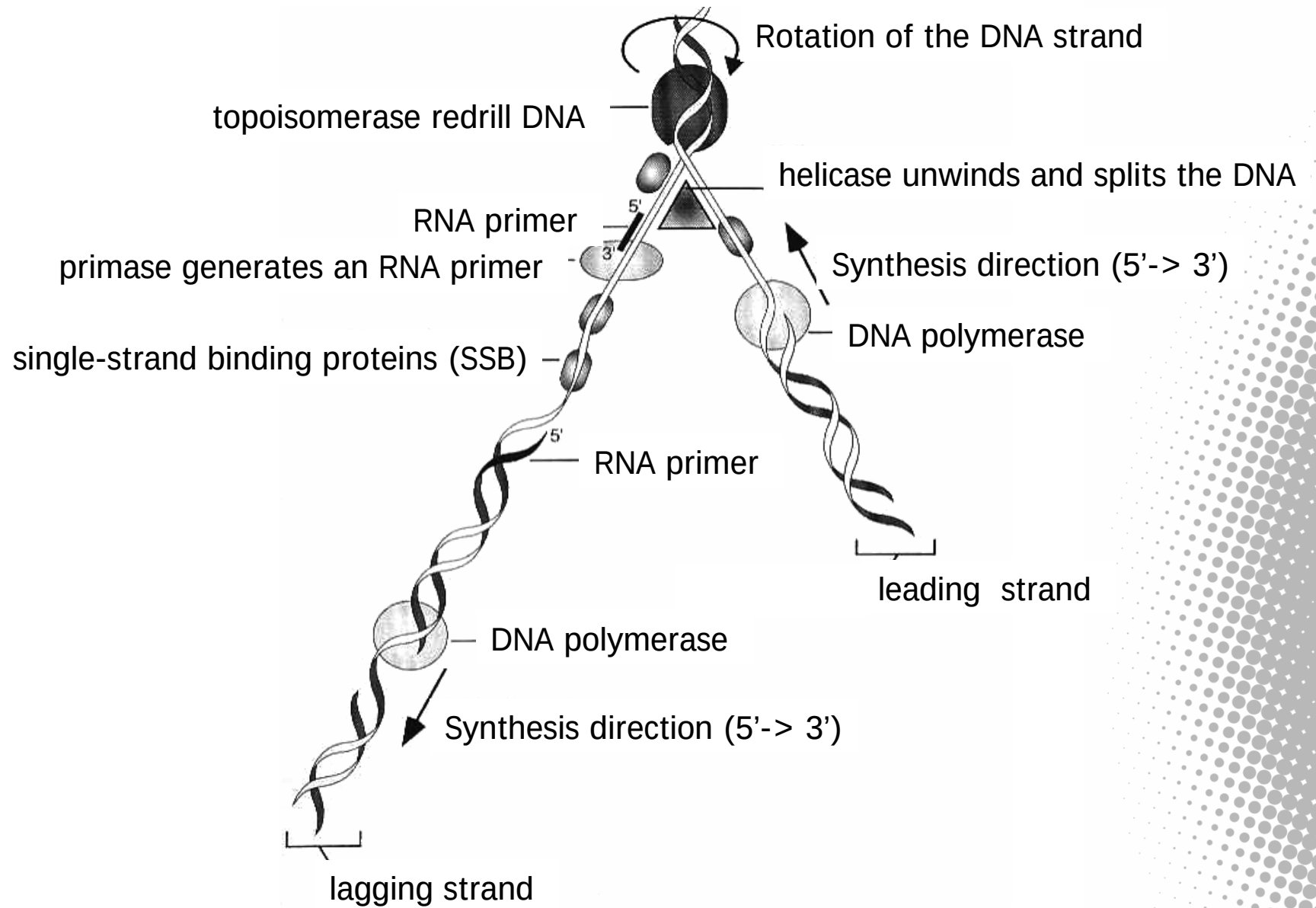



codogenic strand
**DNA**
non-codogenic strand

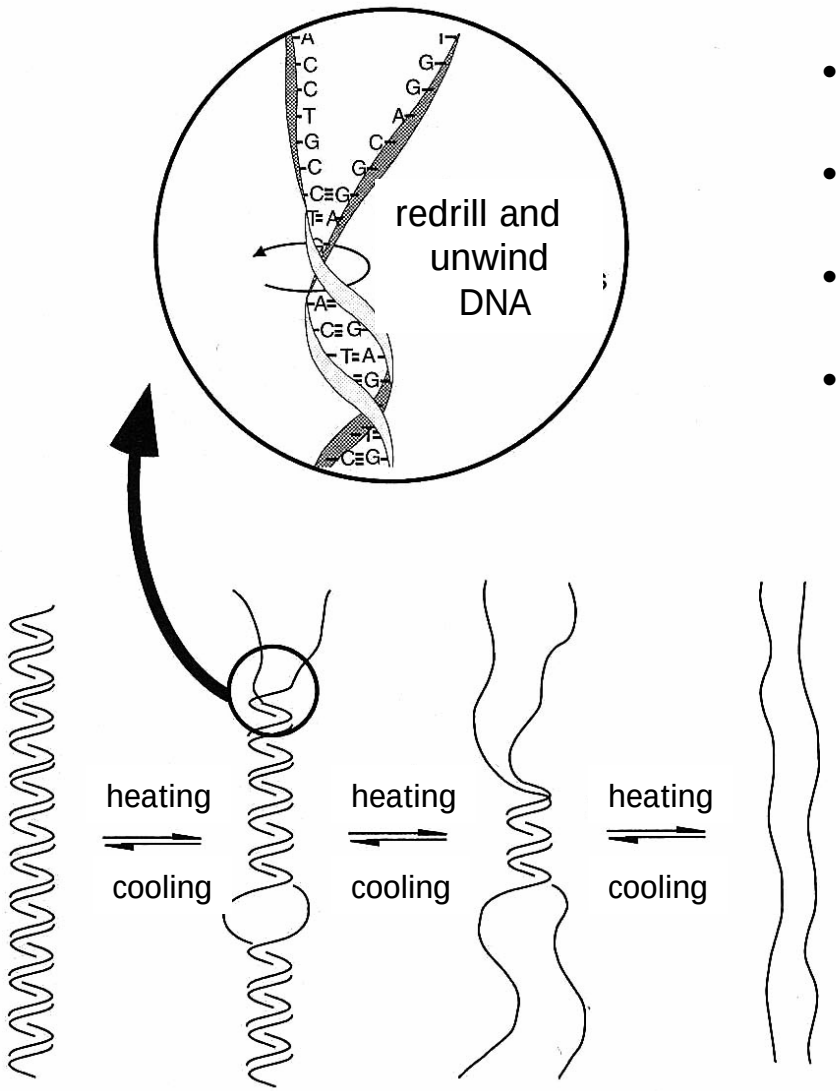
mRNA

Protein

# Basics: DNA replication

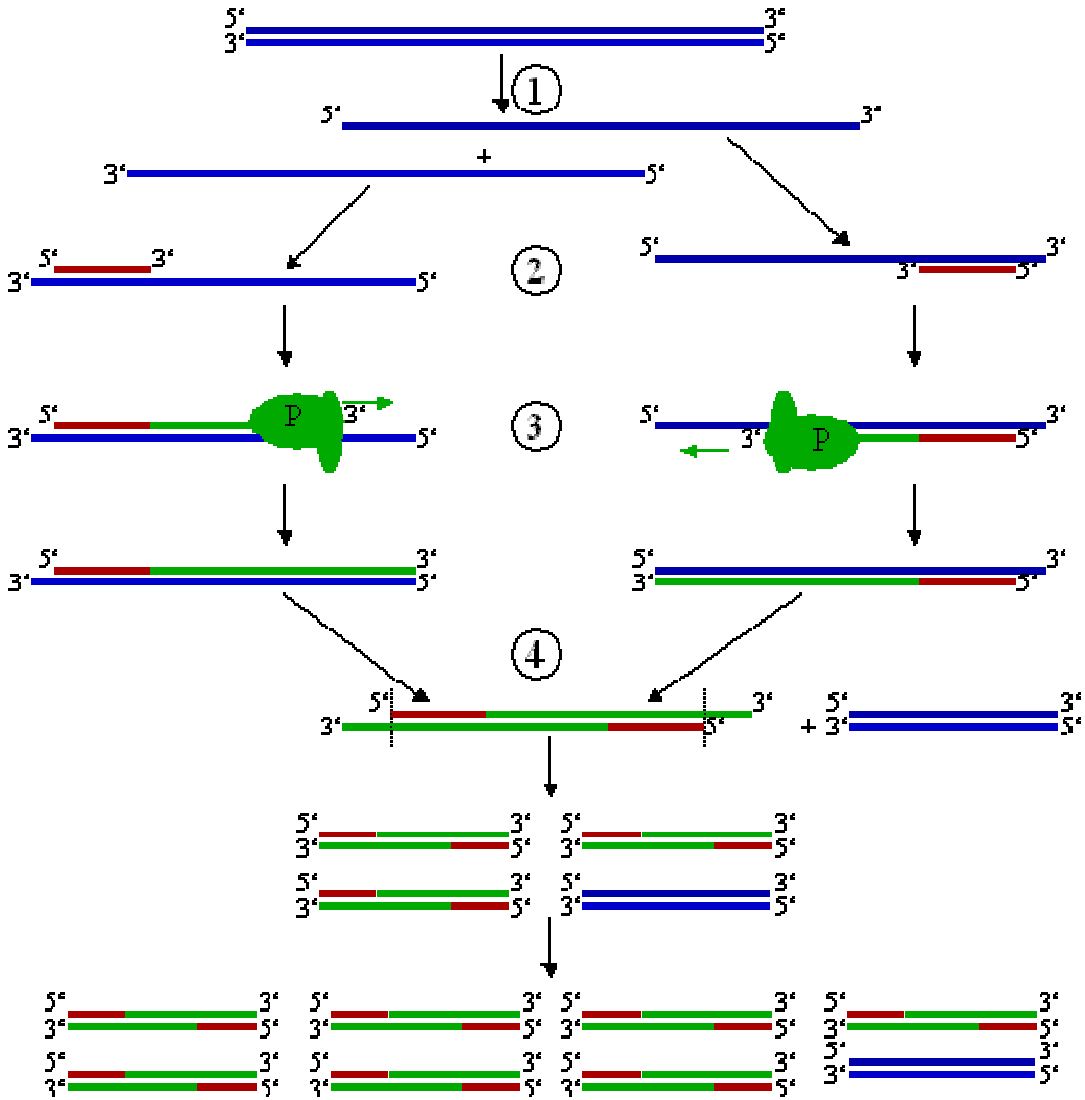


# Basics: DNA replication
# polymerase chain reaction - PCR




- *DNA templates,* which contains the region of the DNA fragment to be amplified
- *Primers,* which determine the beginning and end of the region to be amplified
- DNA-Polymerase, which copies the region to be amplified
- *Nucleotides,* from which the DNA-Polymerase builds the new DNA
- *Buffer,* which provides a suitable chemical environment for the DNA-Polymerase

# Basics: DNA replication
# polymerase chain reaction - PCR

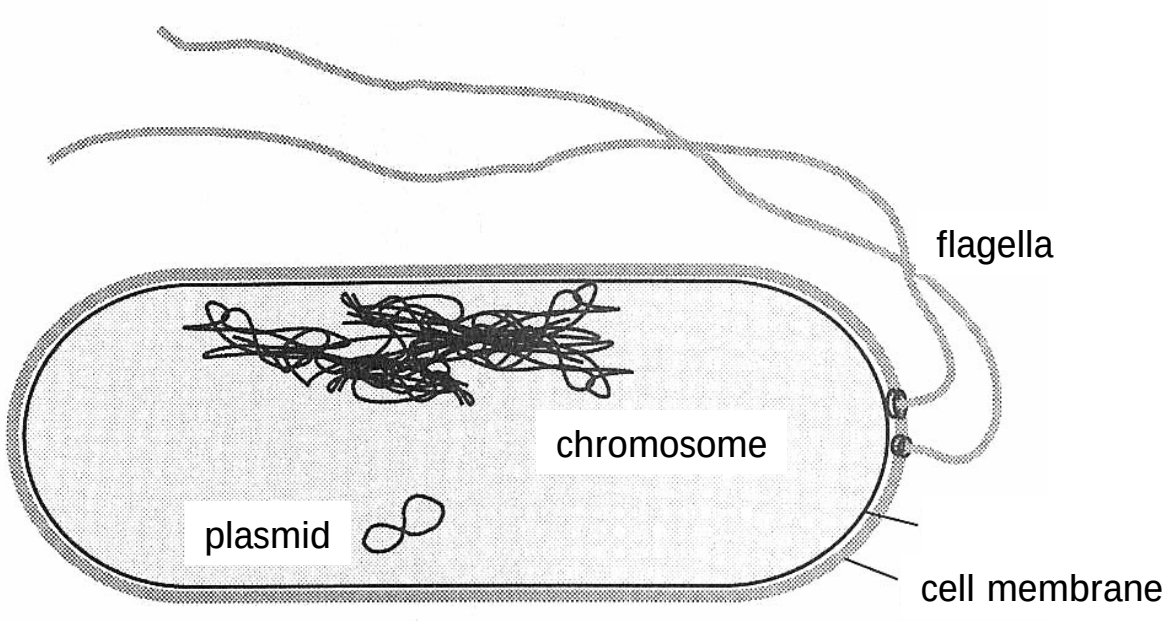


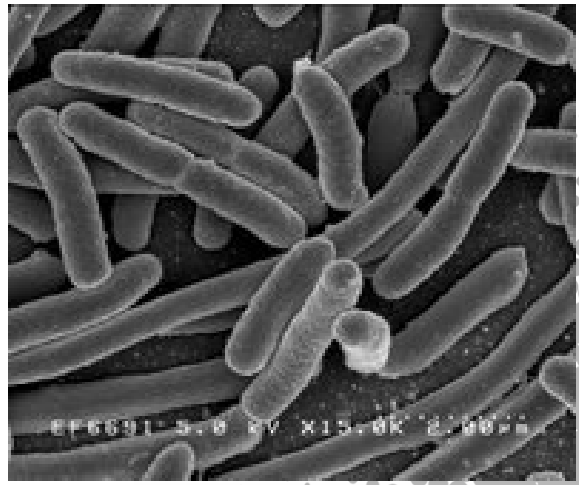

Schematic drawing of the PCR cycle.

(1) Denaturing at 96°C.
(2) Annealing at 68°C.
(3) Elongation at 72°C (P=Polymerase).
(4) The first cycle is complete.

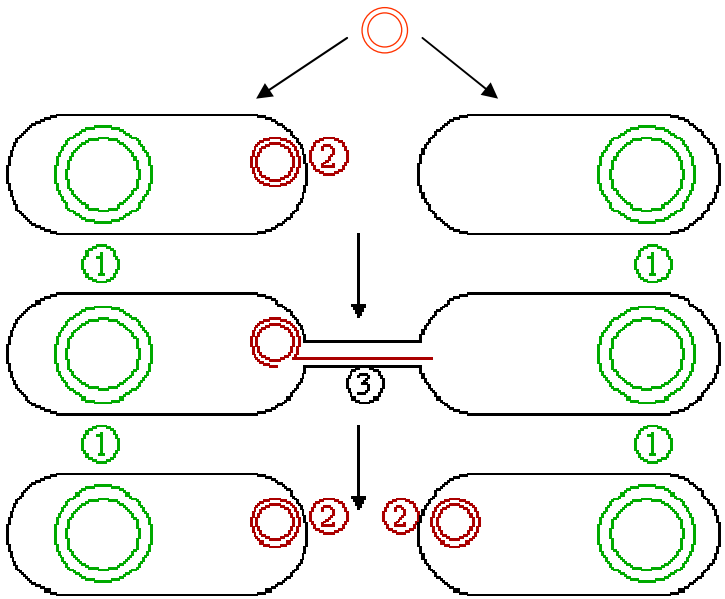
The two resulting DNA strands make up the template DNA for the next cycle, thus doubling the amount of DNA duplicated for each new cycle.

# Basics: cell structure of bacteria



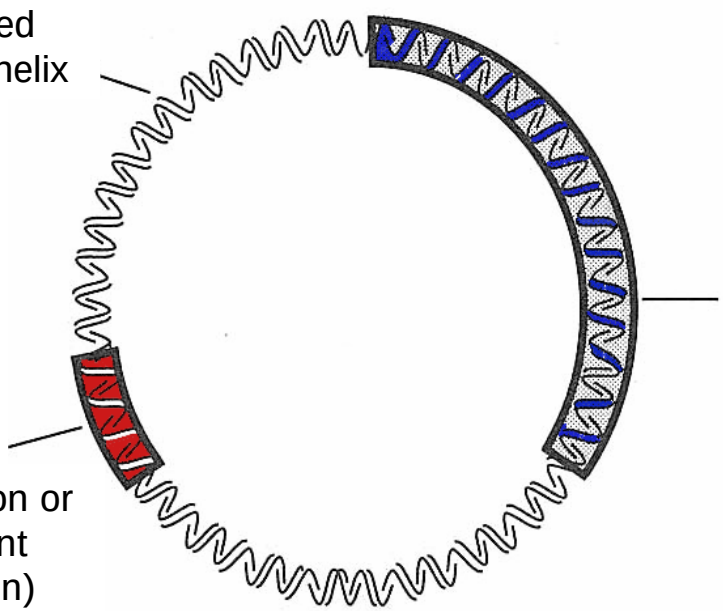



Escherichia coli



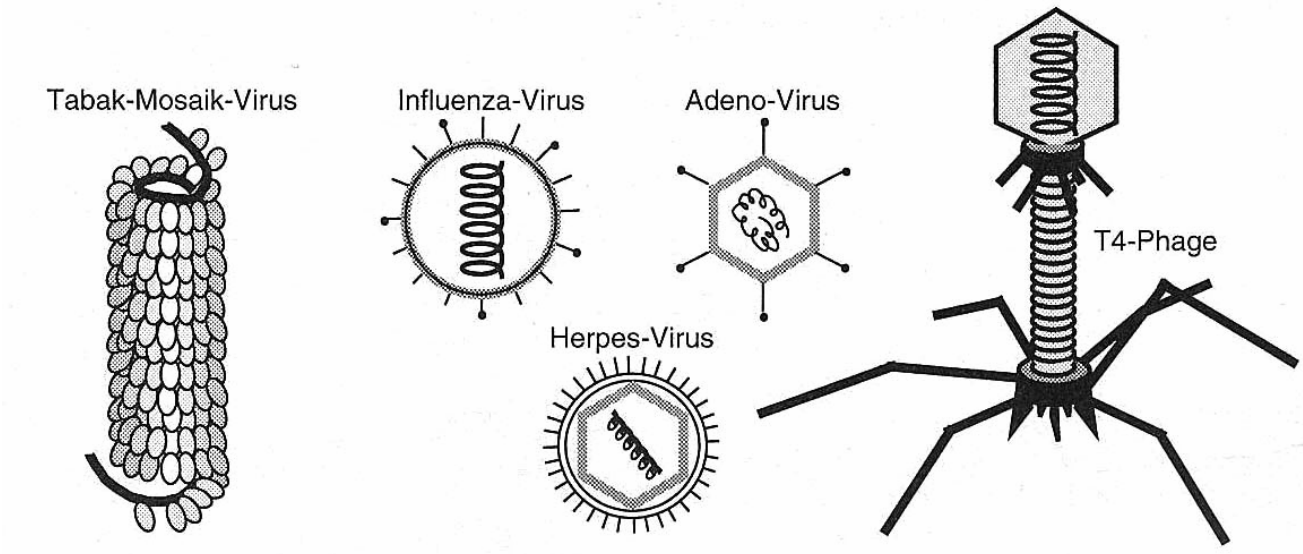

Schematic drawing of bacterial conjugation.
1 Chromosomal DNA.
2 Plasmids.
3 Pilus.

# Basics: schematic drawing of a plasmid with antibiotic resistances

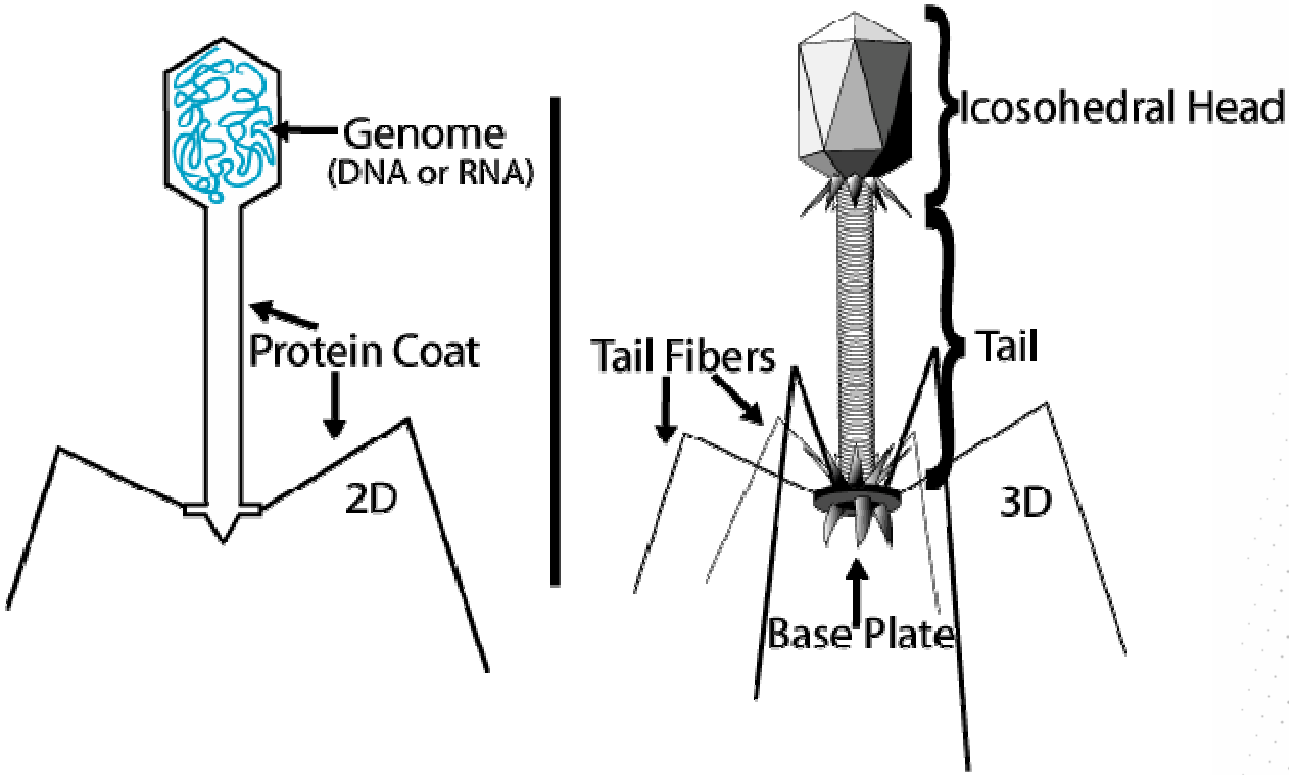



isolated/discovered in 1950
discovery: plasmids are responsible for the resistant for antibiotics

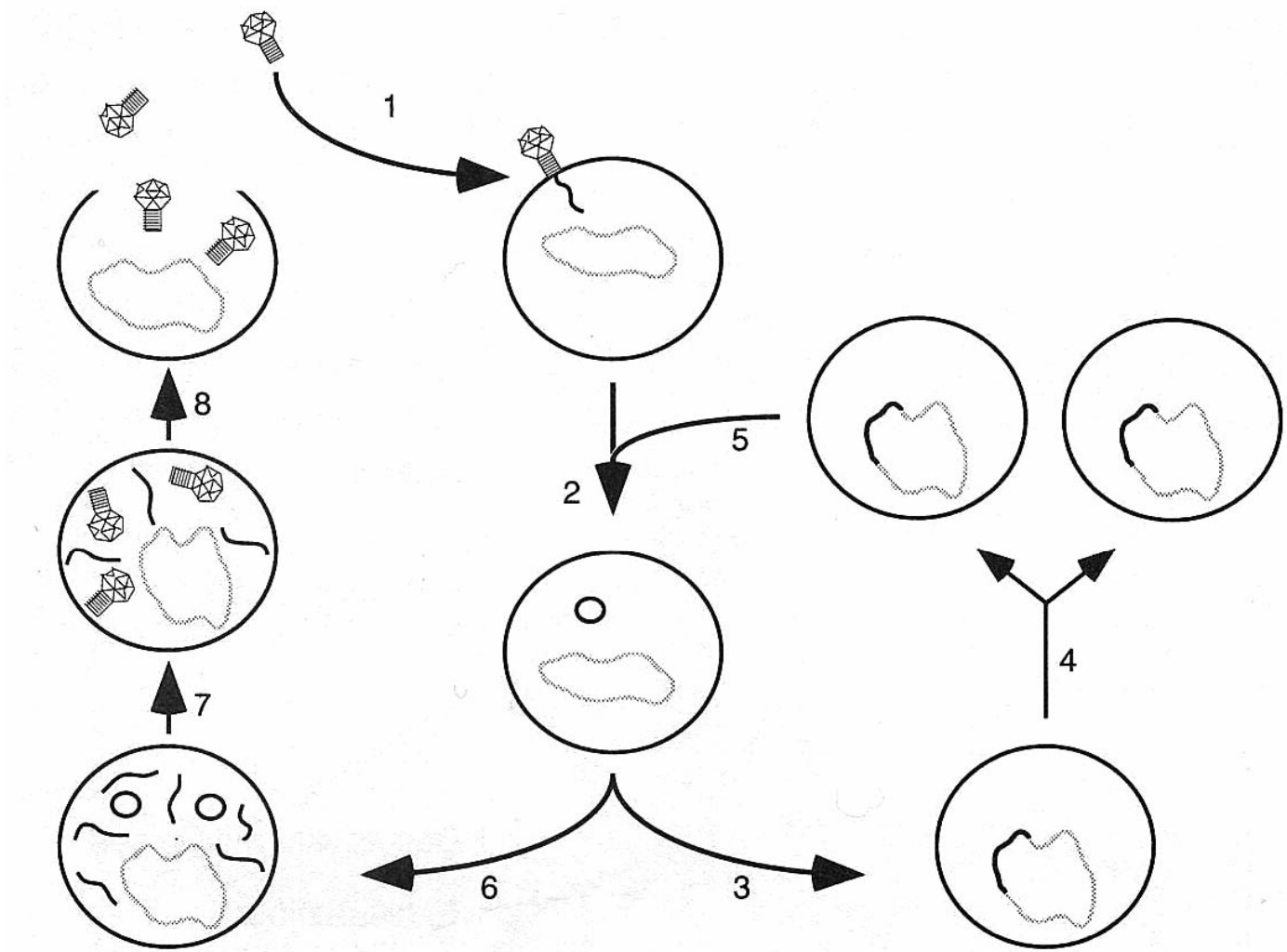
# Virus: schematic drawing of different viruses



# Structural overview of bacteriophage



# Reproduction of bacteriophages

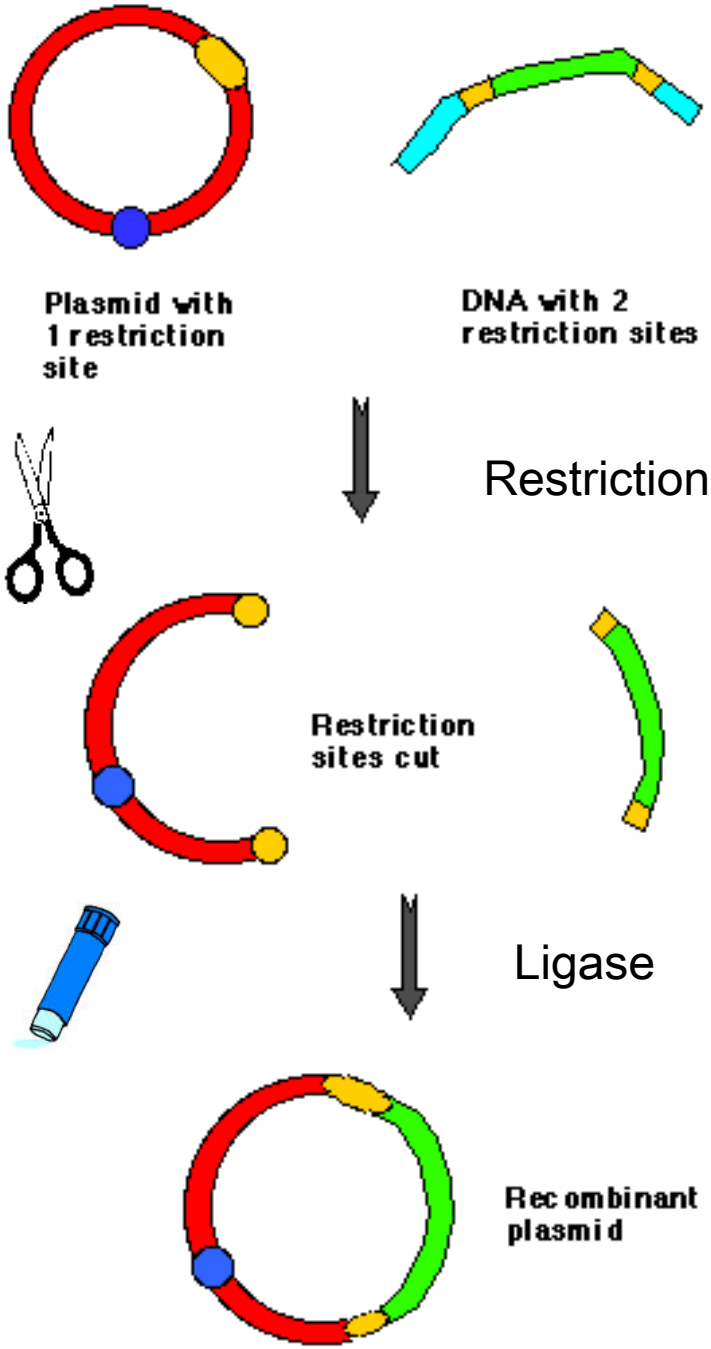



lytic cycle
(cell destroying cycle)

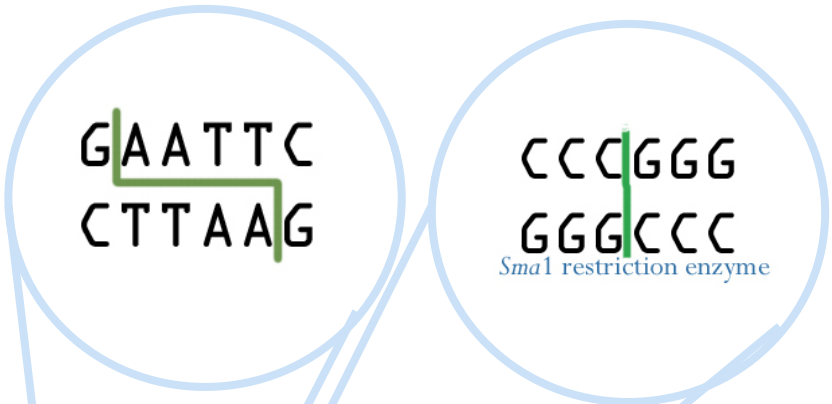
lysogenic cycle
(prophage cycle)

# Ligation

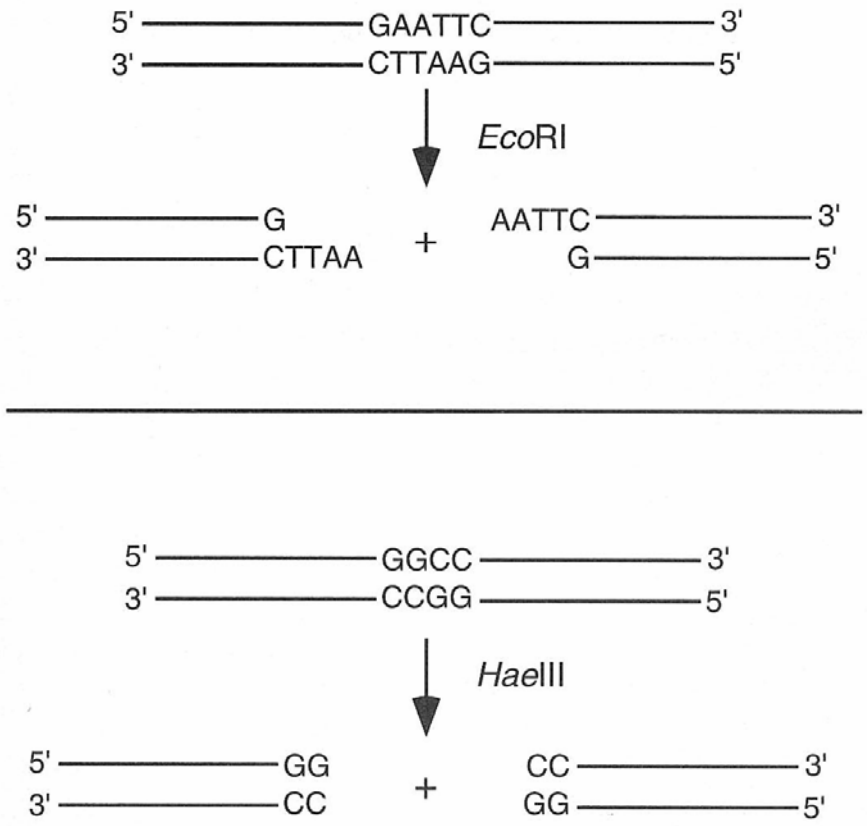
Vector



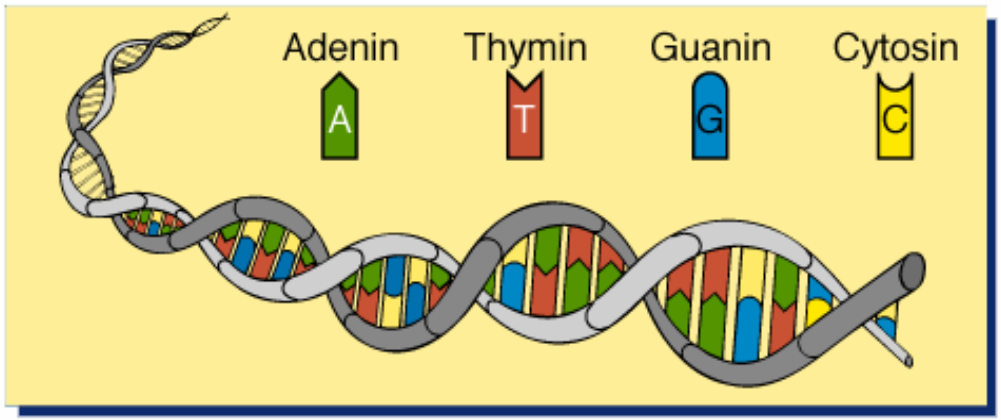
# Examples for Restriction Endonucleases




| Enzyme | recognition sequence | Enzyme | recognition sequence |
|---|---|---|---|
| *Ava*I | G / ACGCC | *Kpn*I | GGTAC / C |
| *Alu*I | AG / CT | *Nco*I | C / CATGG |
| *Apa*I | GG GCC / C | *Pvu*I | CGAT / CG |
| *Bal*I | TGG / CCA | *Pvu*II | CAG / CTG |
| *Bam*HI | G / GATCC | *Pst*I | CTGCA / G |
| *Bgl*I | C CCNNNN / NGGC | *Sma*I | CCC / GGG |
| *Bgl*II | A / GATCT | *Sau*3A | / GATC |
| *Bst*EII | G / GTNACC | *Sal*I | G / TCGAC |
| *Cla*I | AT / CGAT | *Sph*I | GCATG / C |
| *Eco*RI | G / AATTC | *Sst*I | GAGCT / C |
| *Eco*RII | / CCA GG | *Sst*II | CCGC / GG |
| *Hae*III | GG / CC | *Taq*I | T / CGA |
| *Hind*III | A / AGCTT | *Xba*I | T / CTAGA |
| *Hpa*I | GTT / AAC | *Xho*I | C / TCGAG |
| *Hpa*II | C / CGG | *Xma*I | C / CCGGG |

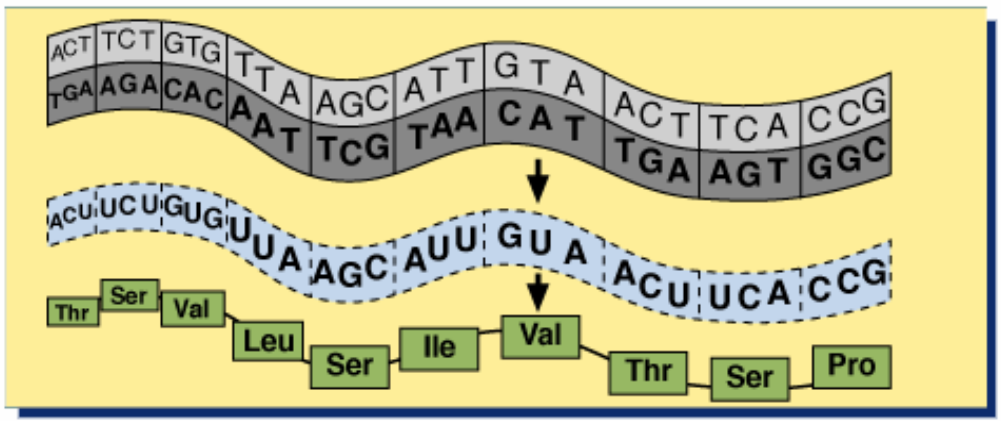


# The Genetic Code - Alphabet of Life




Nucleotides

codogenic strand
**DNA-Helix**
non-codogenic strand




codogenic strand
**DNA**
non-codogenic strand

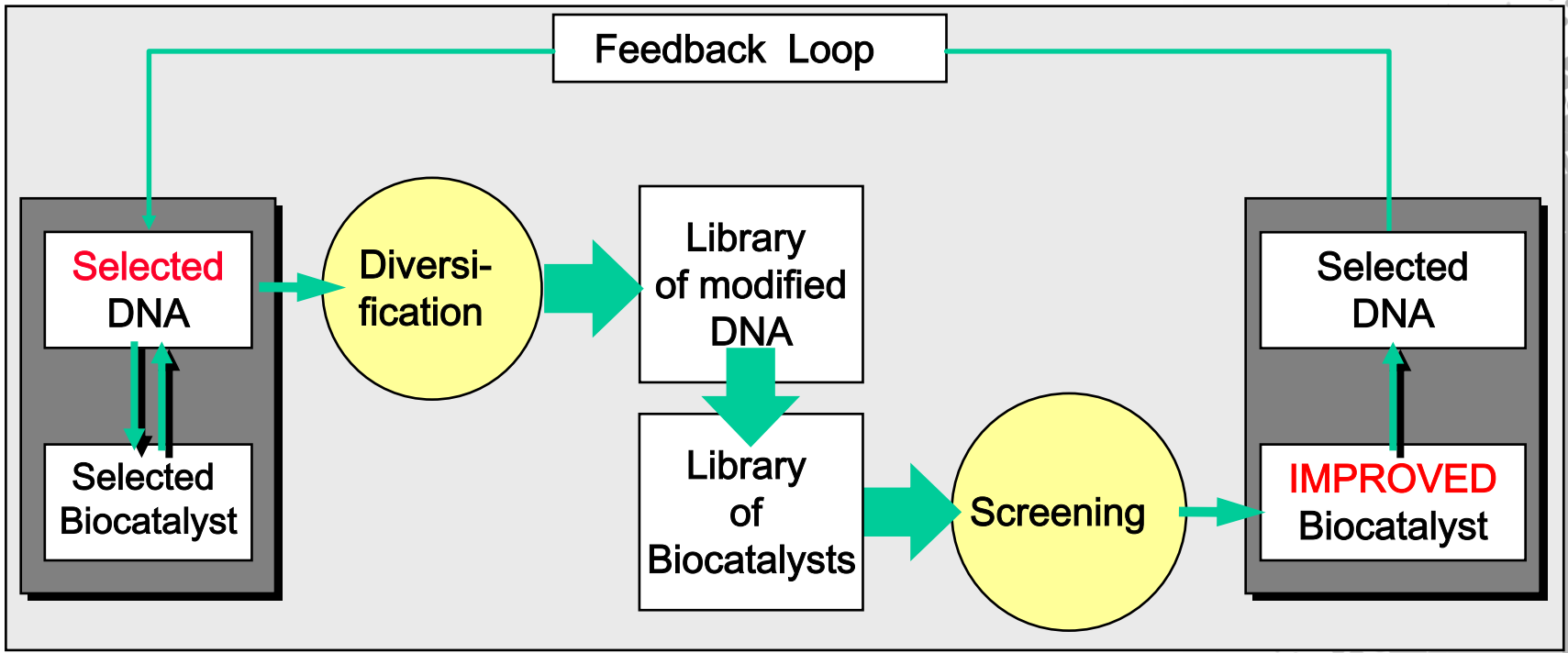
mRNA

Protein

Transcription (polymerase)

Translation (ribosome)

# Biocatalysis: Automated Screening

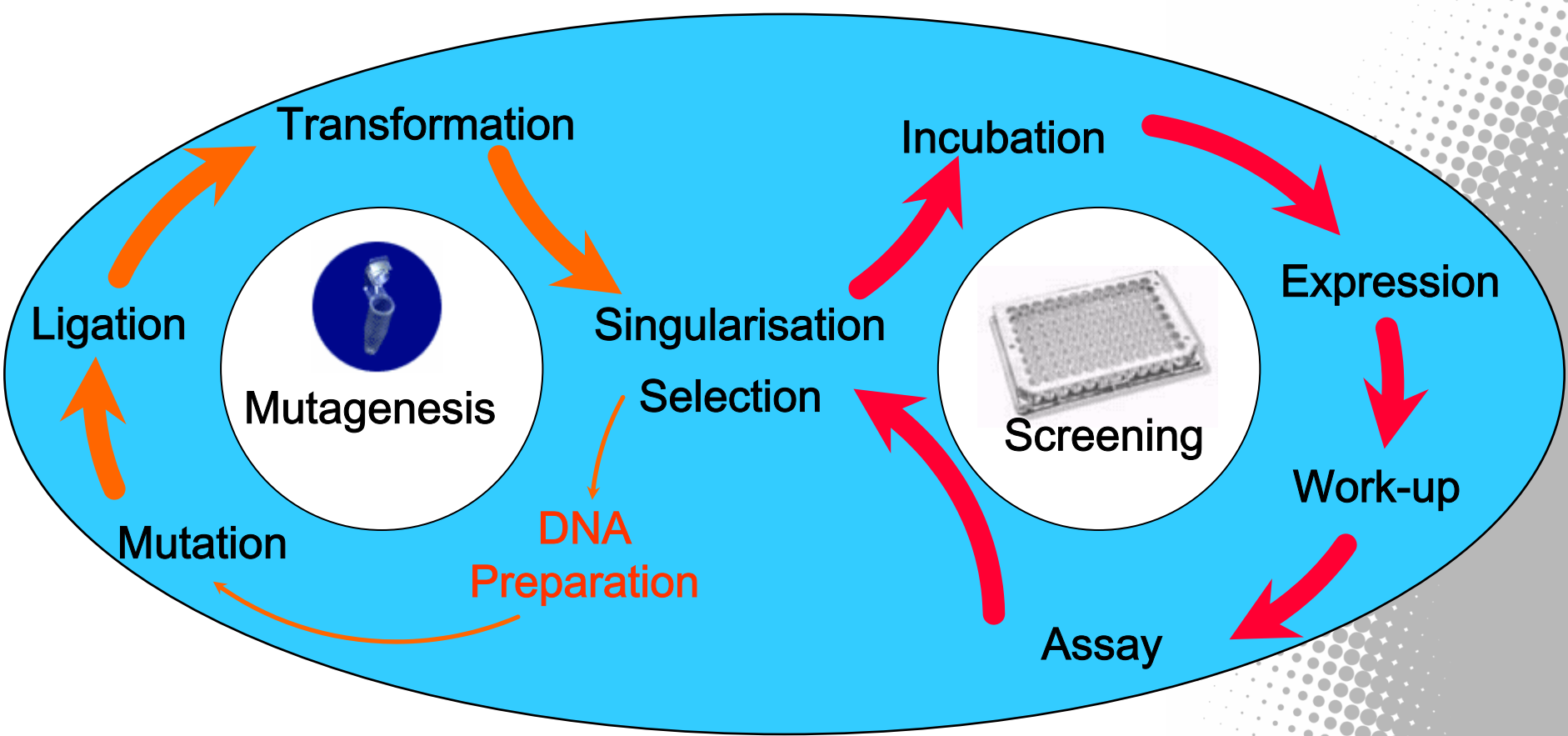



Gene & Biocatalyst – Library

Mutagenesis

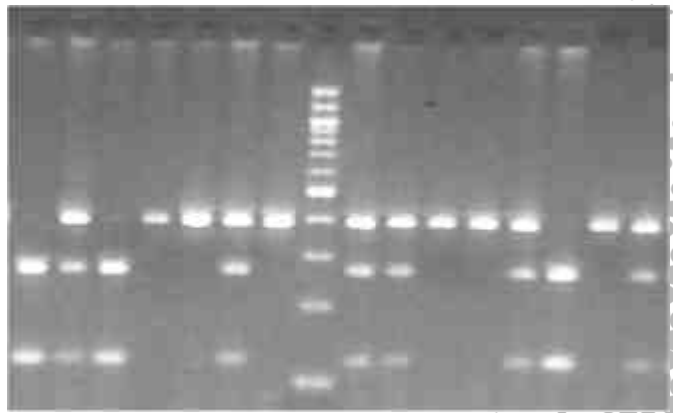
Automation, Assay Development , Screening

# Evolutionary Methods: How it works in practice



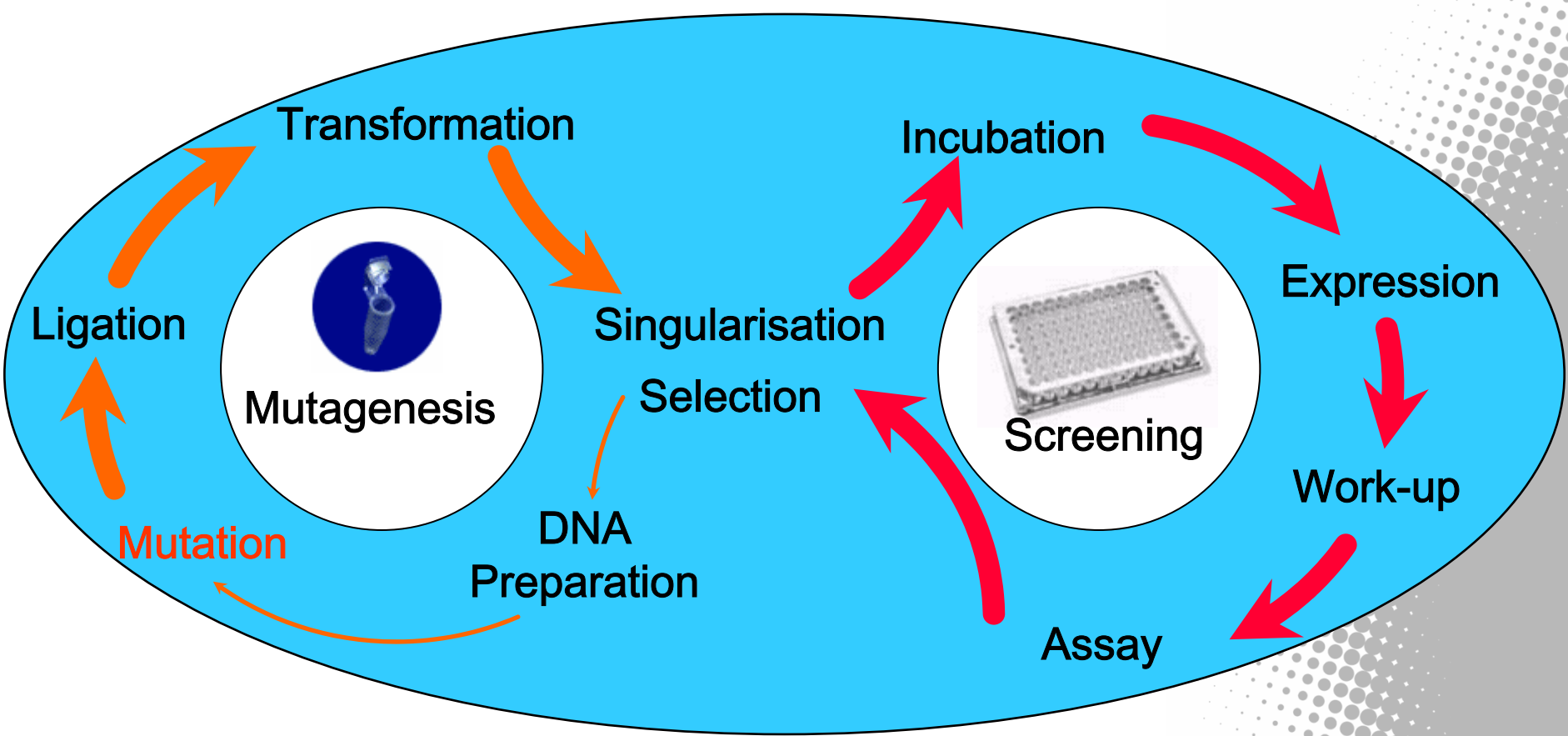
# DNA Preparation

DNA Strand

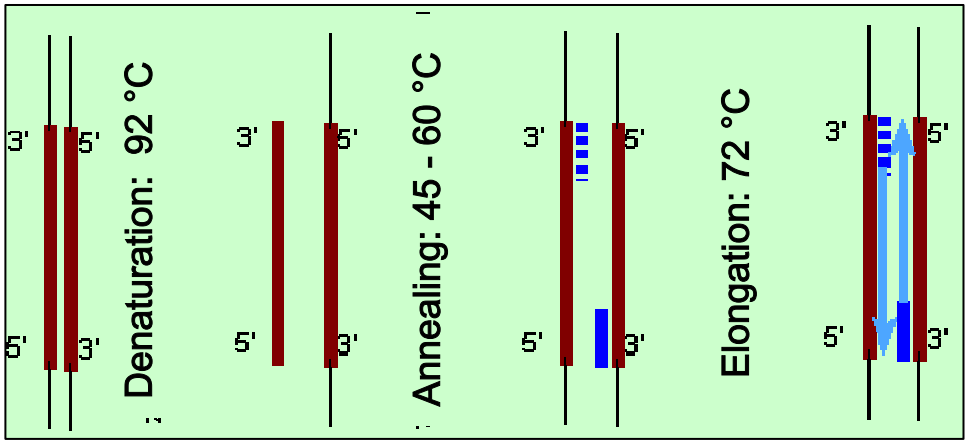


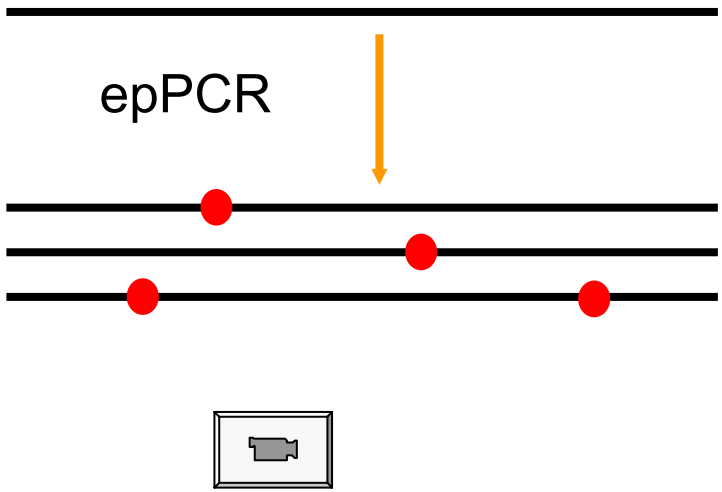
Analysis by
Gel Electrophoresis

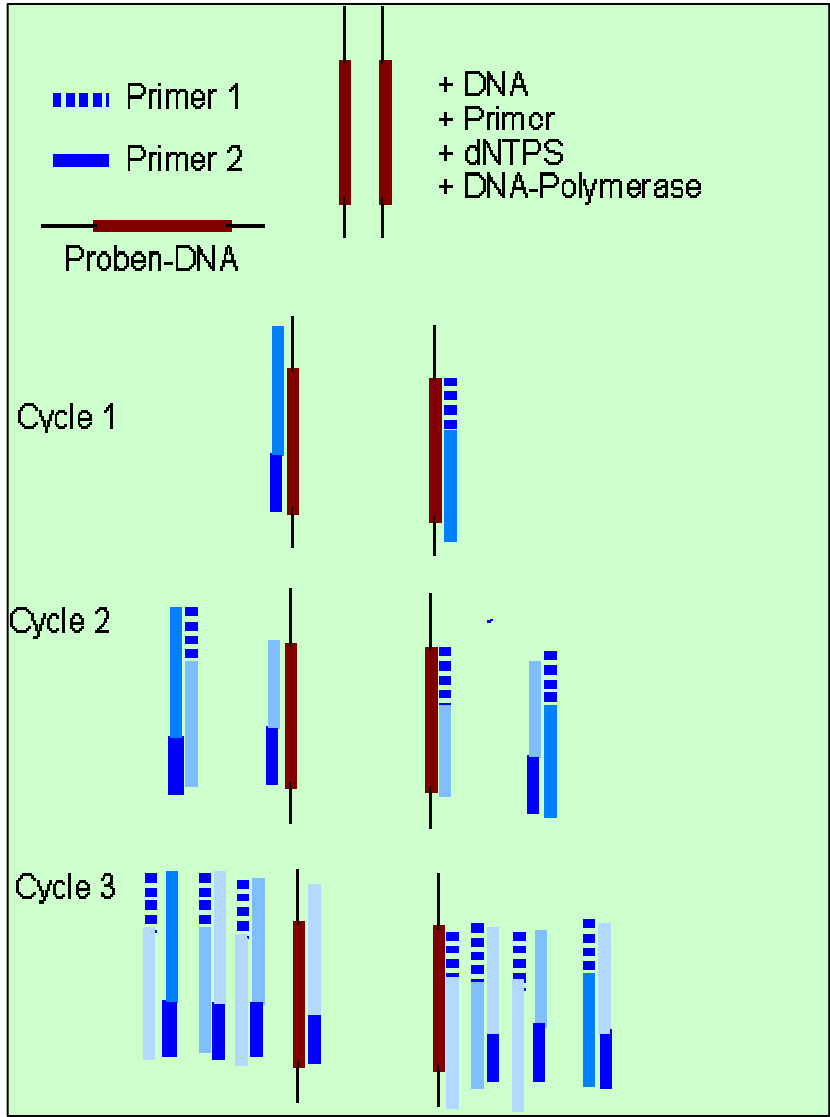
# Evolutionary Methods: Practical Performance



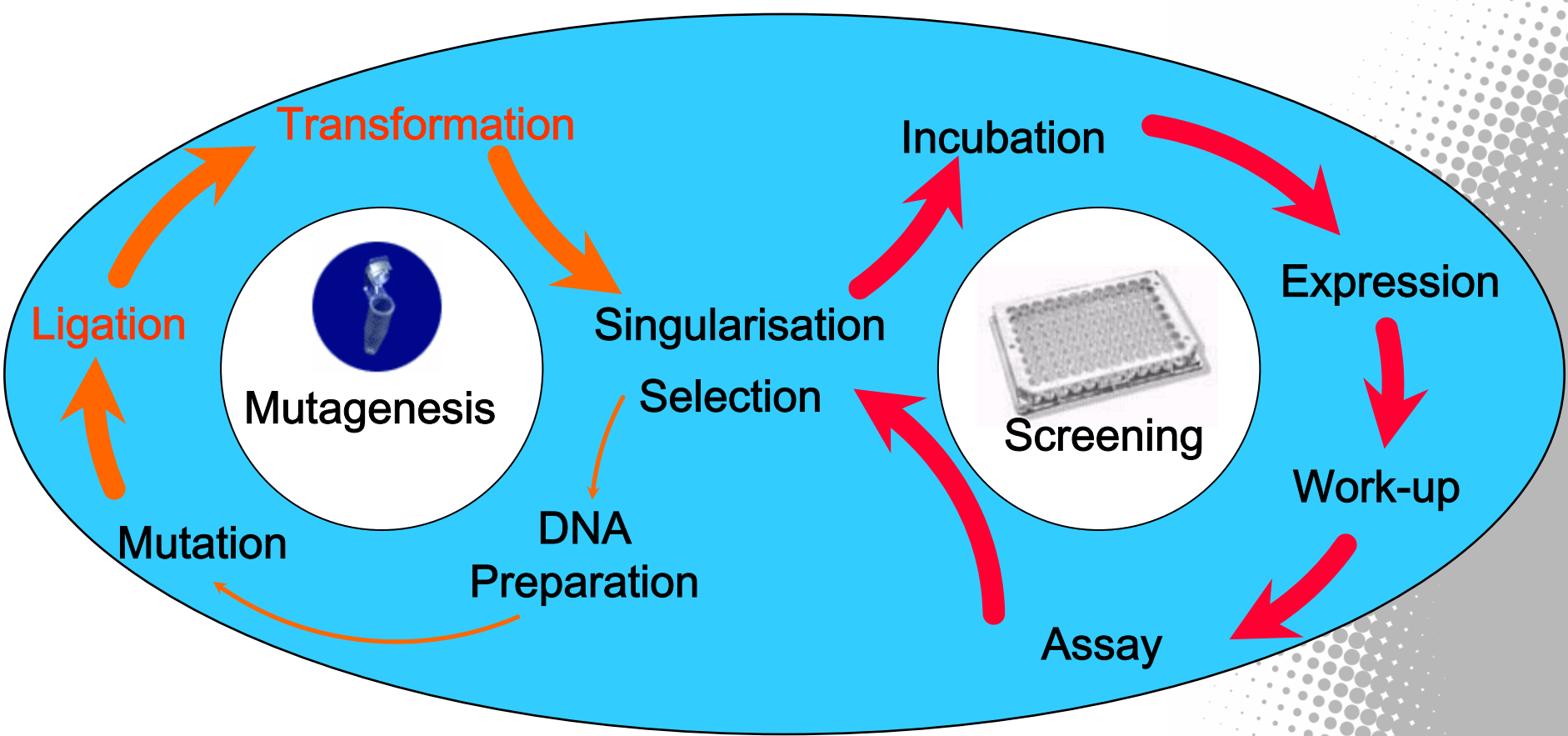
# Mutation by error-prone PCR (polymerase chain reaction )



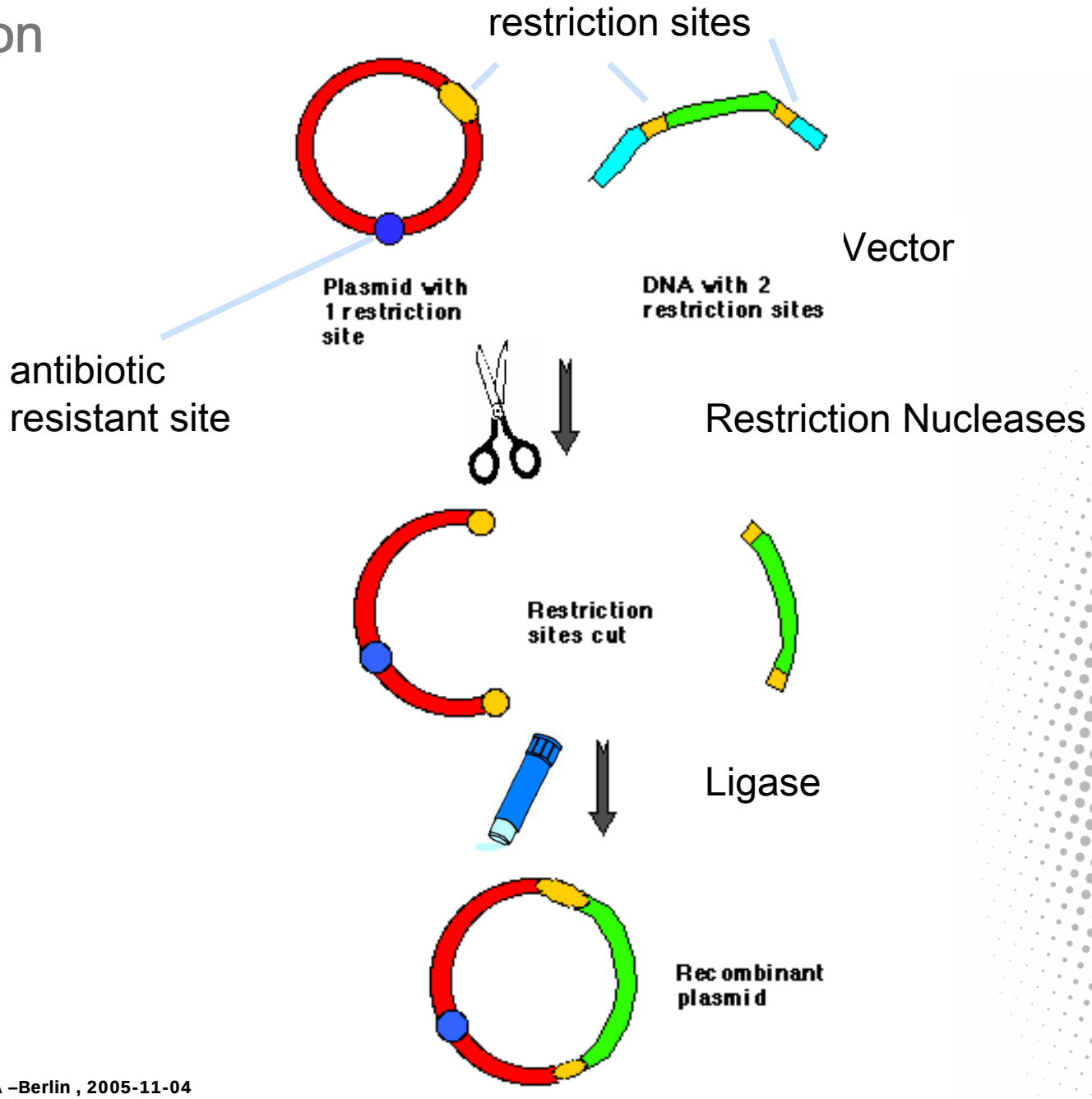



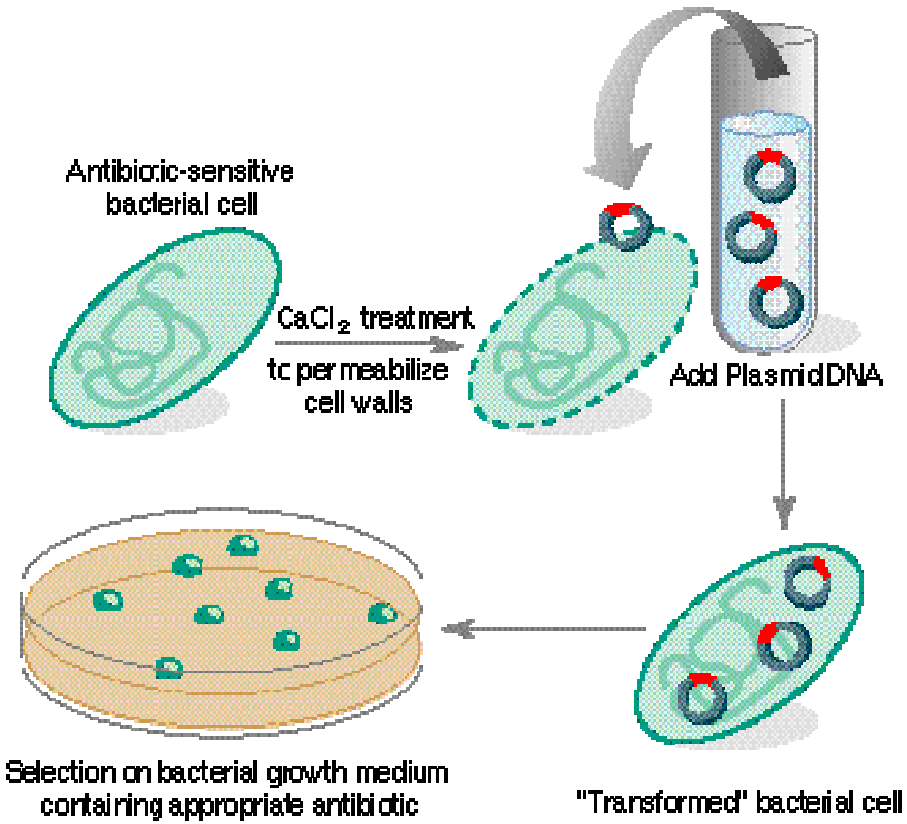
# Evolutionary Methods: Practical Performance



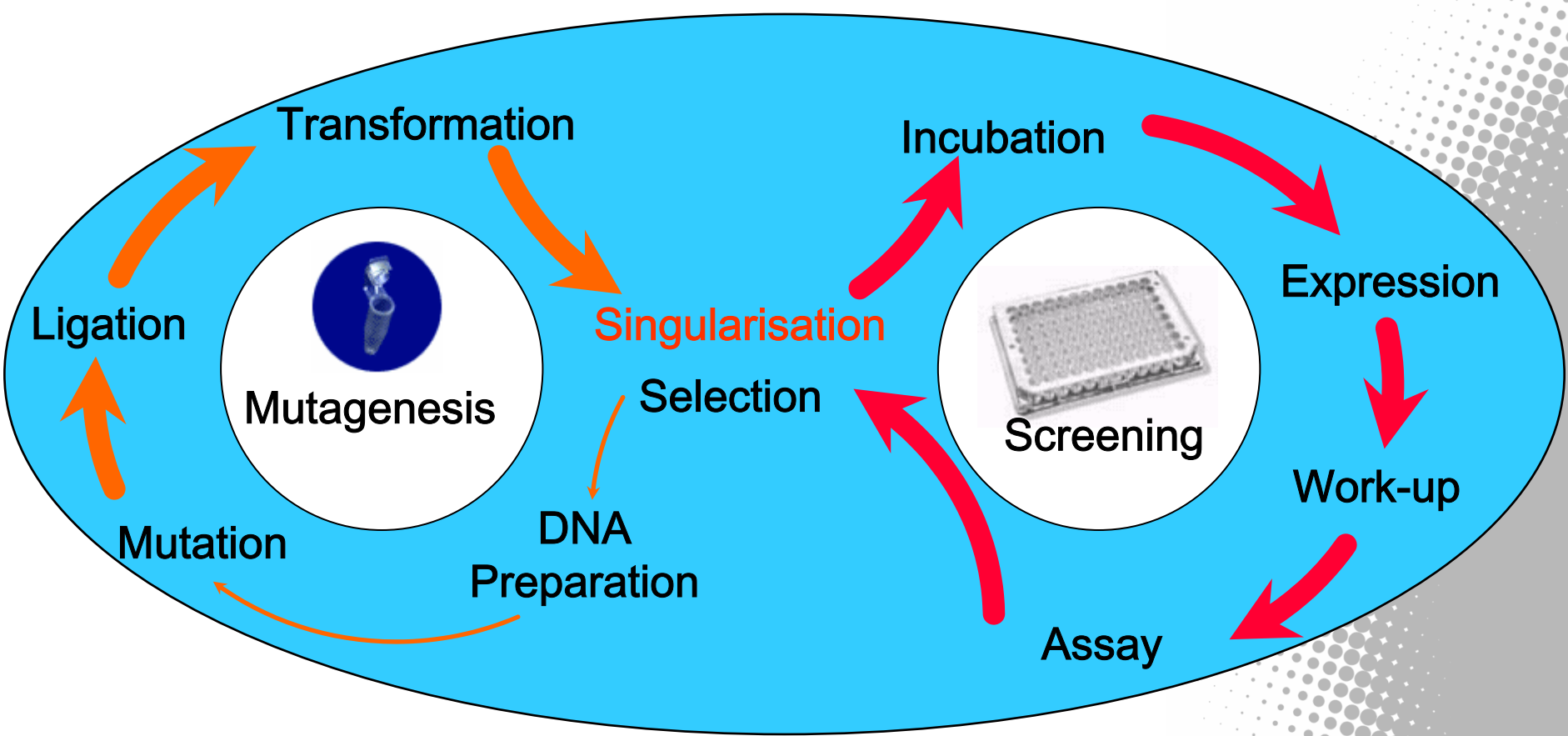
# Ligation


restriction sites
Vector
Plasmid with 1 restriction site
DNA with 2 restriction sites
antibiotic resistant site
Restriction Nucleases
Restriction sites cut
Ligase
Recombinant plasmid

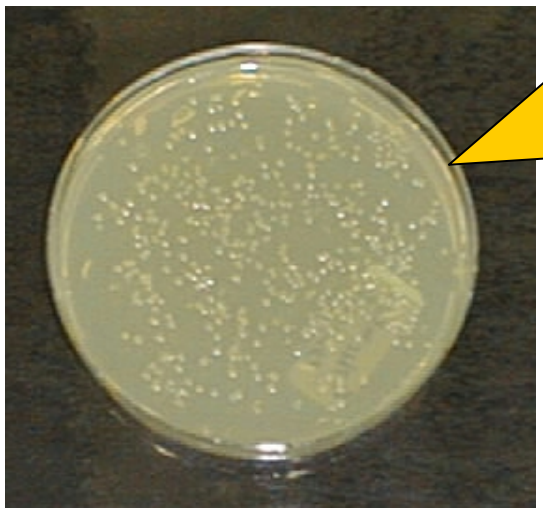
# Bacterial Transformation



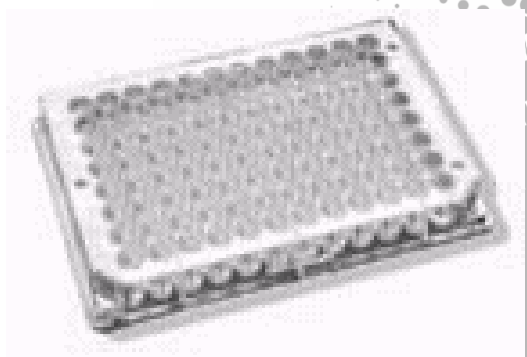
# Evolutionary Methods: Practical Performance



# Singularization



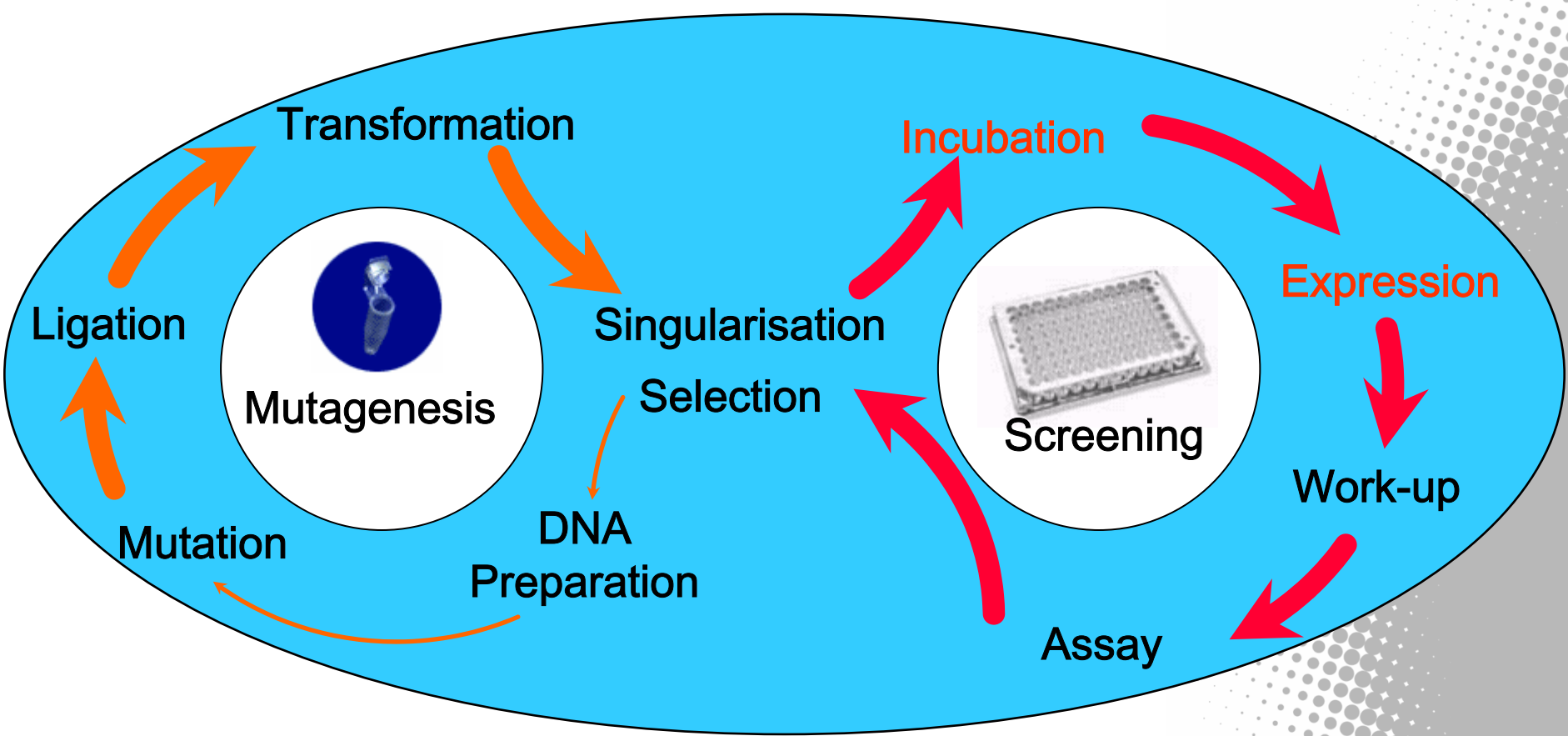
Each of the colonies represents a new clone!



- Statistical distribution
- Localization only by relativ position
- Solid medium

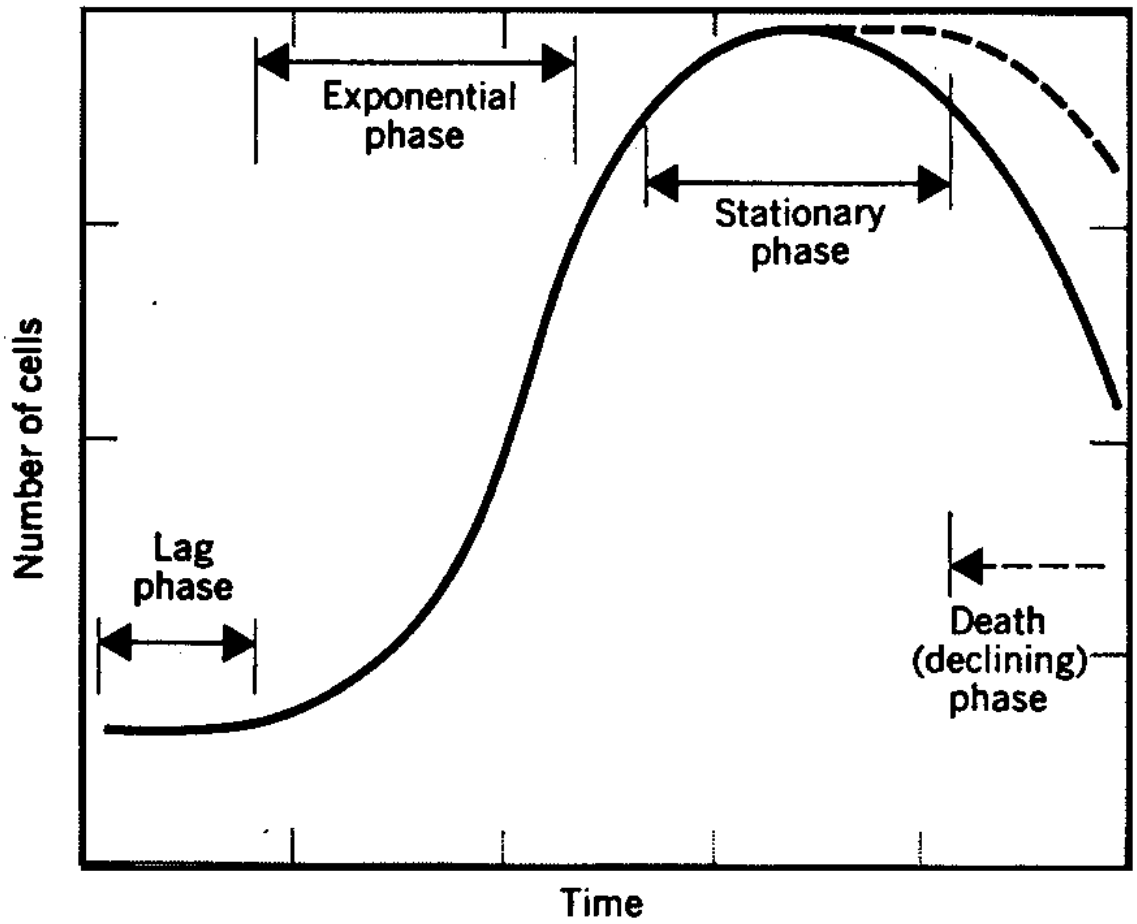
- Well defined position
- Storable
- Solution

# Evolutionary Methods: Practical Performance

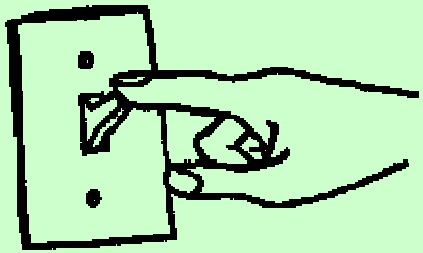


# Robot

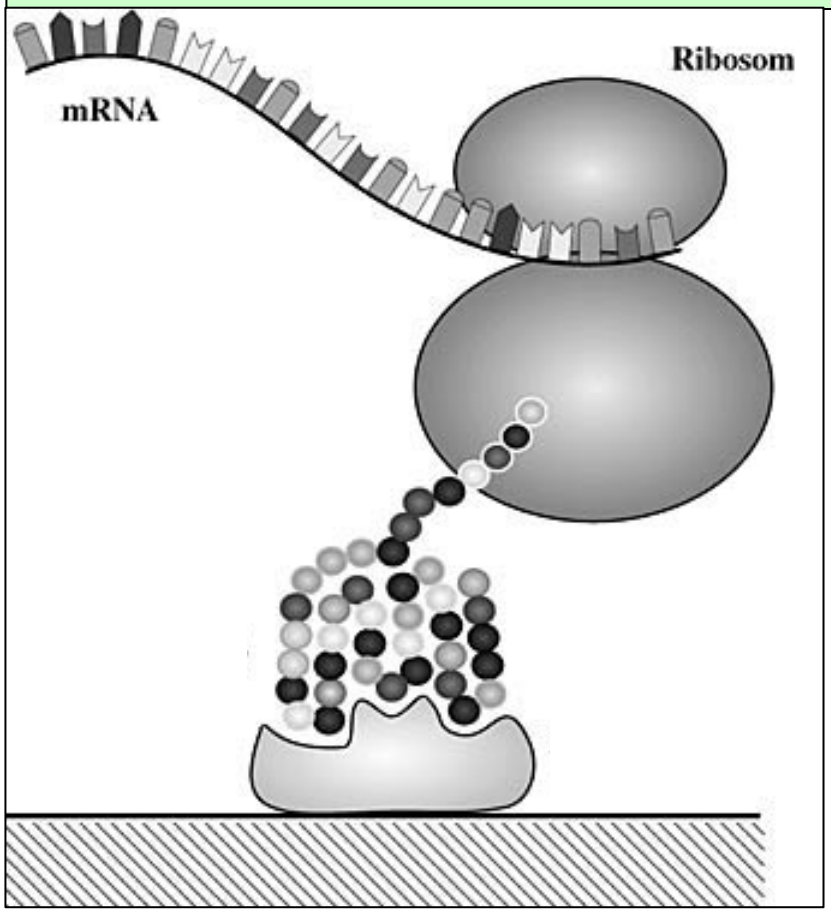
# Incubation


Exponential phase
Stationary phase
Lag phase
Death (declining) phase
Number of cells
Time

# Expression



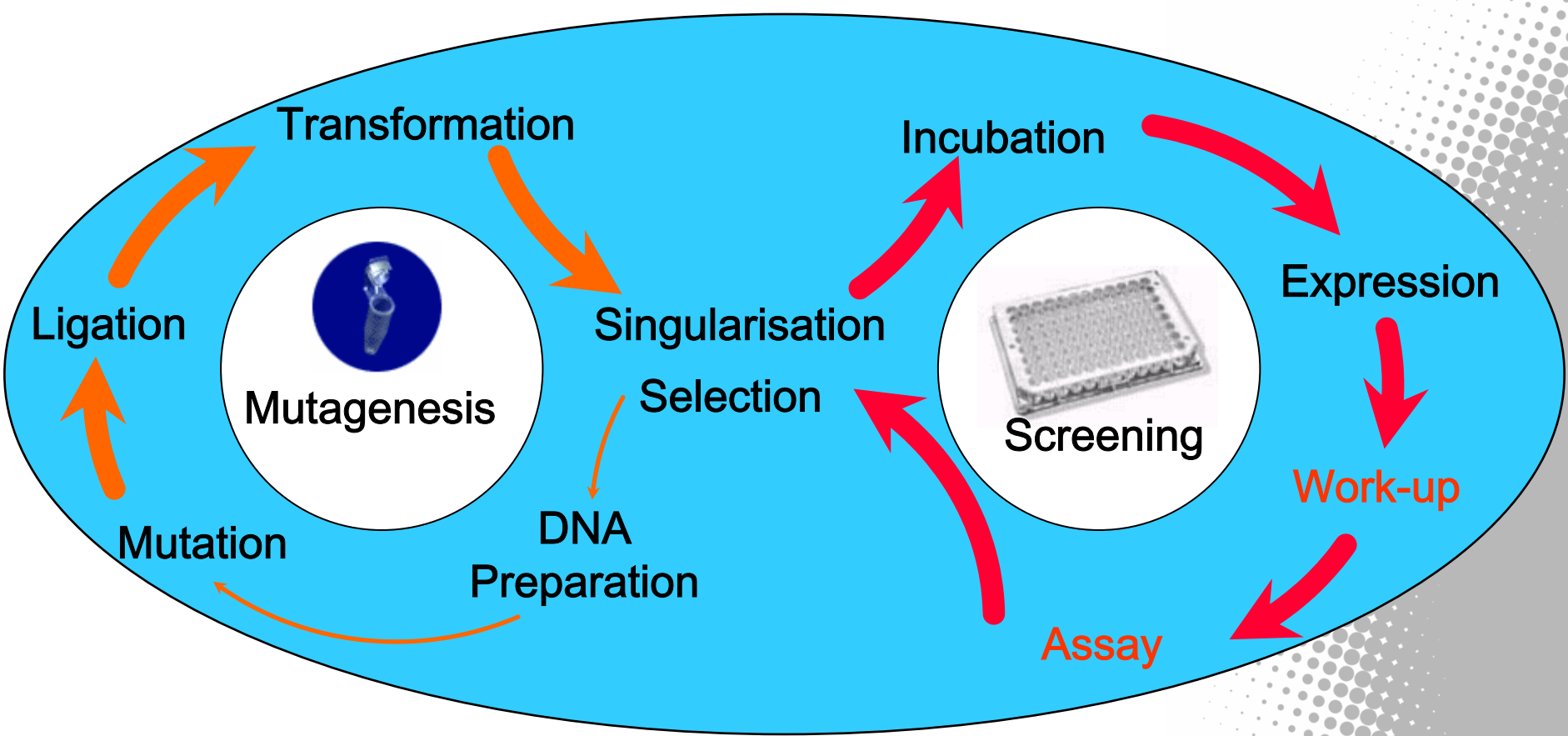
## Molecular Switch for Initiation of Protein Synthesis




Induction by

- Temperature shift
- Chemicals like
  - iso-Propyl-ß-Thiogalactosid (IPTG)
  - Arabinose
  - Rhamnose

# Evolutionary Methods: Practical Performance



# Work-up

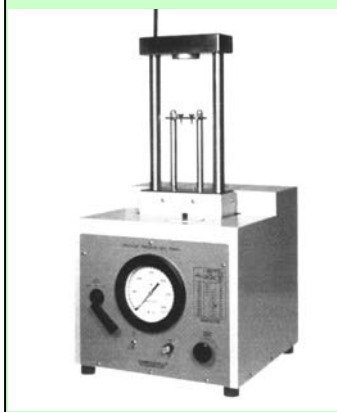
Cell Disruption: Destruction of cell membranes to bring the expressed protein in contact with the substrate

## Principles

**Solvents**
- *Osmotic*
- *Enzymatic*

**Mill**
*Mechanical stress*

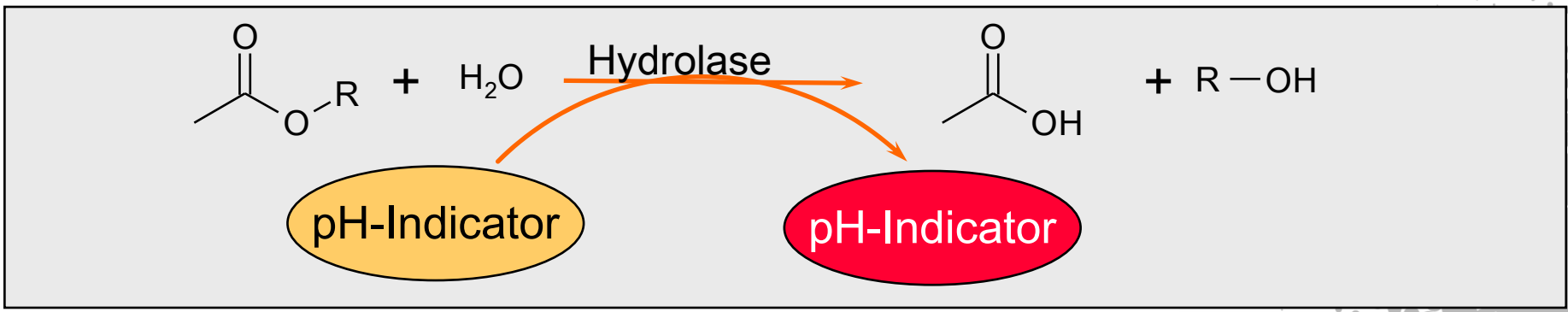
**French-Press**
*Pressure*



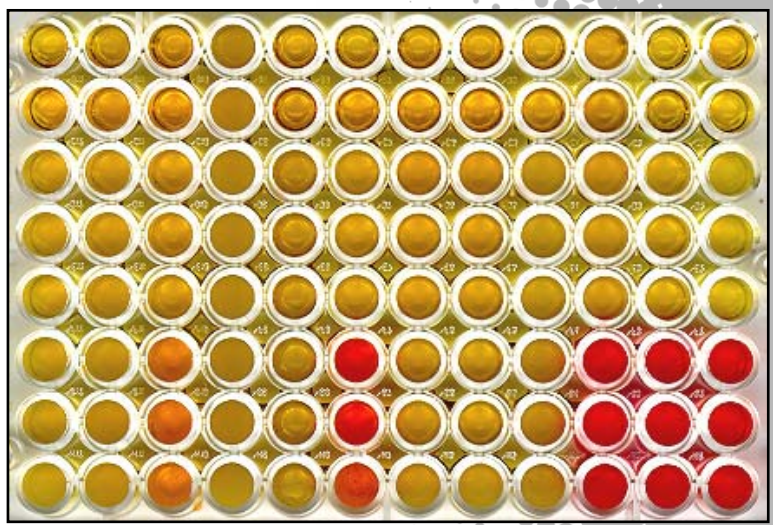
**Sonifier**
*Ultrasound*

**Freezing/ Thawing**

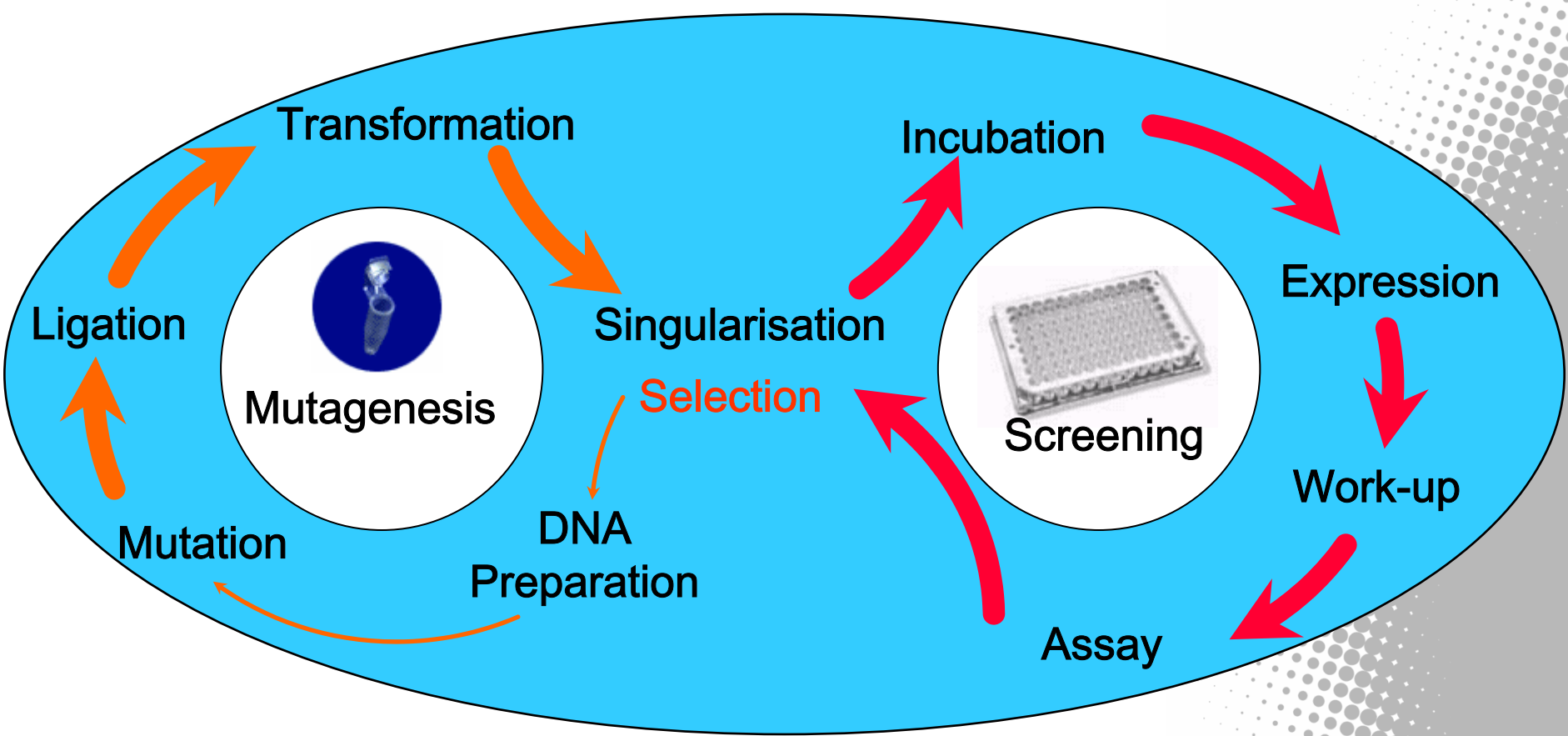
# Assay



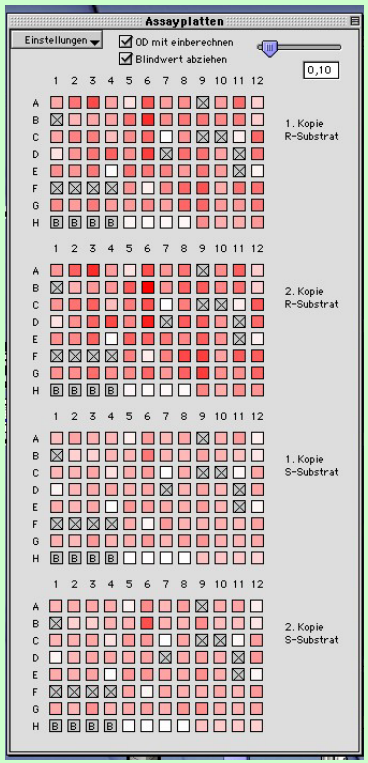

- You only get what you screen for
- Fast
- Colorimetric



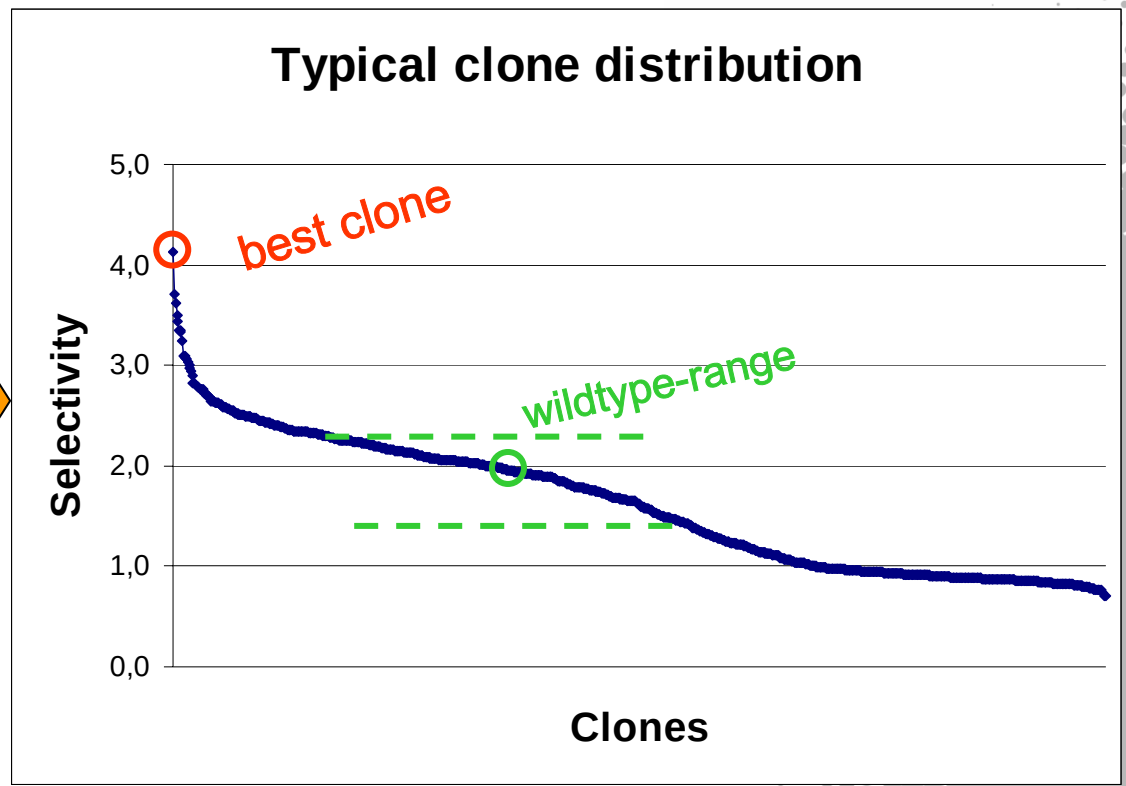
# Evolutionary Methods: Practical Performance



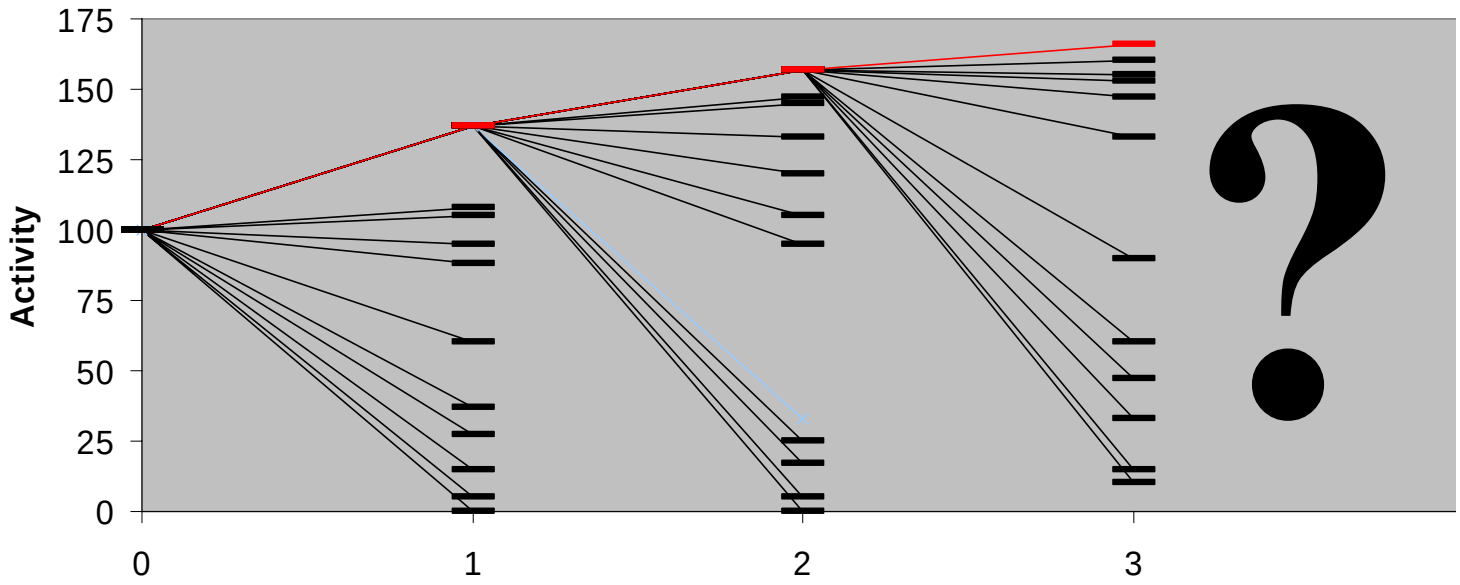
# Data-Analysis and Selection

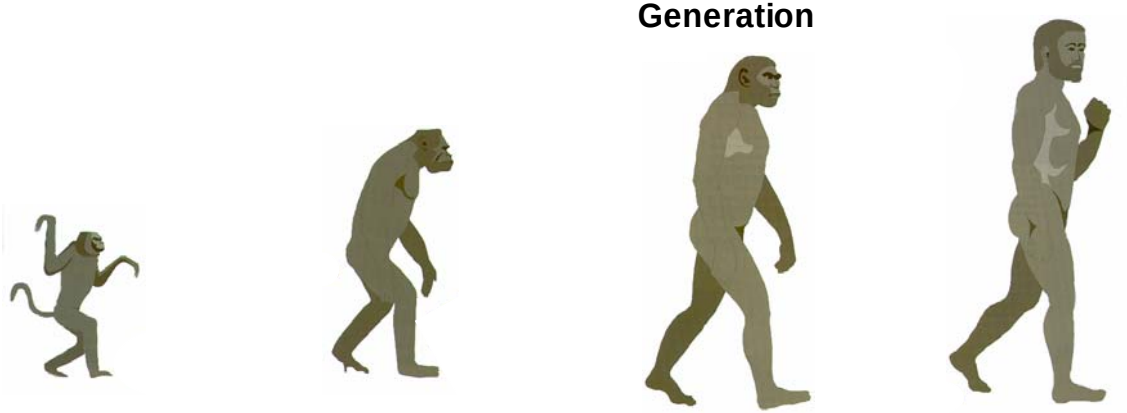



Data-Management

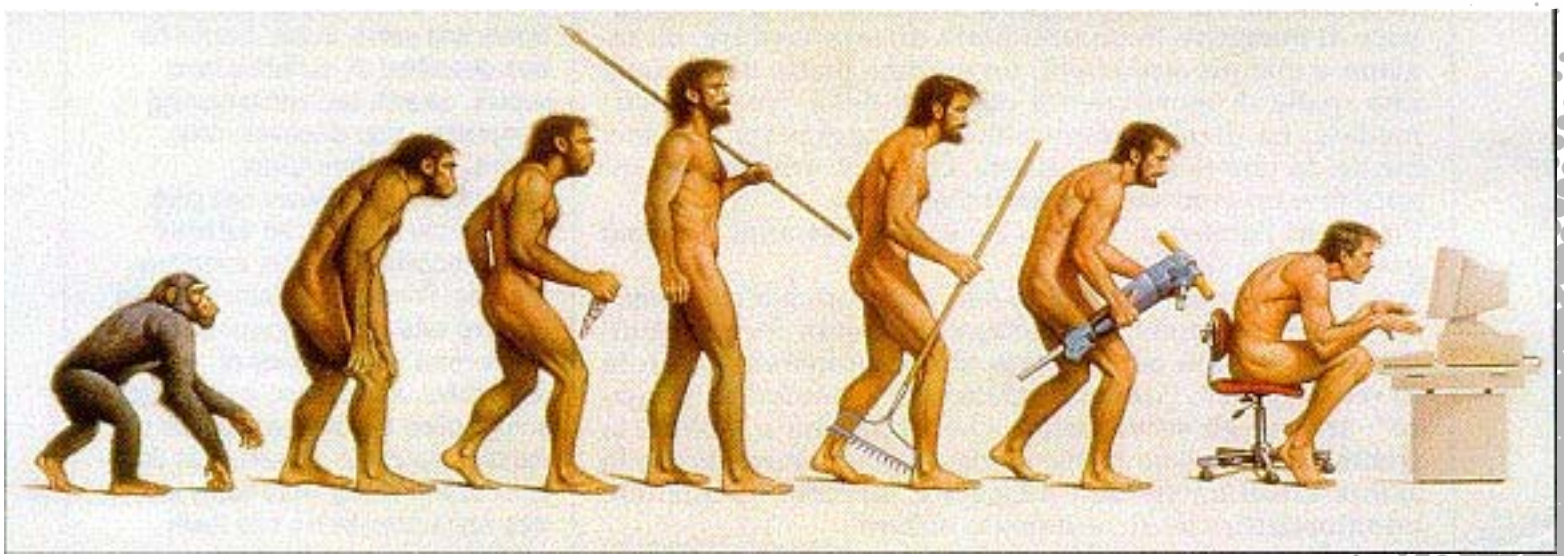


# Directed Evolution



# Directed Evolution –natures outcome



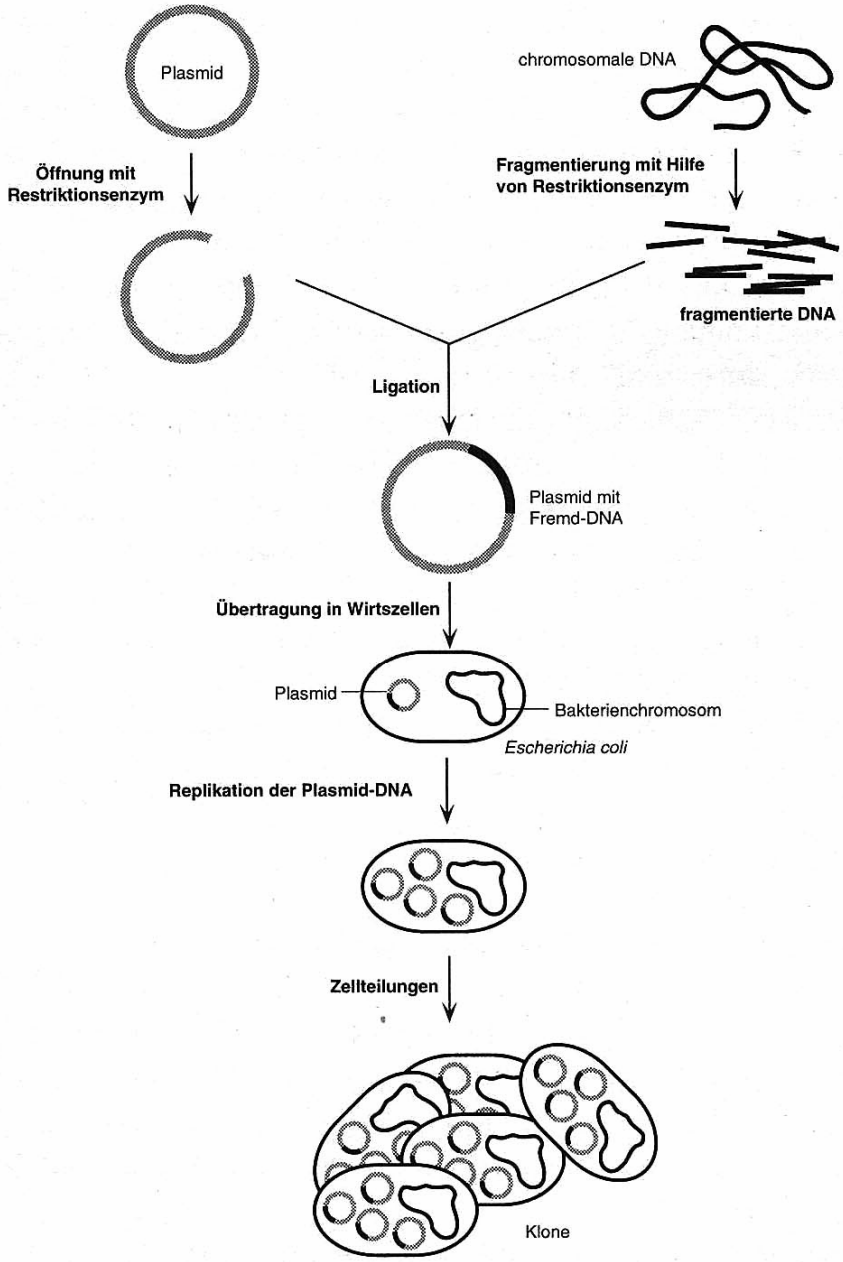



**Abbildung 36: Schematische Darstellung der Klonierung eines DNA-Frag-**

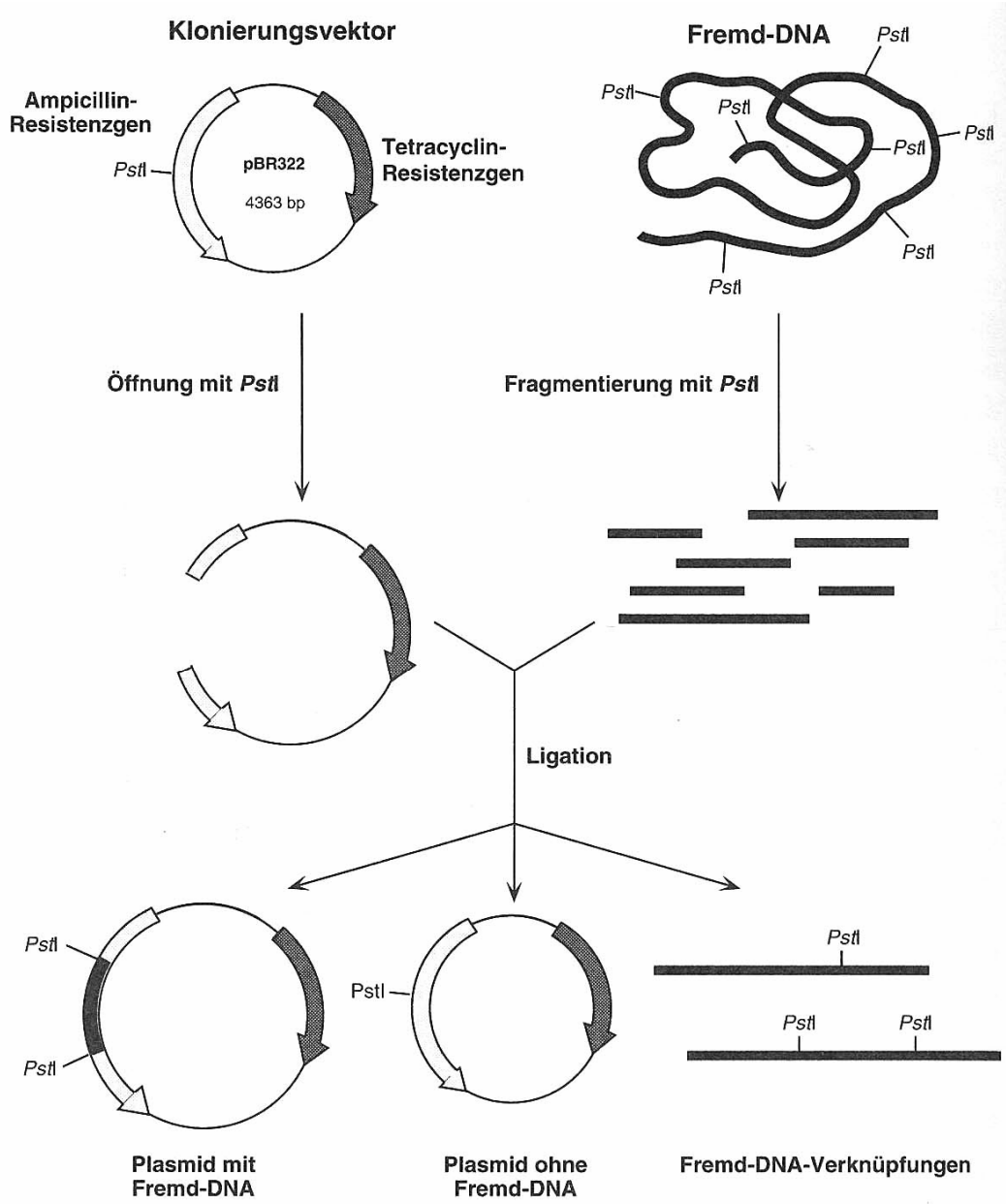
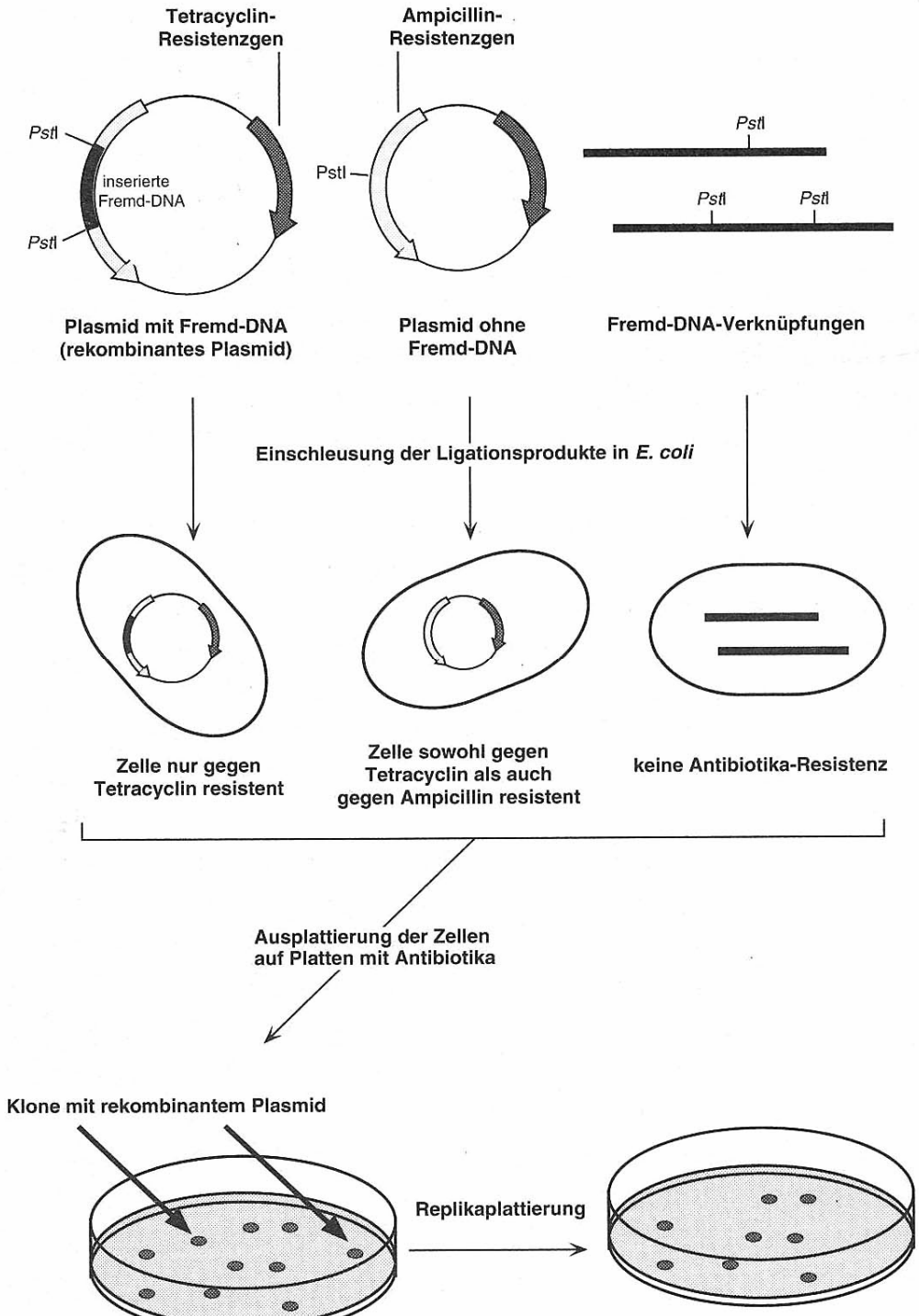



**Abbildung 39: Schema zum Einfügen von Fremd-DNA in das Plasmid pBR322**

Das Tetracyclin-Resistenzgen ist durch einen dunkel gerasterten Balken und das


Tetracyclin-Resistenzgen
Ampicillin-Resistenzgen
PstI
inserierte Fremd-DNA
PstI
PstI
PstI
PstI
PstI
Plasmid mit Fremd-DNA (rekombinantes Plasmid)
Plasmid ohne Fremd-DNA
Fremd-DNA-Verknüpfungen
Einschleusung der Ligationsprodukte in E. coli
Zelle nur gegen Tetracyclin resistent
Zelle sowohl gegen Tetracyclin als auch gegen Ampicillin resistent
keine Antibiotika-Resistenz
Ausplattierung der Zellen auf Platten mit Antibiotika
Klone mit rekombinantem Plasmid
Replikaplattierung