# Restriction / Modification Systems

## Host-controlled restriction and modification

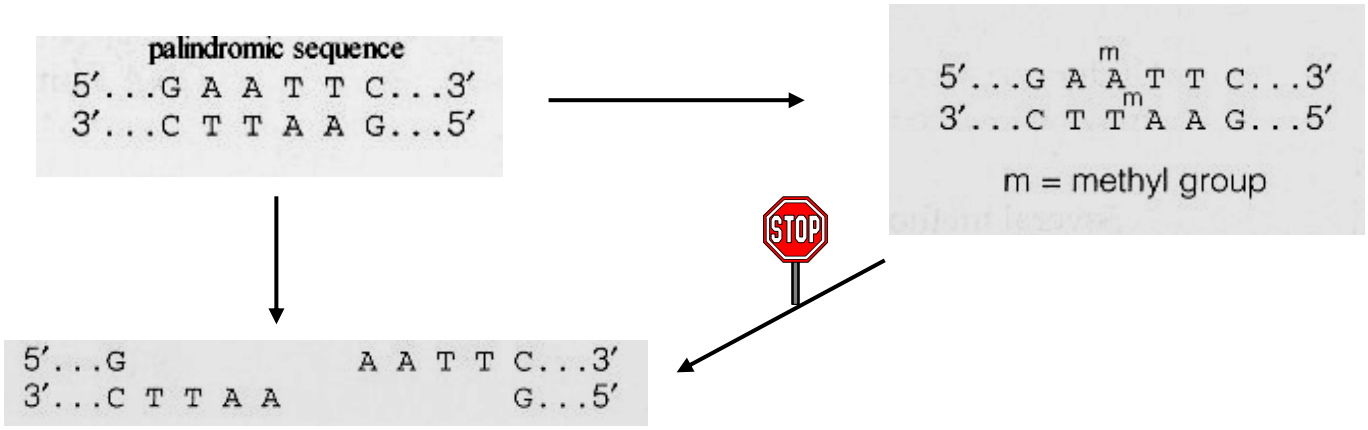
**Restriction** systems allow bacteria to monitor the origin of incoming DNA and to destroy it, if it is recognized as foreign. Restriction **endonucleases** recognize **SPECIFIC** sequences in the incoming DNA and cleave the DNA into fragments, either at specific sites or more randomly, thus preventing it from successfully replicating and parasitizing the cell.

(Immunity system? Self vs. non-self DNA recognition).

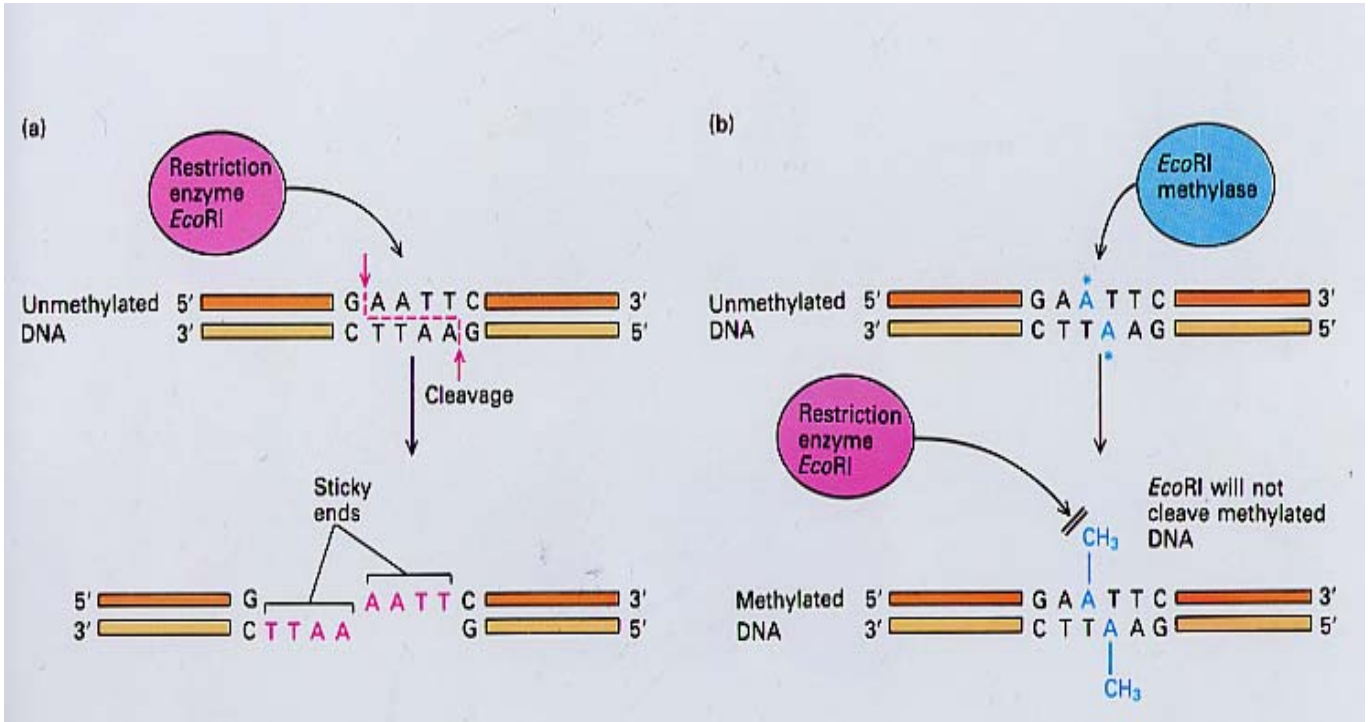
The restrictive host must, of course, **protect** its own DNA from the potentially lethal effects of the endonuclease and so its DNA must be appropriately **modified**. Modification involves methylation of certain bases at a very limited number of sequences within DNA.

Together, a restriction endonuclease and its 'cognate' modification methyl-transferase form a **restriction-modification** (R-M) **system.**

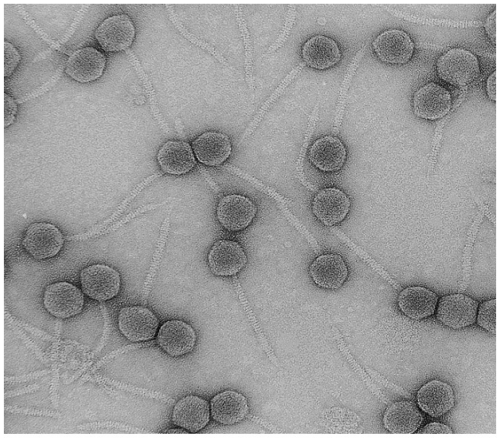
## EcoRI restriction endonuclease-methylase system (type II)

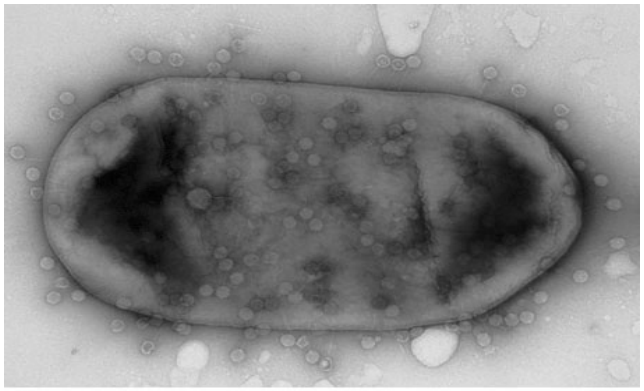



Note: this class of enzymes is distinct from general nucleases, which cleave DNA **randomly**, either from the ends (**exonucleases**) or at internal sites (**endonucleases**).

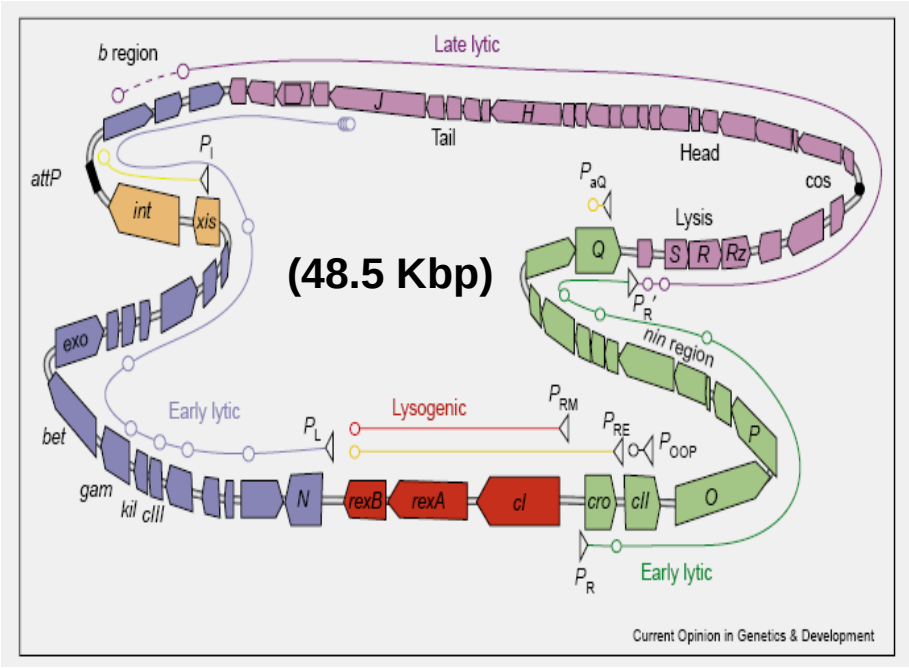


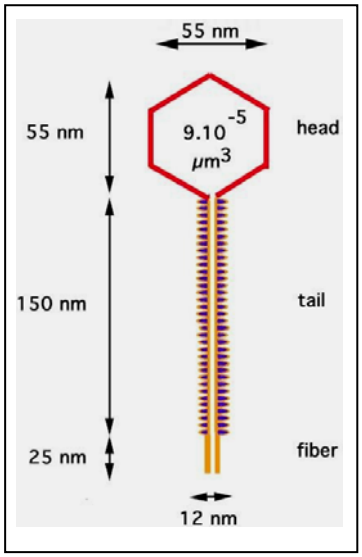
# All started with … lambda phage

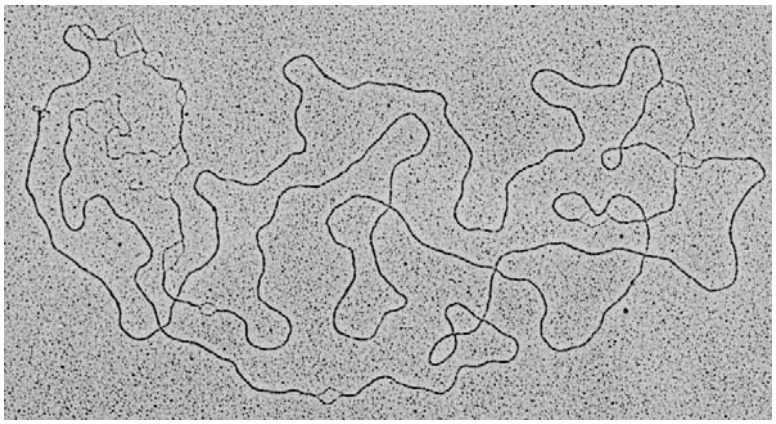




EM pictures of Lambda phages

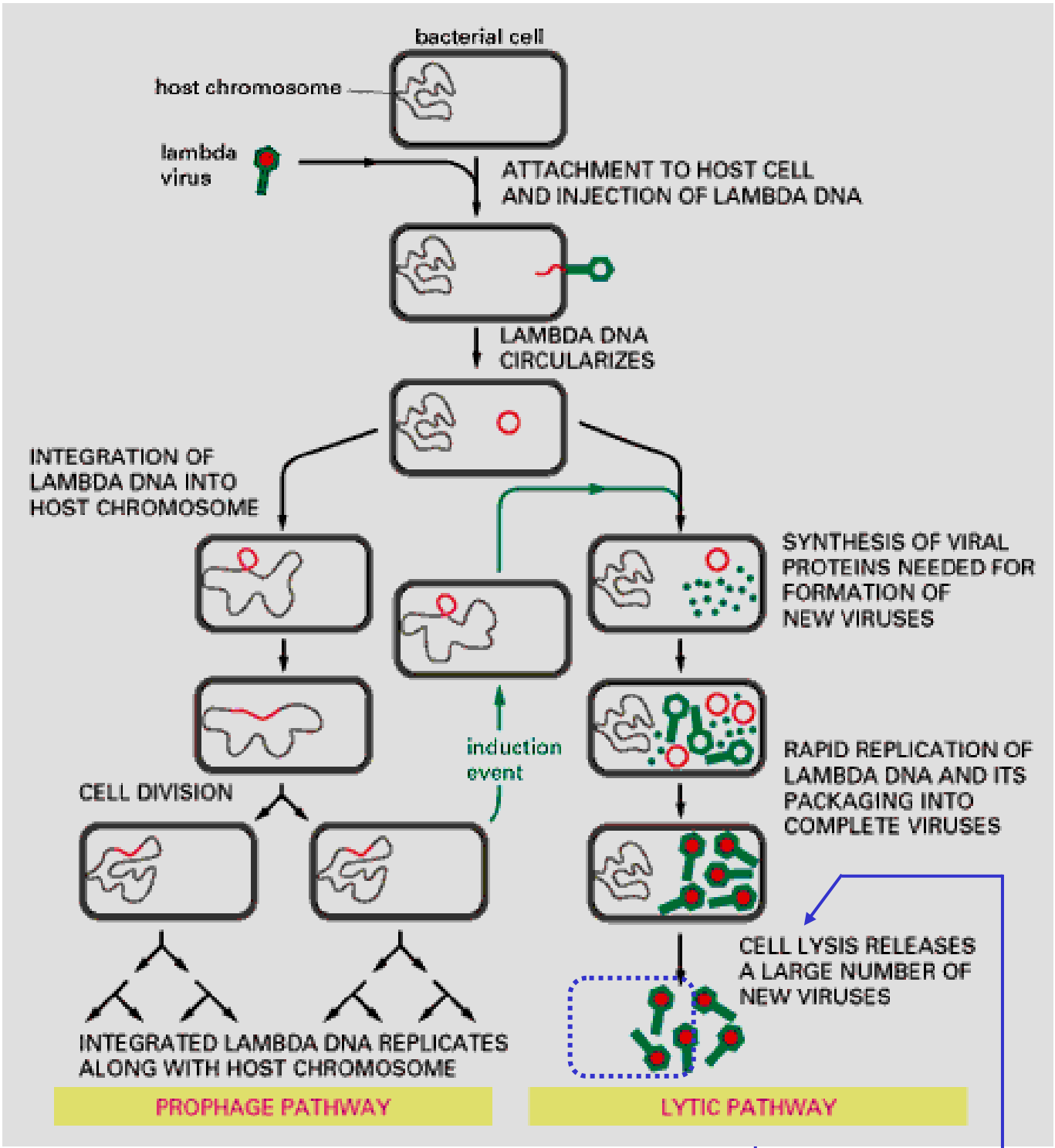


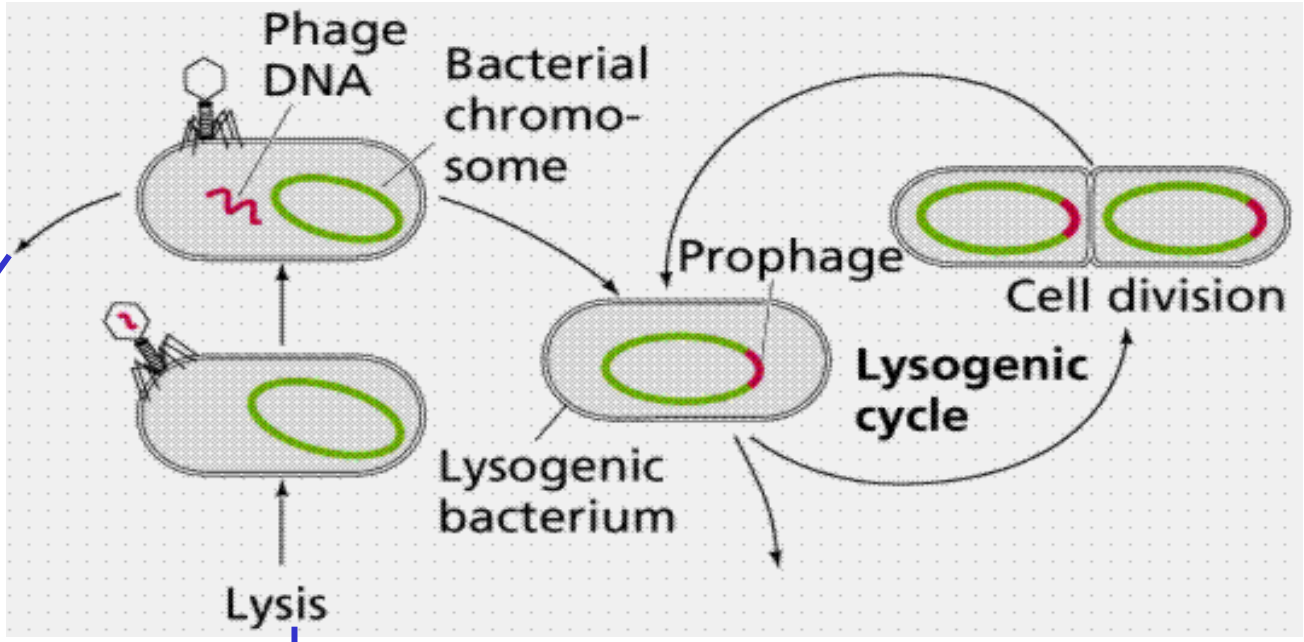



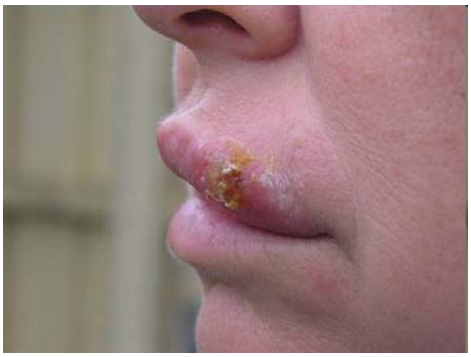
Partially denatured Lambda DNA

# The Life Cycle of Bacteriophage Lambda



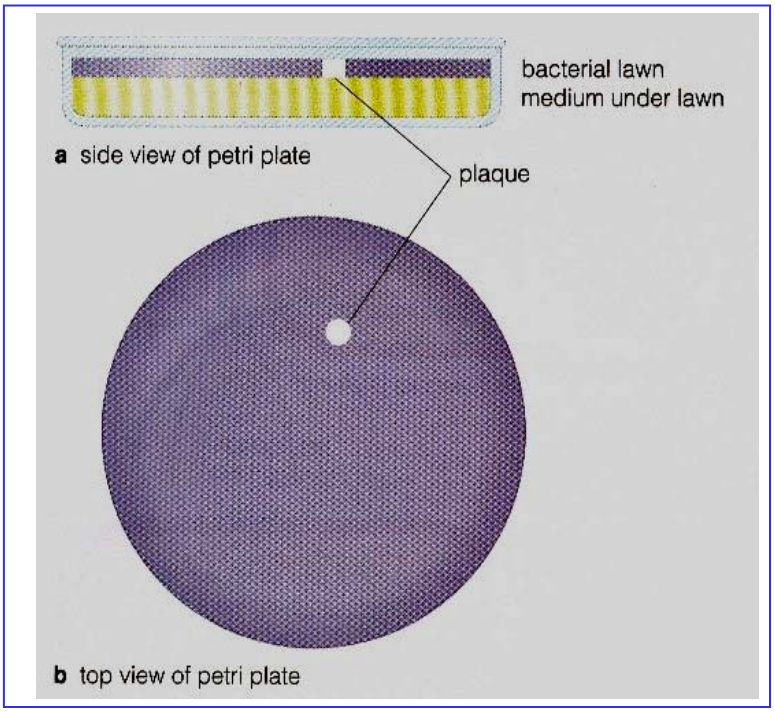

**LYSOGENIC PATHWAY**



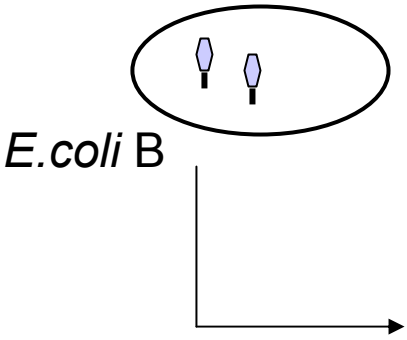
Analogies with Herpes Simplex Virus

Petri plate with a uniform lawn of *E.coli* cells, and one phage plaque

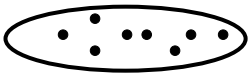


The phenomena of restriction and modification were well illustrated and studied by the behaviour of phage lambda on two *E.coli* strains.

(Luria,S.E., Human,S.L. 1952; *J. Bacteriol.*)
(Bertani,G., Weigle,J.J. 1953; *J. Bacteriol.*)



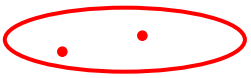

Isolation of lambda



*E.coli* B

EOP= 1

?

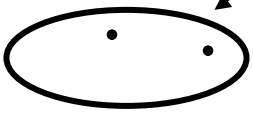


*E.coli* K

EOP = $1/10^4$ = 0.0001

Phages are said to be *restricted* by the second host strain (through the action of a nuclease)

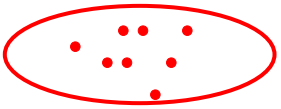
Isolation of lambda



*E.coli* B

EOP= $1/10^4$ = 0.0001

?

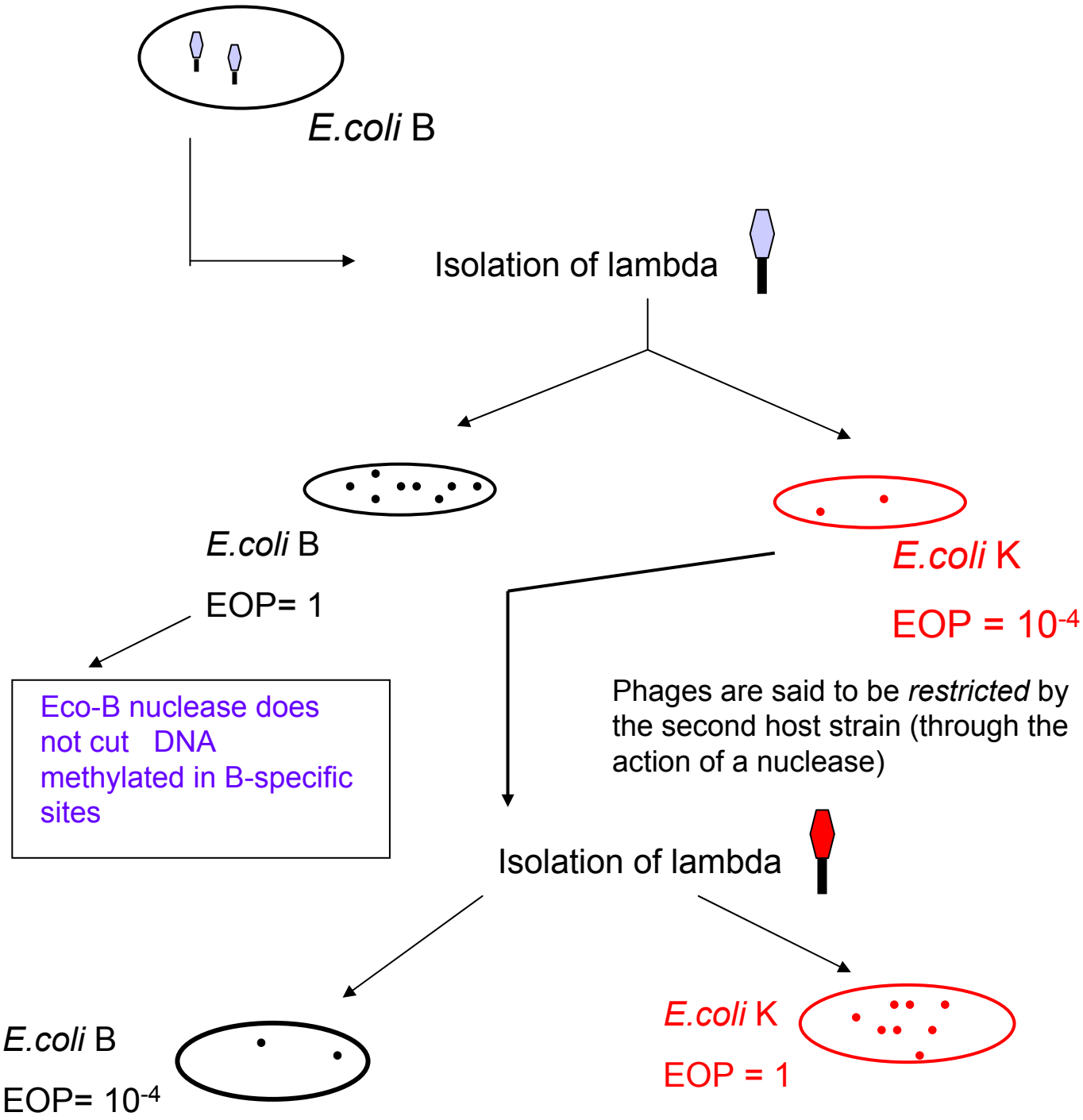


*E.coli* K

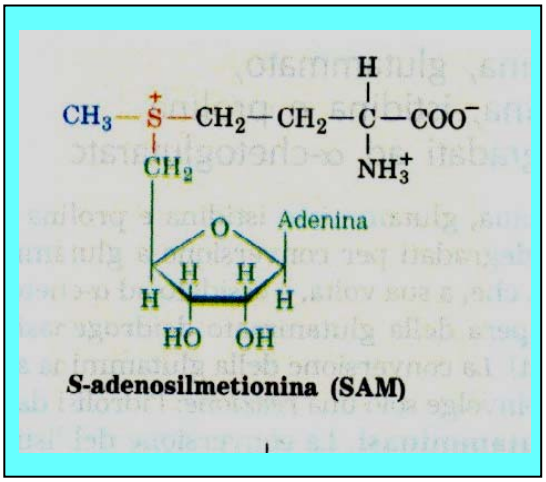
EOP = 1

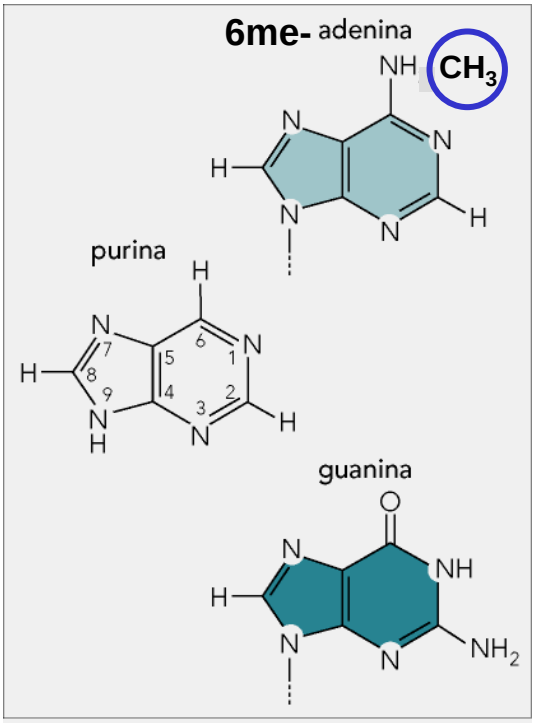
Phages replated on *E.coli* K are no longer restricted.

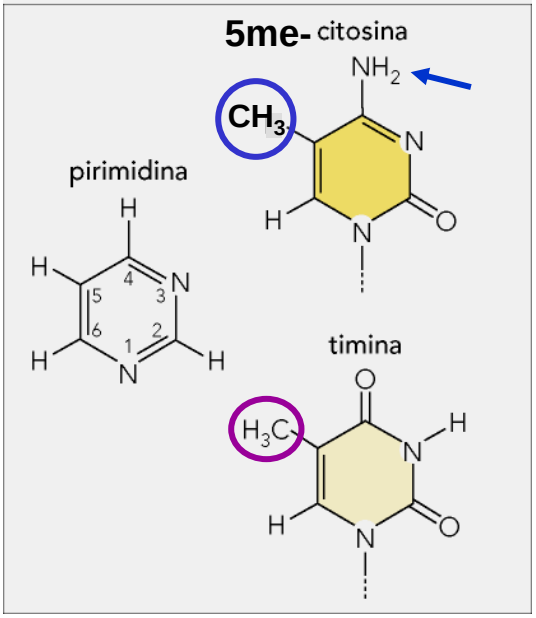
This non heritable change conferred upon the phage by the second host strain is called *modification (methylase)*


E.coli B
Isolation of lambda
E.coli B
EOP= 1
E.coli K
EOP = 10-4
Eco-B nuclease does not cut DNA methylated in B-specific sites
Phages are said to be restricted by the second host strain (through the action of a nuclease)
Isolation of lambda
E.coli B
EOP= 10-4
E.coli K
EOP = 1
Eco-K nuclease cuts DNA methylated in B-specific sites; few phages escape the hydrolysis and are modified in K-specific sites
Phages replated on E.coli K are no longer restricted. This non heritable change conferred upon the phage by the second host strain is called modification (methylase)

**DNA methyltransferases** (MTases) comprise a biologically important class of enzymes that are found in most organisms, from bacteria to mammals.




They catalyze the transfer of the activated methyl group from the cofactor S-adenosyl-L-methionine (AdoMet) to specific DNA sequences of:

-the exocyclic amino group of adenine (**N**6-adenine DNA MTases → **6meA**)

- the exocyclic amino group of cytosine (**N**4-cytosine DNA MTases, **less common** → **4meC**)

- the 5 position of cytosine (**C**5-cytosine DNA MTases → **5meC**)

Between 2% and 7% of the **Cytosines** of animal cell DNA are methylated (the value varies with the species).

Most of the methyl groups are found in CG 'doublets', and usually the C on both strands are methylated, giving the structure

**5' C- G 3'** ⟶ **5' $^{m}$C- G 3'**
**3' G- C 5'** ⟶ **3' G- C$^{m}$ 3'**

At present, it is not fully clear the relationship between the presence of **methyl groups** and the control of gene expression (in *Drosophila* can be detected only in the embryonic stages; *C.elegans* and *Yeast* are considered to lack DNA methylation).

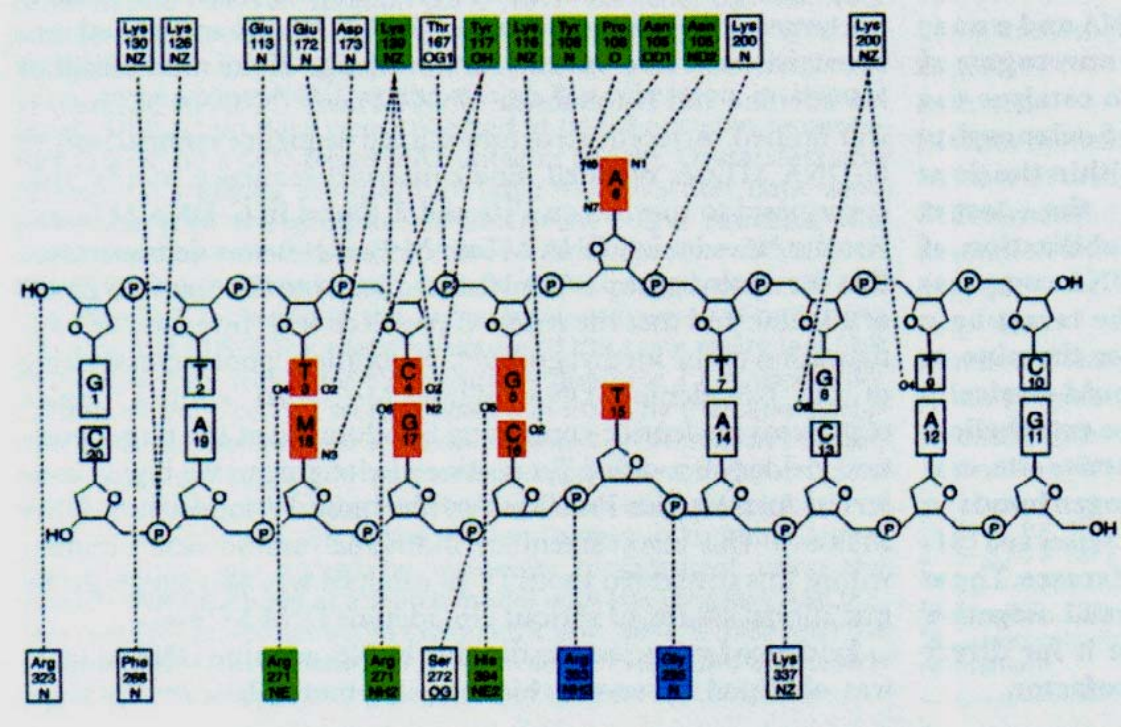
---

The catalytic mechanism involves a **rotation** of the target adenine/cytosine---deoxyribose out of the DNA helix → a process called **'base flipping'.**

How is DNA base flipping initiated is unknown; the base itself has nothing to do with this process. The structure of M.*Hha*I (isolated from *Haemophilus haemolitycus*) and a DNA substrate having an abasic (apurinic/apyrimidinic) site reveals that the enzyme still rotates the **deoxyribose** to the 'flipped out' position.

Schematic representation of **hydrogen bonds** and **salt bridges** between M.*Taq*I and the DNA substrate (10 bp duplex oligodeoxynucleotide)

Direct contacts between amino acid residues (green) of the catalytic domain and the bases of the recognition sequence (5'-TCGA-3') are formed within the widened minor groove



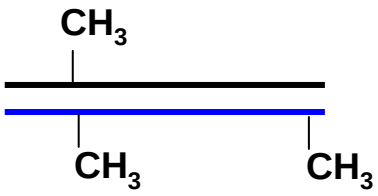
Note: Surprisingly, restriction and modification enzymes from the same R-M system share little amino acid sequence similarity even though they recognize the same DNA sequence

In **eukaryotes** methylation has a different purpose: distinguishing genes in different functional conditions.

Gene expression is associated with demethylation. It has been demonstrated that a given gene is **inactive when methylated**, but becomes active if it is non-methylated [ON/OFF switch under normal/pathological conditions].

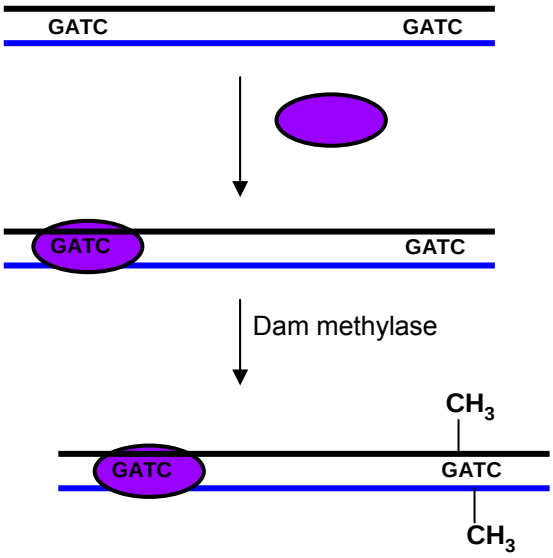
DNA may exist in:

- doubly methylated form,
- unmethylated form,
- hemi-methylated (mono-methylated) form in the cell.




The modification enzymes are generally monomeric proteins; their primary reaction is the restoration of full methylation following DNA replication.

[ DNA methylation can be regarded as an **increase** in the **information content** of DNA ]




Regulatory proteins that bind to non-methylated target sites can modulate gene expression by protecting them from methylation.

Because **no changes** occur within the **DNA** primary **sequence**, this type of heritable gene regulation is considered "**EPIGENETIC**"

Characteristics of **restriction endo-nucleases**

| | Type I | Type II | Type III |
|---|---|---|---|
| Restriction and modification activities | Single multifunctional enzyme | Separate endonuclease and methylase | Separate enzymes with a subunit in common |
| Protein structure of restriction endonuclease | Three different subunits | Simple | Two different subunits |
| Requirements for restriction | ATP, $Mg^{2+}$ S-adenosylmethionine | $Mg^{2+}$ | ATP, $Mg^{2+}$ (S-adenosylmethionine) |
| Sequence of host specificity sites | *Eco*B: $TGAN_8TGCT$ *Eco*K: $AACN_6GTGC$ | Rotational symmetry (not in type IIs) | *Eco*P1: AGACC *Eco*P15: CAGCAG |
| Cleavage sites | Possibly random, at least 1000 bp from host specificity site | At or near host specificity site | 24–26 bp to 3′ of host specificity site |
| Enzymatic turnover | No | Yes | Yes |
| DNA translocation | Yes | No | No |
| Site of methylation | Host specificity site | Host specificity site | Host specificity site |

N = any nucleotide.

1%

**A** is the target

< 1%

## Type I

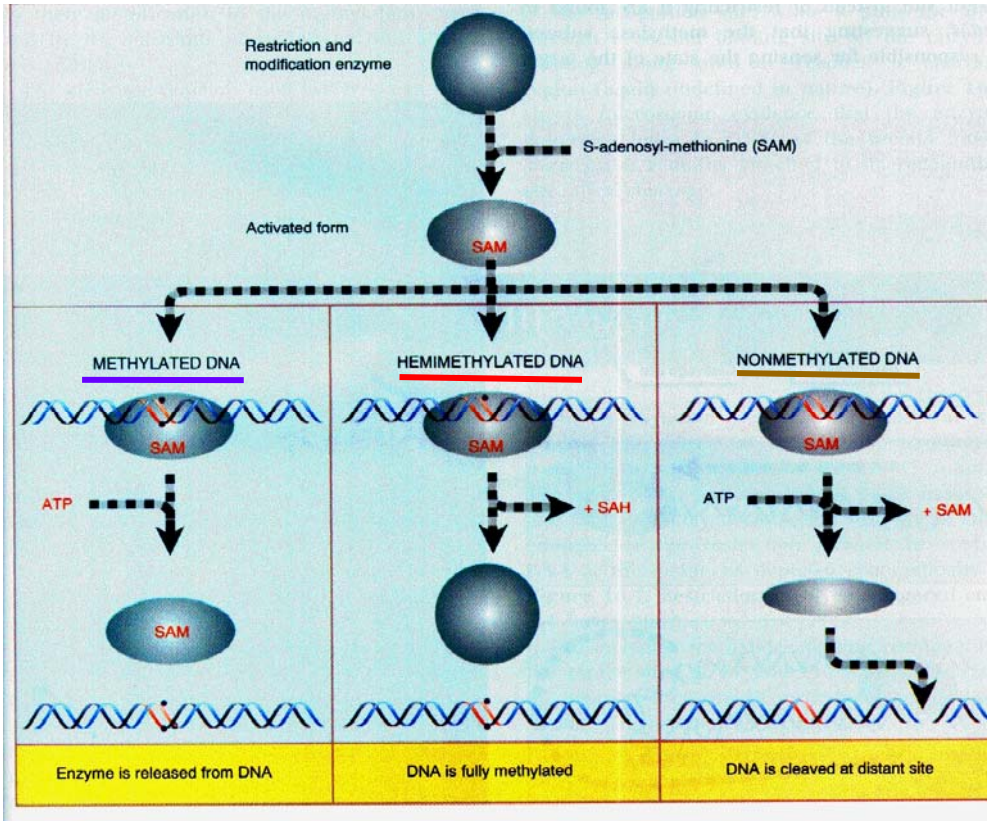
The type I enzymes, represented by the *Eco*K and *Eco*B activities of *E.coli* strains K and B, were the first to be discovered. *Eco*K restriction enzyme, encoded by *hsd*RMS genes, attacks DNA that is **not** protected by **adenine methylation** at the appropriate recognition site.

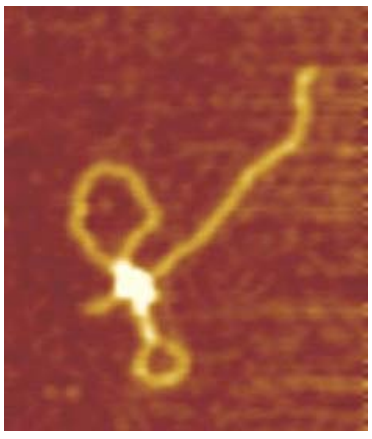
They require $Mg^{++}$ ions, ATP and S-adenosylmethionine (SAM). The enzymes are composed of three subunits: a specificity subunit (S) which determines the recognition specificity, a modification subunit (M), and a restriction subunit (R).

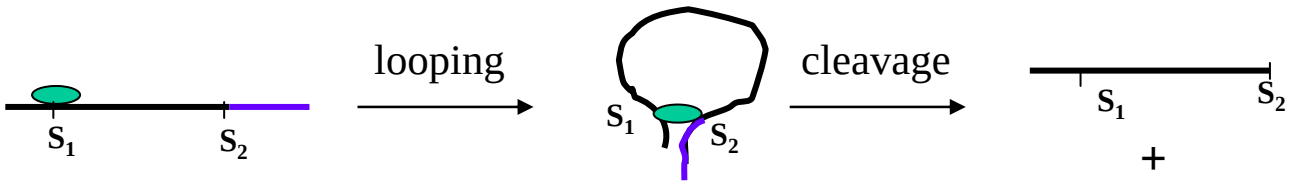
*hsd*R mutations abolish restriction but not methylation (r-m+), mutations in *hsd*S and *hsd*M are r-m- (they can neither modify nor restrict DNA).

If the recognition sequence is:

- methylated in both strands (met-**Adenosine**) → ATP stimulates the dissociation of enzyme from DNA;
- hemi-methylated (met-**Adenosine** in only one strand) → ATP stimulates the methylation of the other strand, with SAM being the methyl donor; [When this happens?]
- non methylated → cleavage occurs. The enzyme first translocates along the DNA (or pulls the DNA ?) by looping it; cleavage then takes place at a random site several kilobases from the recognition site





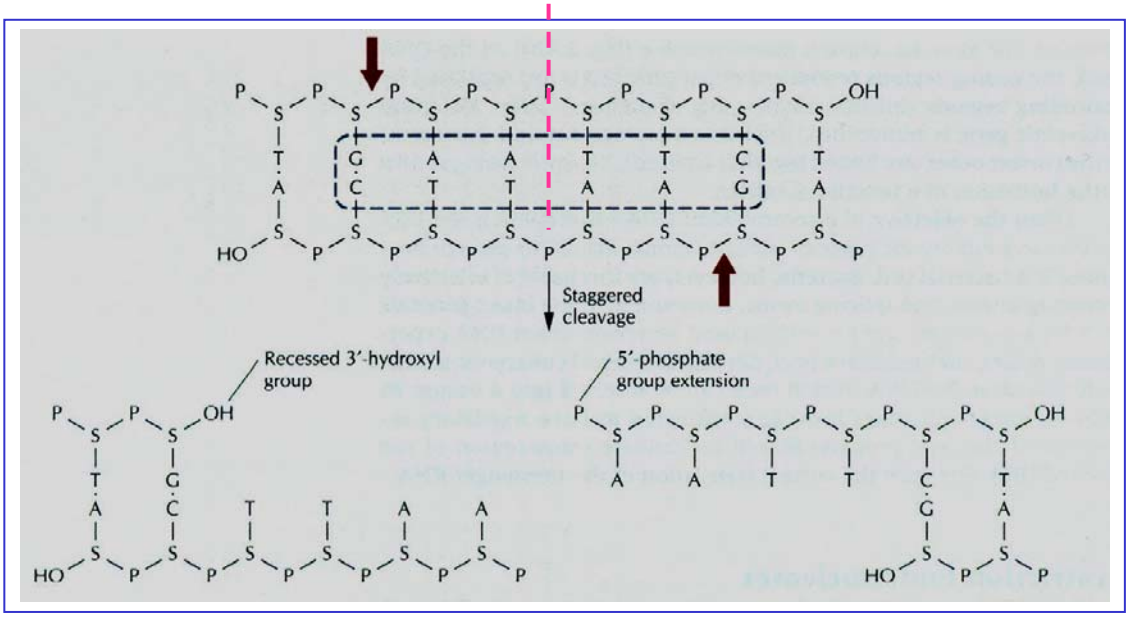
## Type II (binary systems)

This class of enzyme is particularly useful for gene manipulation, and present day DNA technology is totally dependent upon our ability to **cut DNA molecules at specific sites**, using type II restriction enzymes.

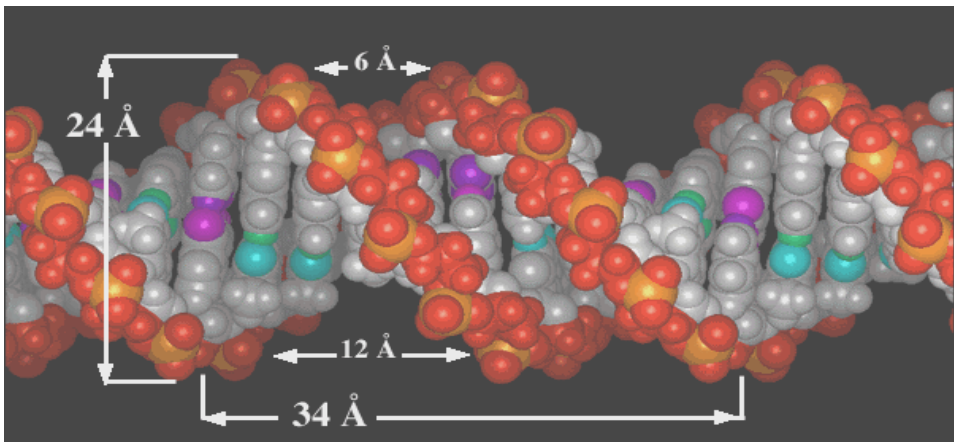
These enzymes recognize a particular target sequence (of 4 ; 5 ; 6; 7 or 8 nucleotides) in a DNA molecule and cleave **both strands** of the duplex within, or near to, that sequence to give rise to discrete DNA fragments of defined length and sequence. They consist of a single polypeptide and require only **Mg** ions.

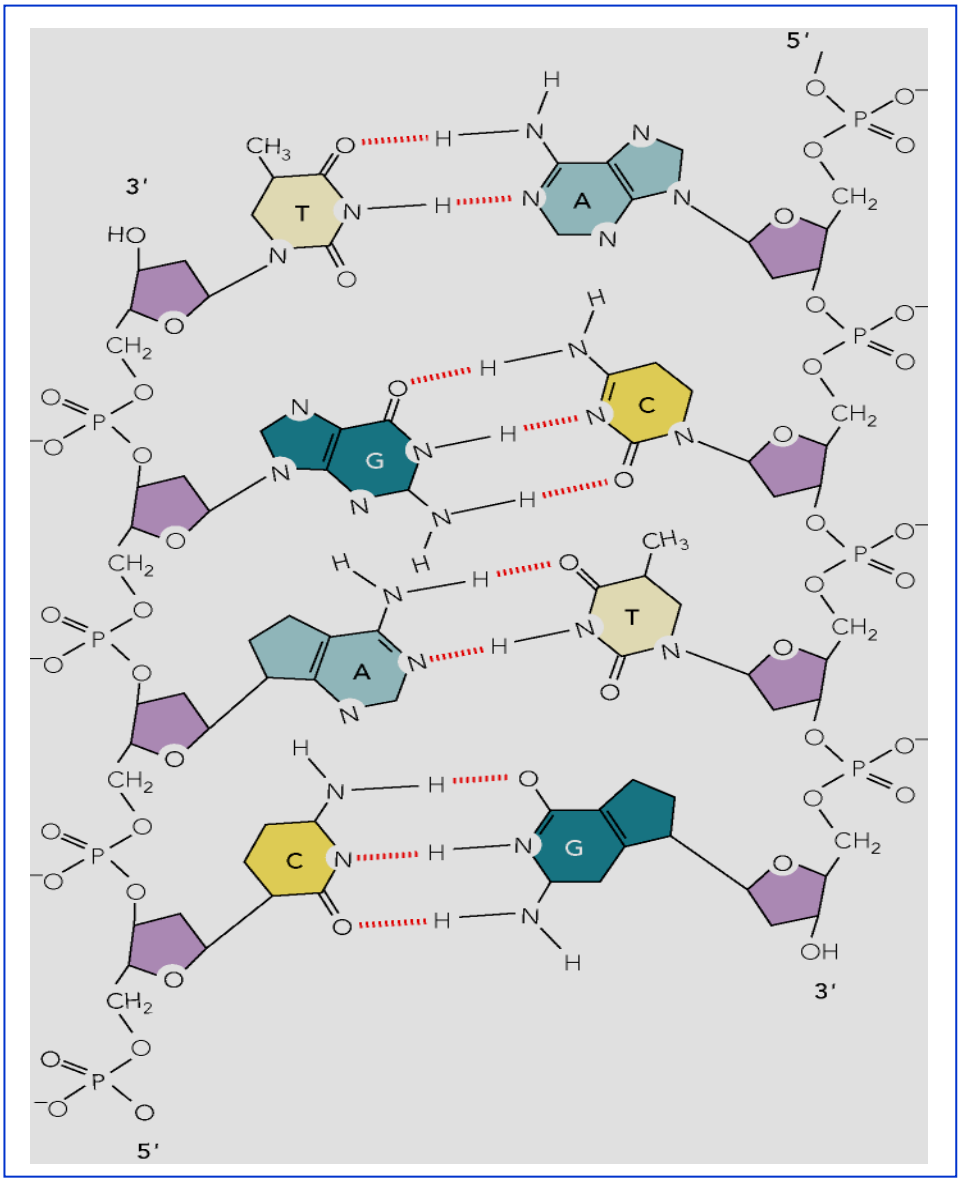
Recognition sequences are symmetric, some sequences are continuous (e.g. **G A T C**), some are interrupted (e.g. **G A N T C**). **Type II $_S$** (**s**hifted cleavage) systems, like MboII or FokI, differ from standard type II system in having asymmetric recognition sequences. Cleavage occurs only on one side, at a point some distance away ( < than 20 nucleotides).

Two-fold axis of simmetry: the 5' → 3' sequence in the 'top' strand is the same as that in the 'bottom' strand




Symmetrical, staggered cleavage of a short fragment of DNA by the type II restriction endonuclease *Eco*RI. The bold arrows show the sites of cleavage in the DNA backbone.
(S=deoxyribose sugar; P=phosphate group)




Detailed structure of a polynucleotide chain

Note the opposite orientation of the two strands

*Eco*RI restriction endonuclease bound to the

5’ GAATTC 3’ specific sequence of dsDNA

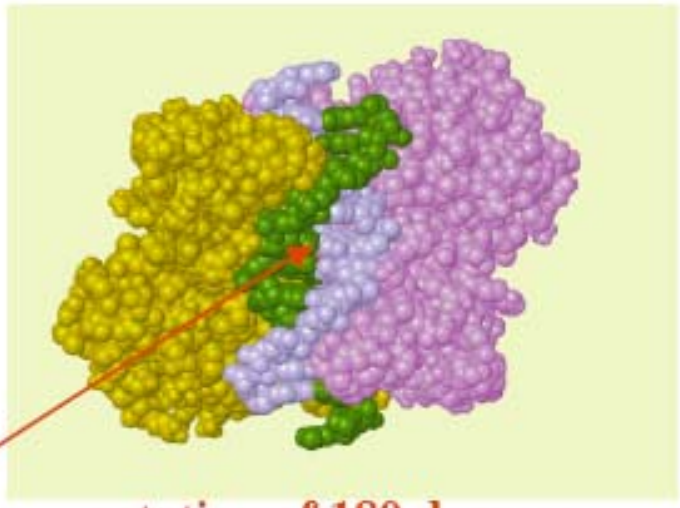
3’ CTTAAG 5’

► Homodimer of two identical protein subunits (purple and yellow)

► Bound to a palindromic DNA sequence

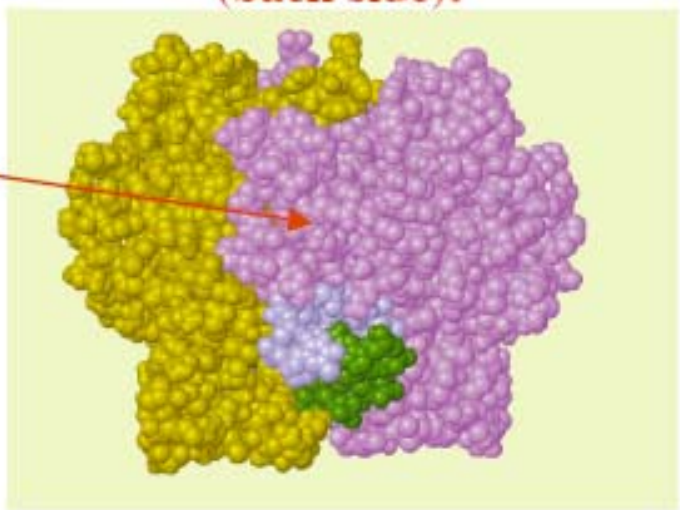
(same sequence 5’ → 3’ on light-blue and green DNA strands)

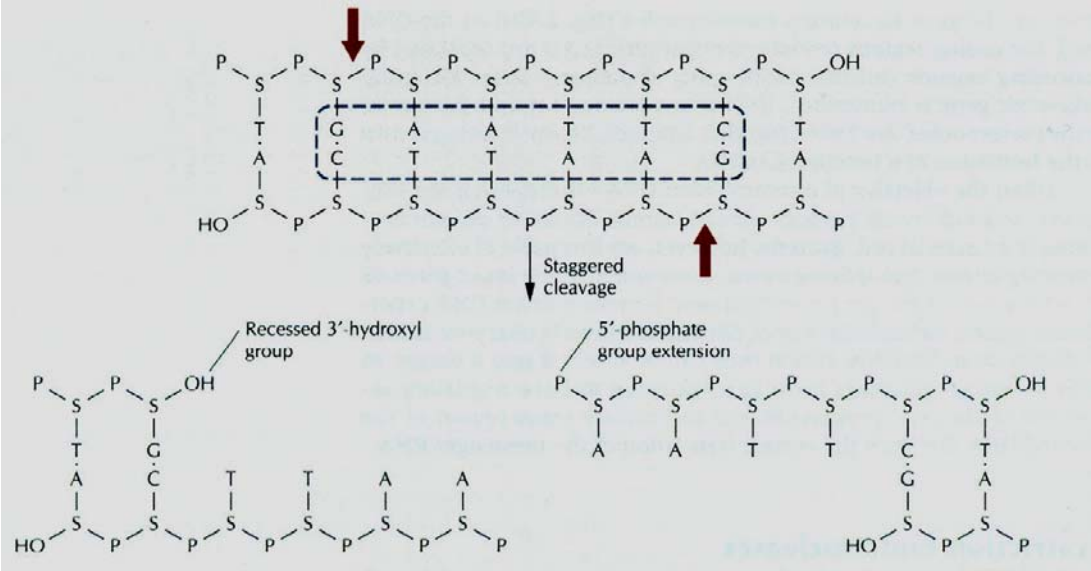
Looking into the **minor groove**



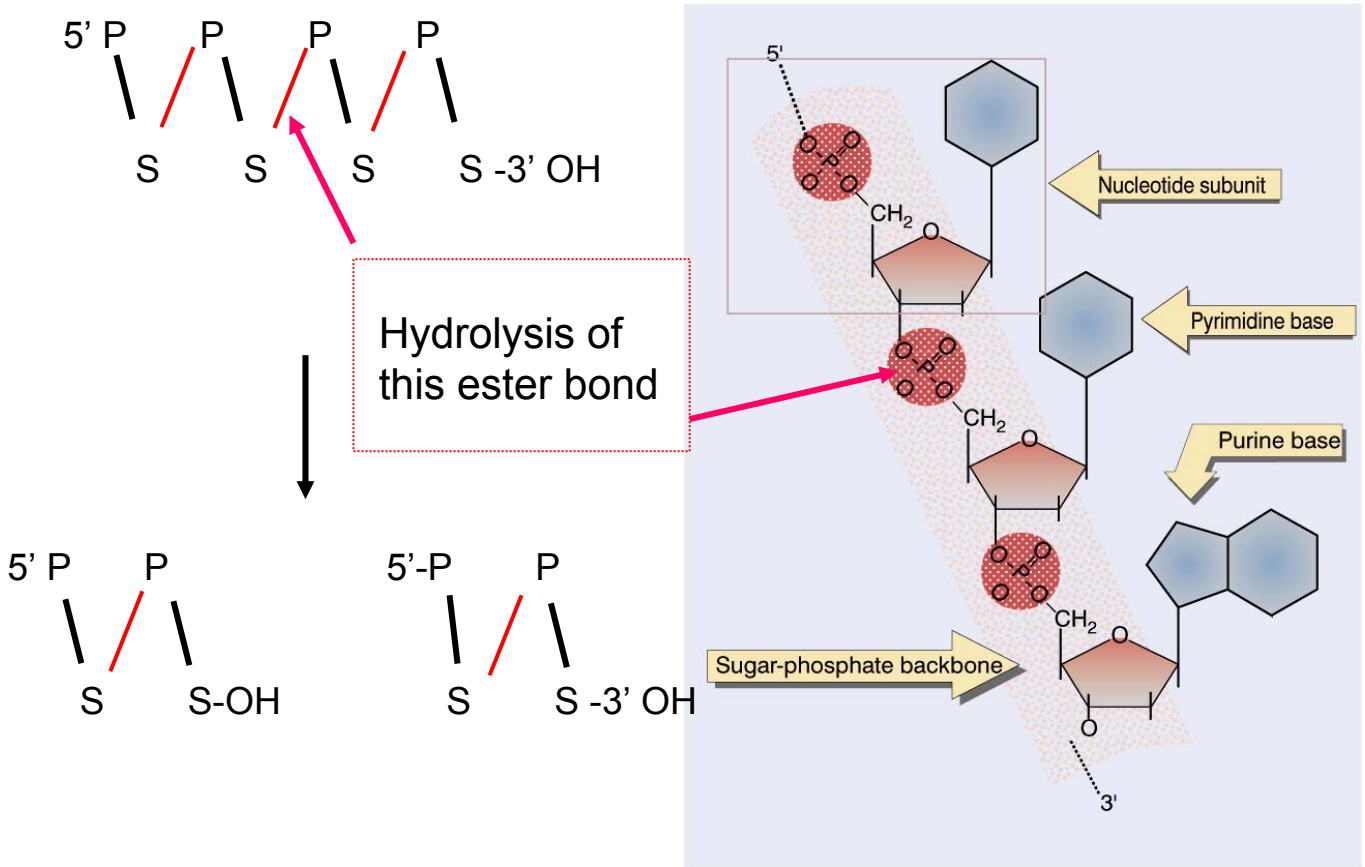
rotation of 180 degrees (back side):

Looking into the **major groove**






**Details** of the cleavage reaction occurring on the backbone of **one** DNA strand:




The cleavage occurs at the ester bond between the sugar and the phosphate group, following the path 5’ ⟶ 3’

(Red bond — in the scheme; **NOT** the black bond!)

# “Sticky” ends

**5′ overhang (*Eco*RI)**


**3′ overhang (*Pst*I)**


# “Blunt” ends

**Blunt-ends (*Sma*I)**

# Type II enzymes

Interrupted sequence

## RESTRICTION ENDONUCLEASES

### *Sfi* I
MEDIUM 50°C

5'...GGCCNNNNNGGCC...3'
3'...CCGGNNNNNCCGG...5'

**Description:** Isolated from *Streptomyces fimbriatus.*

**Ligation and Recutting:** After 10-fold over digestion with *Sfi* I, greater than 95% of the DNA fragments can be ligated and recut with this enzyme.

**Assay Conditions:** Medium buffer. Incubate at 50°C.

**Reference:** Quiang, B. Q. and Schildkraut, I. (1984) *Nucleic Acids Res.* **12**, 4507-4515

| *Catalogue No:* | *Quantity:* |
|---|---|
| *AB-0403-a* | *750u* |
| *AB-0403-b* | *2,000u* |

3'-OH protruding;

8-bases recognition sequence
(rare cutting enzyme)

### *Sma* I
TA 25°C

5'...CCCGGG...3'
3'...GGGCCC...5'

**Description:** Isolated from *Serratia marcescens.*

**Ligation and Recutting:** After 2-fold over digestion with *Sma* I, greater than 90% of the DNA fragments can be ligated and recut with this enzyme.

**Assay Conditions:** T.A. buffer. Incubate at 25°C.

| *Catalogue No:* | *Quantity:* |
|---|---|
| *AB-0314-a* | *4,500u* |
| *AB-0314-b* | *15,000u* |

Blunt-end

### *Spe* I
MEDIUM 37°C

5'...ACTAGT...3'
3'...TGATCA...5'

**Description:** Isolated from *Sphaerotilus species.*

**Ligation and Recutting:** After 10-fold over digestion with *Spe* I, greater than 90% of the DNA fragments can be ligated and recut with this enzyme.

**Assay Conditions:** Medium buffer. Incubate at 37°C.

**Reference:** Comb, D.G., and Schildkraut, I. Cited in Roberts, R.J. (1985) *Nucleic Acids Res.* **13**, r165-r200

| *Catalogue No:* | *Quantity:* |
|---|---|
| *AB-0243-a* | *400u* |
| *AB-0243-b* | *1,000u* |

5'-P protruding

(sticky ends)

### *Sph* I
MEDIUM 37°C

5'...GCATGC...3'
3'...CGTACG...5'

**Description:** Isolated from *Streptomyces phaeochromogenes.*

**Ligation and Recutting:** After 10-fold over digestion with *Sph* I, greater than 95% of the DNA fragments can be ligated and recut with this enzyme.

**Assay Conditions:** Medium buffer. Incubate at 37°C.

**Reference:** Finch, L.Y., *et al.* (1980) *Gene* **10**, 39-46

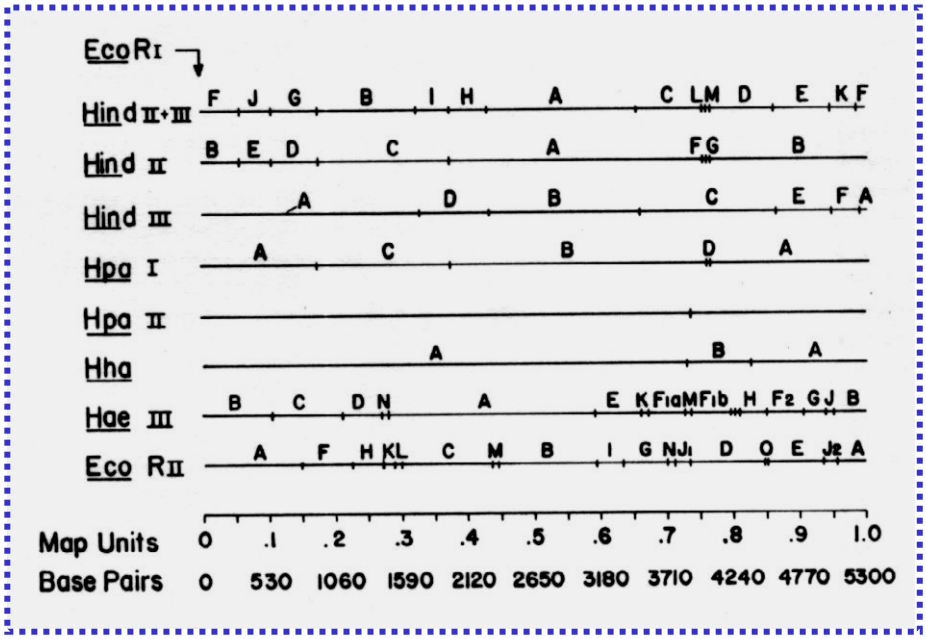
| *Catalogue No:* | *Quantity:* |
|---|---|
| *AB-0526-a* | *200u* |
| *AB-0526-b* | *1,000u* |

3'-OH protruding

(sticky ends)

Use of restriction endonucleases:

✂ physical mapping of chromosomes




Cleavage map of the Simian Virus 40 genome.

The zero point of the map is the *Eco*RI site

Specific endo R cleavage sites or fragments serve as **physical reference** in the map → "cleavage map" or "fragment map"

Once constructed, the map can serve as a framework for localizing functions and genes and for relating nucleotide sequences to the entire genome.

✂ mapping and isolation of genes

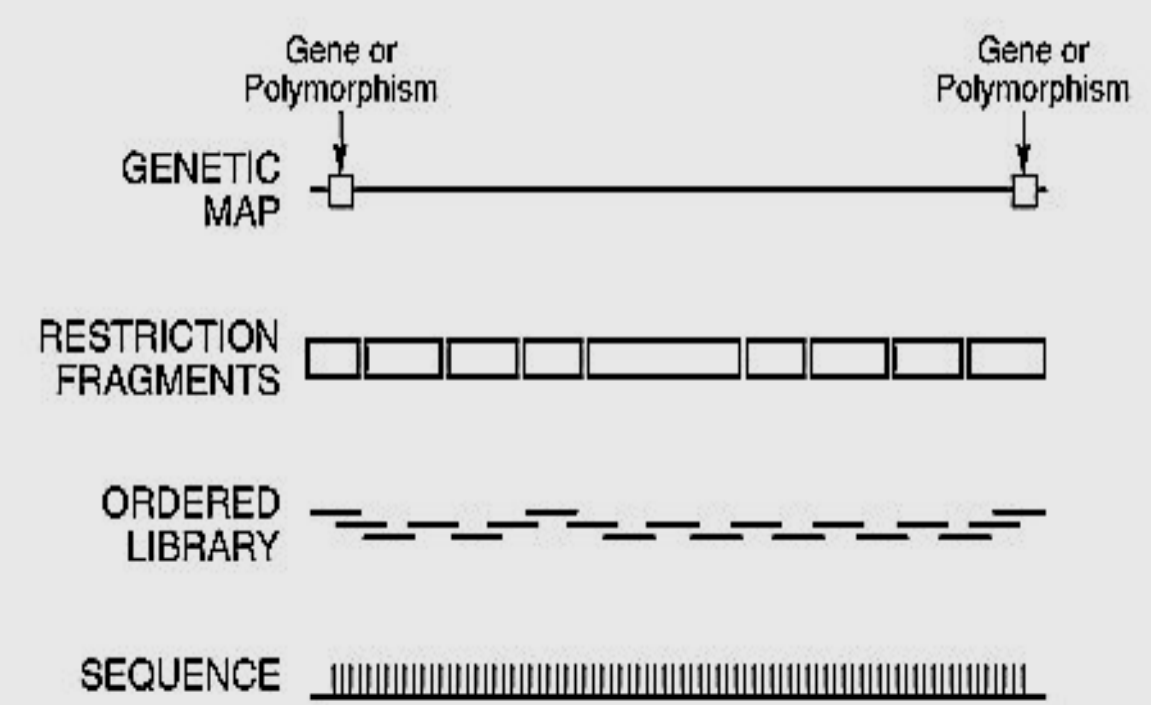
✂ location of protein binding sites (if a DNA fragment contains a protein-binding site, it can be retrieved as a protein-DNA complex on a membrane filter)

✂ nucleotide sequence analysis

✂ *in vitro* restructuring and cloning of DNA molecules

✂ generation of mutant (e.g. excision of DNA segments → production of deletion mutants)

… remember this ?



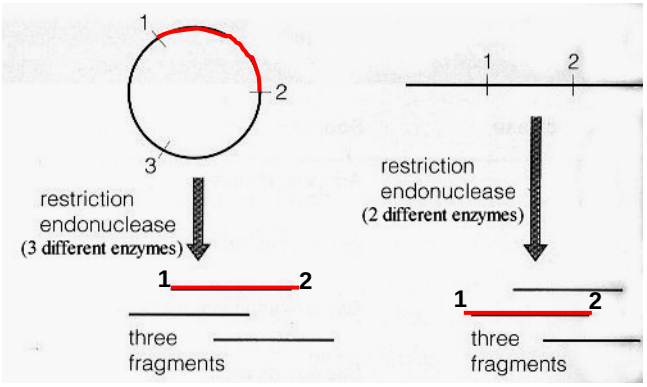

**Genetic Map** → measuring recombination frequencies of "linked markers" (genes or polymorphisms, whose pattern of transmission can be tracked) → low resolution

**Restriction Map** → alignment of 1-2 Mbp DNA fragments → medium resolution

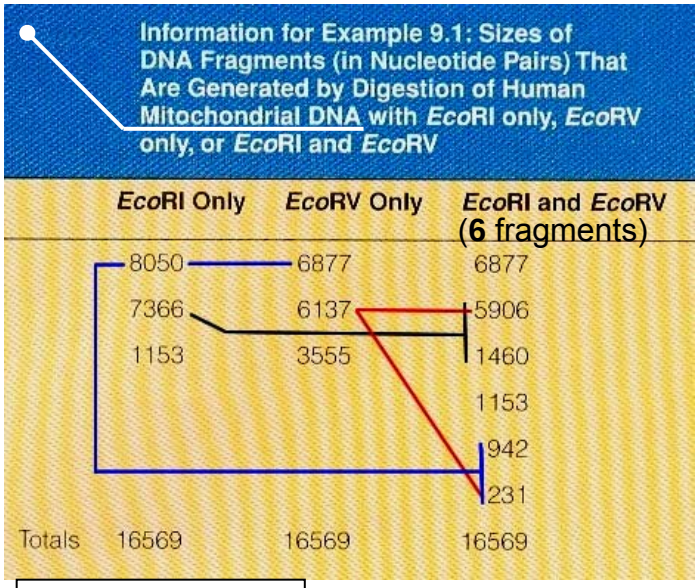
**Libraries** → 40-400 Kbp DNA fragments inserted into artificial chromosomes (YAC, BAC, cosmids) → high resolution

**Nucleotide Sequence** → "is the ultimate physical map"

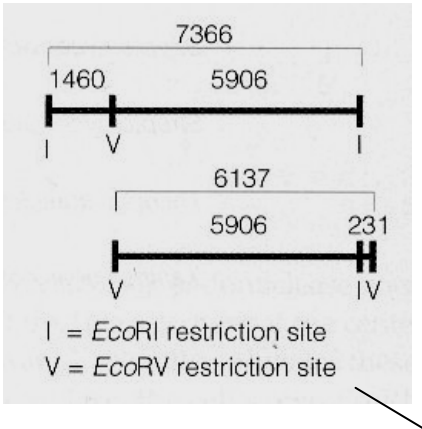
# Restriction mapping of DNA molecules



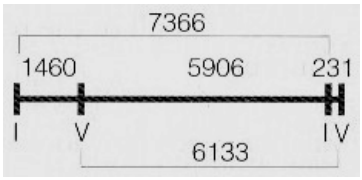

**Circular DNA** **Linear DNA**

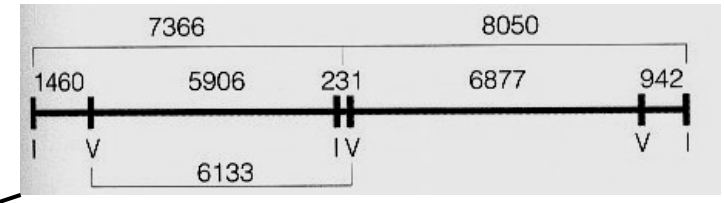


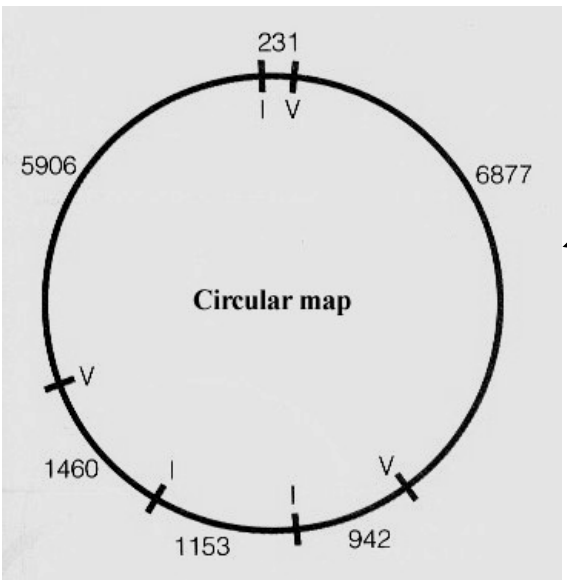

5’ GAATTC 3’
3’ CTTAAG 5’




5’ GATATC 3’
3’ CTATAG 5’




6 fragments

Nearly 3400 restriction enzymes have been found, exhibiting over 220 distinct specificities. Most of the enzymes found today turn out to be duplicates (**isoschizomers**) of already discovered sequence specificities.

Restriction enzymes are **species non-specific**:

- enzymes of the same specificity occur in different species (e.g **Dra I** from *Deinococcus radiophilus* and **Aha III** from the blue-green alga *Aphanothece halophytica.* DNA target: TTTAAA);

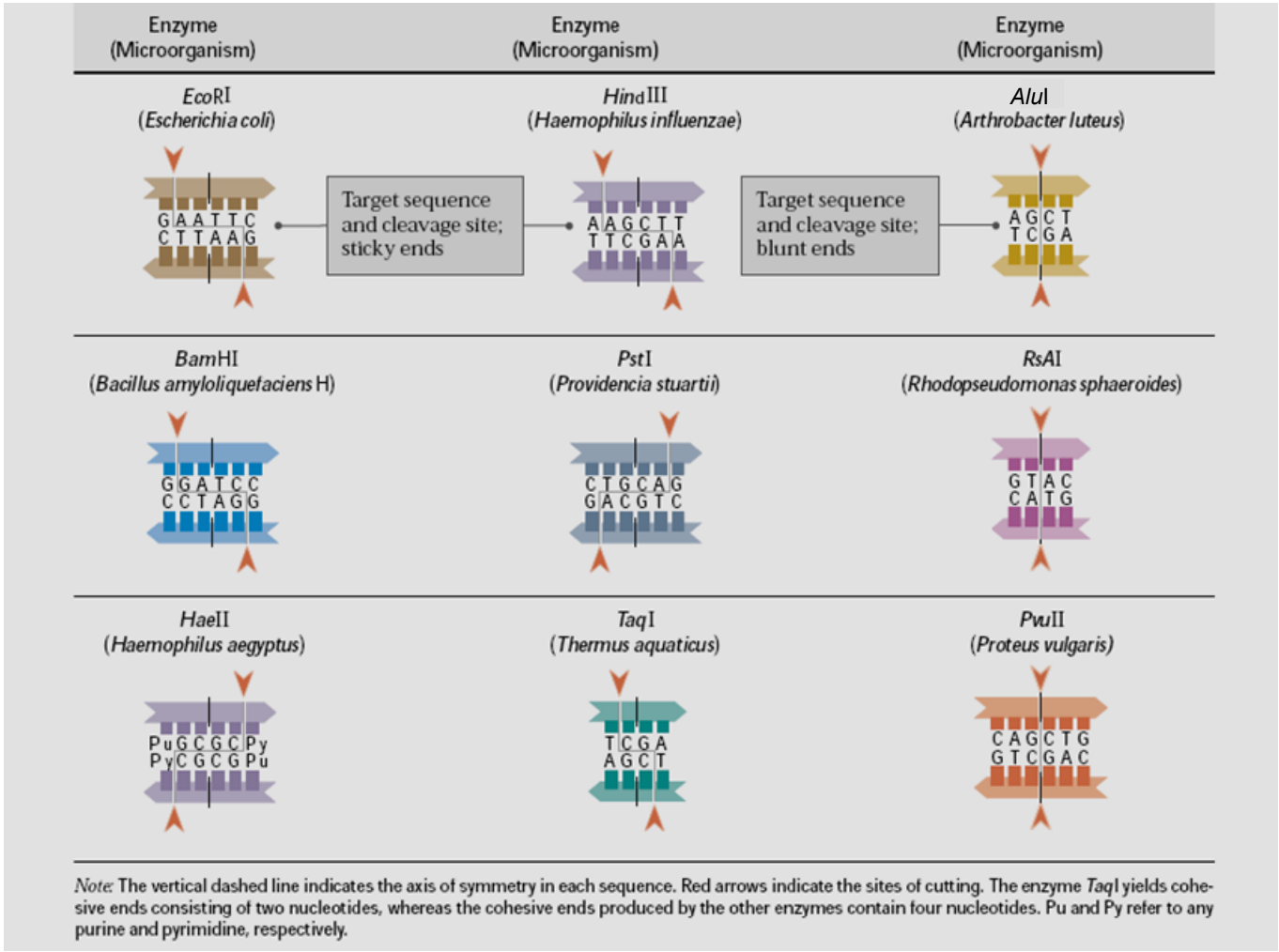
Note the nomenclature: *GENUS-species-* (Capital LETTER)-Roman numeral number

- enzymes of different specificities often occur in different isolates of the same species.

- (almost all lab. strains of *E.coli* are derivatives of wild isolates K-12 or B. They do not carry *Eco*RI or other typeII restriction systems, which were identified in other wild isolates).

Genes for R.E. are often located on the chromosome, sometimes on plasmids and very occasionally located on prophages → it appears that the genes for these enzymes shuffle between microorganisms and that there is a natural selection for variety; some bacteria contain up to 20 **different** restr.-mod. systems

Some restriction endonucleases, their sources, and their cleavage sites



**Construction of Biologically Functional Bacterial Plasmids *In Vitro***

(R factor/restriction enzyme/transformation/endonuclease/antibiotic resistance)

STANLEY N. COHEN*, ANNIE C. Y. CHANG*, HERBERT W. BOYER†, AND ROBERT B. HELLING†

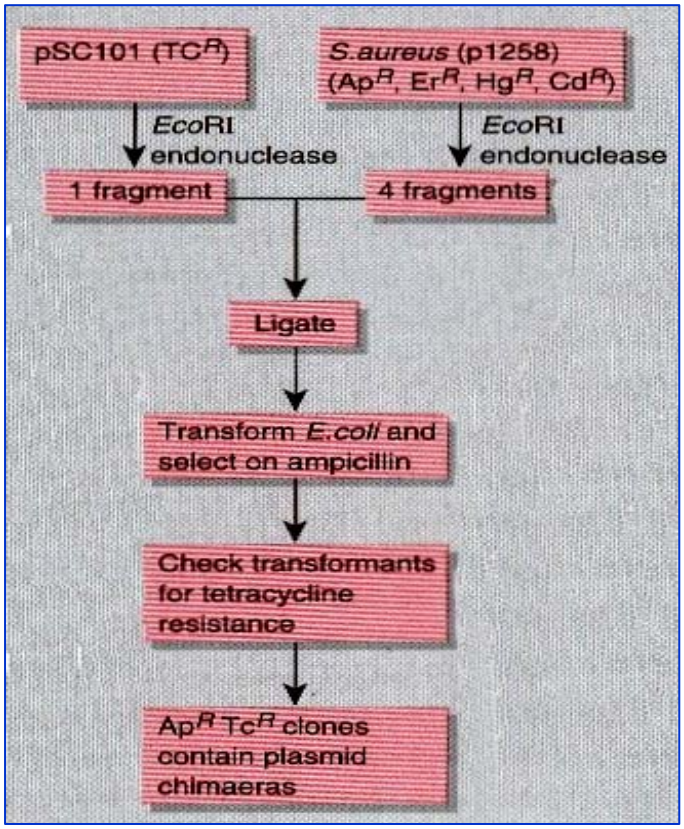
* Department of Medicine, Stanford University School of Medicine, Stanford, California 94305; and † Department of Microbiology, University of California at San Francisco, San Francisco, Calif. 94122

*Proc. Nat. Acad. Sci. USA*
Vol. 70, No. 11, pp. 3240–3244, November 1973

Using plasmid **pSC101** (containing Tetracicline resistance gene and a unique EcoRI site) they inserted exogenous DNA, derived from plasmid **p1258** of *S.aureus* ($Amp^{res}$ and 4 sites for *EcoR*I);

→ trasformation of *E.coli* (selection for $Amp^{res}$ and $Tet^{res}$ )

→ the first chimaeric molecule produced in a test tube.

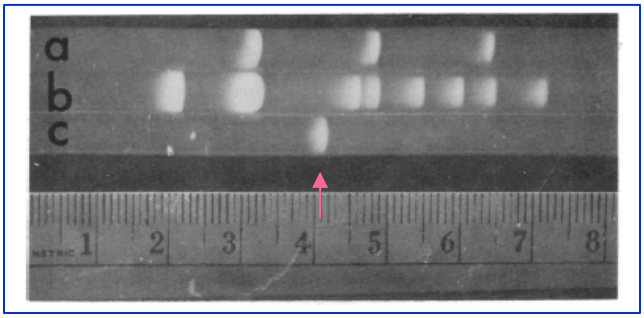


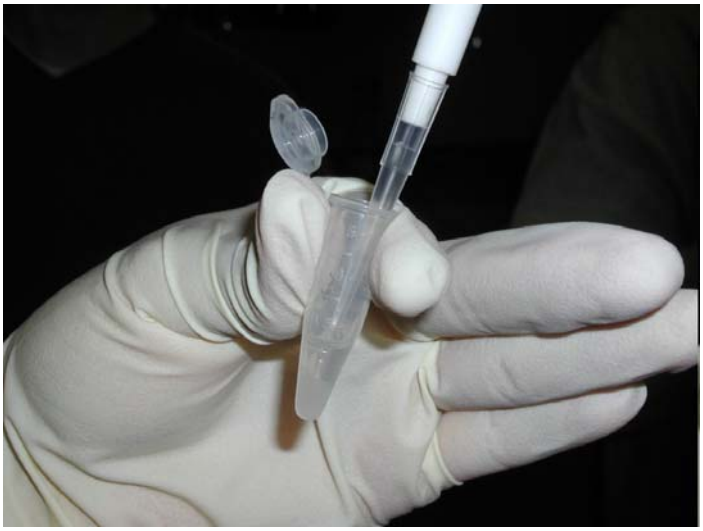

Electrophoretic separation of DNA fragments derived from plasmids:

**a) pSC102**

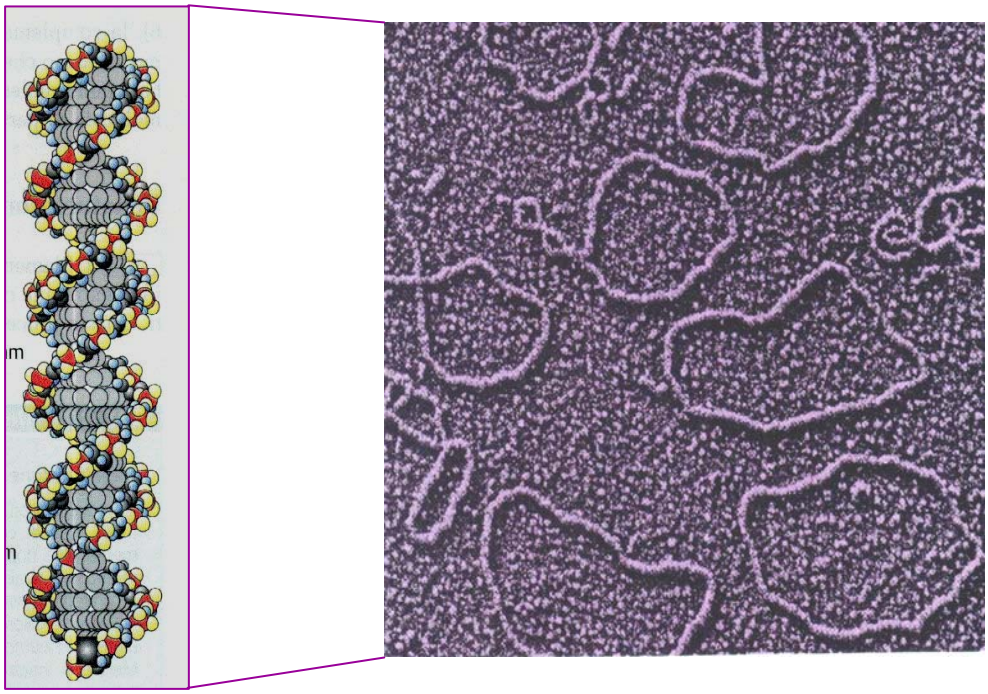
**b) R6-5**

**c) pSC101**



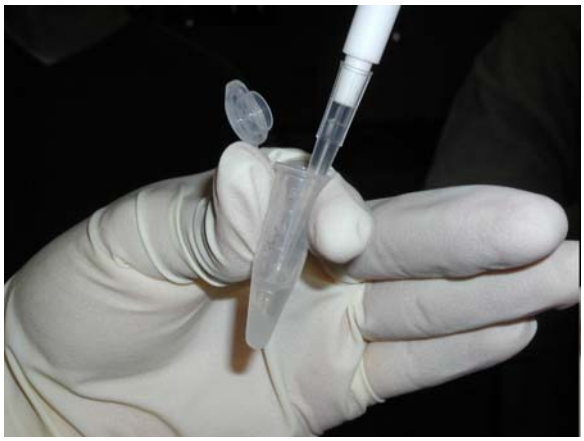


Chimaera



Plasmid **pBR322** (4361 bp)  100.000 X

1973 - Stanley Cohen/Herbert Boyer (Stanford University California)



**CURIOSITY DRIVEN PROJECT**

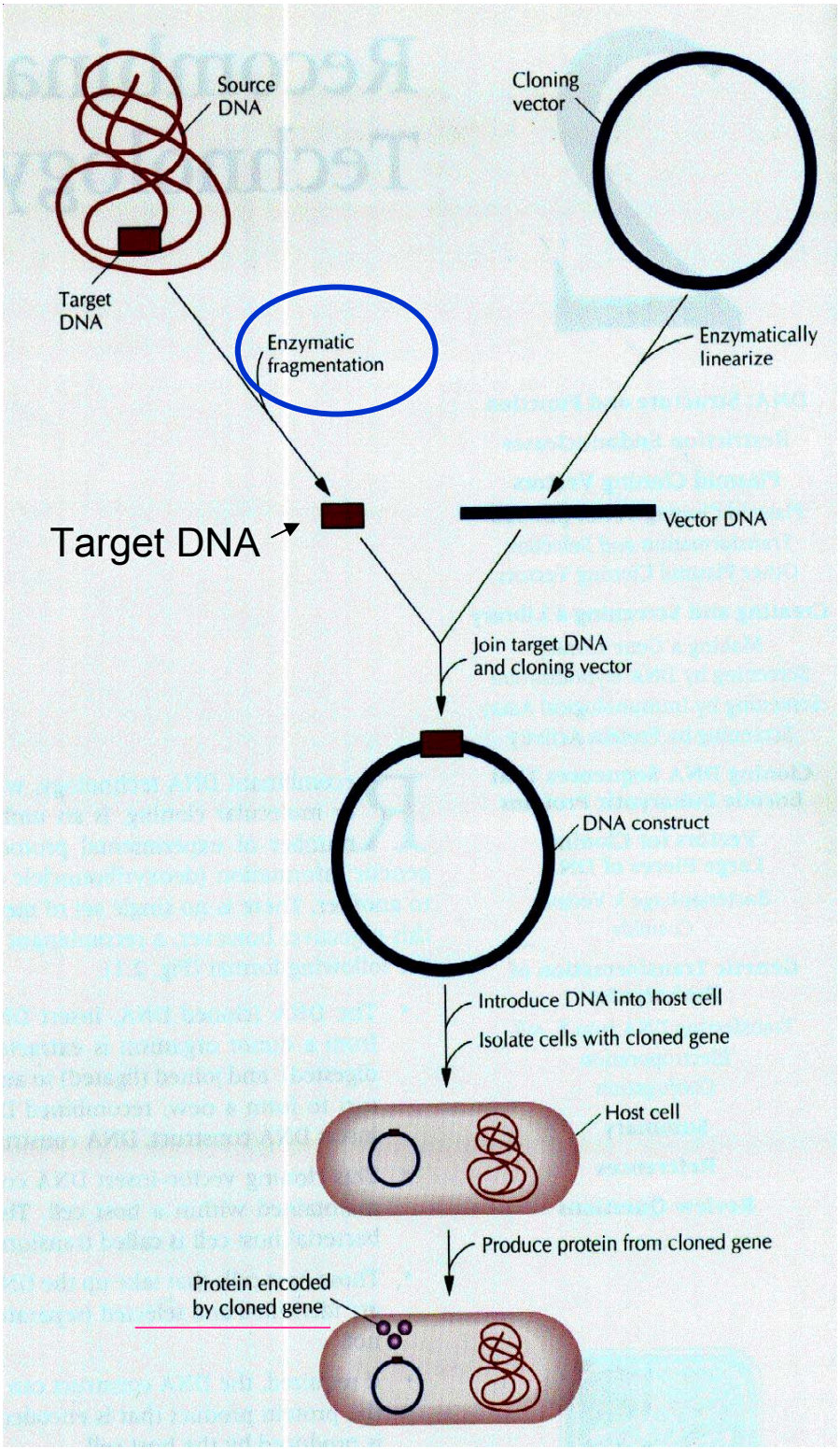
1976 R. Swanson & H. Boyer established **Genentech** (1st biotech company)

1977 Genentech produces the first human protein expressed in a microrganism (**Somatostatin**). **5 mg** of this hormone now can be easily purified from a **1 liter** bacterial culture. The same amount of hormone was obtained from 500.000 sheep' brains

1978 Human **Insulin** gene was cloned in the laboratories of Genetech

1979 Human growth hormone (**H.G.H**.) was cloned in the laboratories of Genetech

# Recombinant DNA cloning procedure




DNA from a source organism is cleaved with a restriction endonuclease and inserted into a cloning vector.

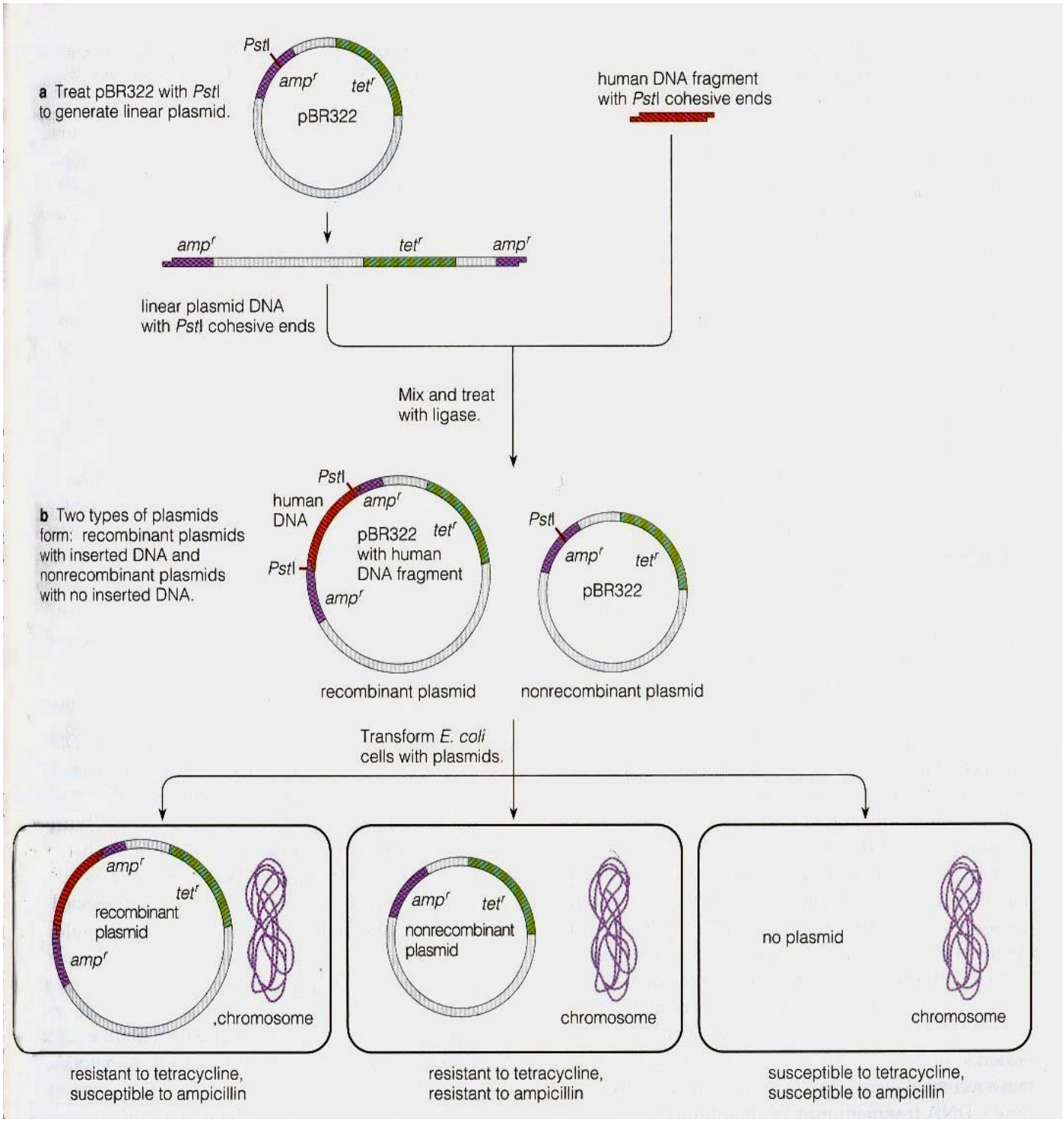
Then, the cloning vector-insert DNA construct is introduced into a target host cell and those cells that carry the construct are identified and grown.

If required, the cloned gene can be expressed in the host cell and its **protein** produced and harvested.

What is a CLONE ?

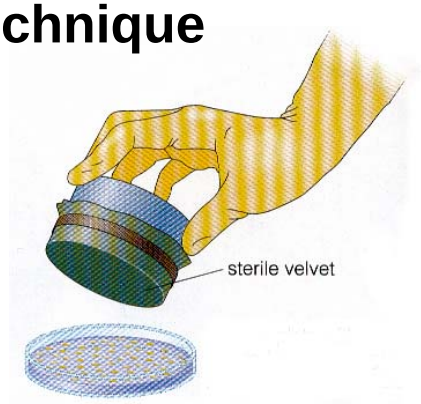
*κλων* (*klön*, twig)

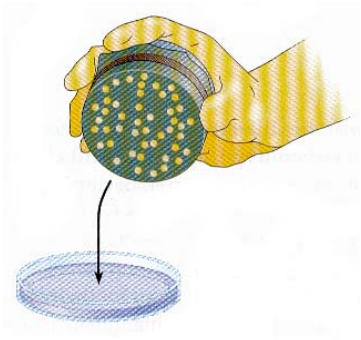
# Construction and identification of recombinant pBR322 plasmids



# Screening by replica plating technique

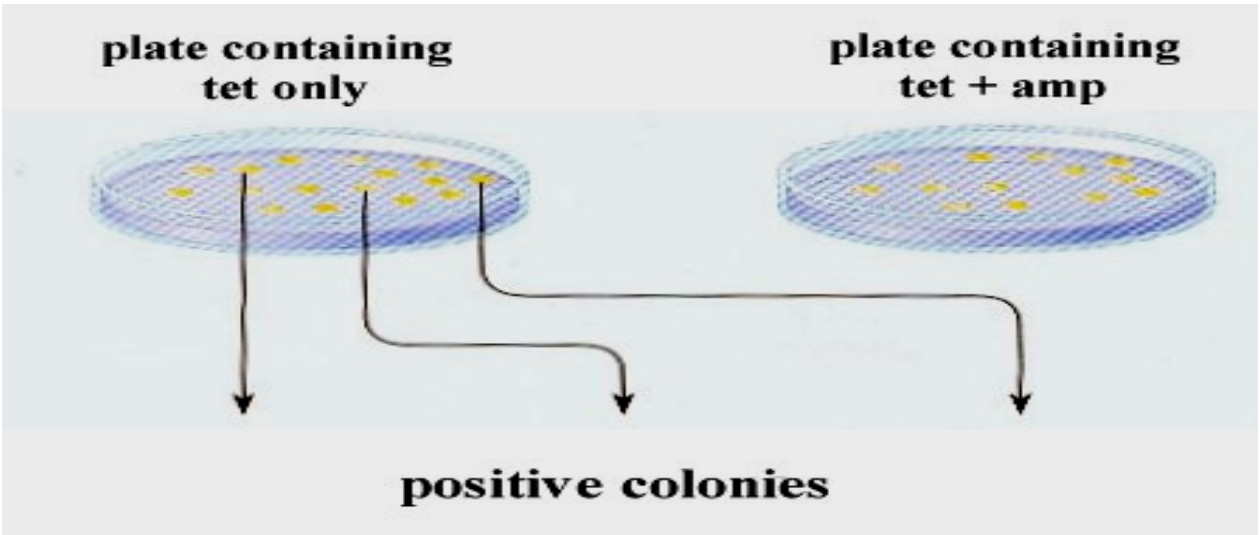
**1) Master plate containing tetracyclin only**

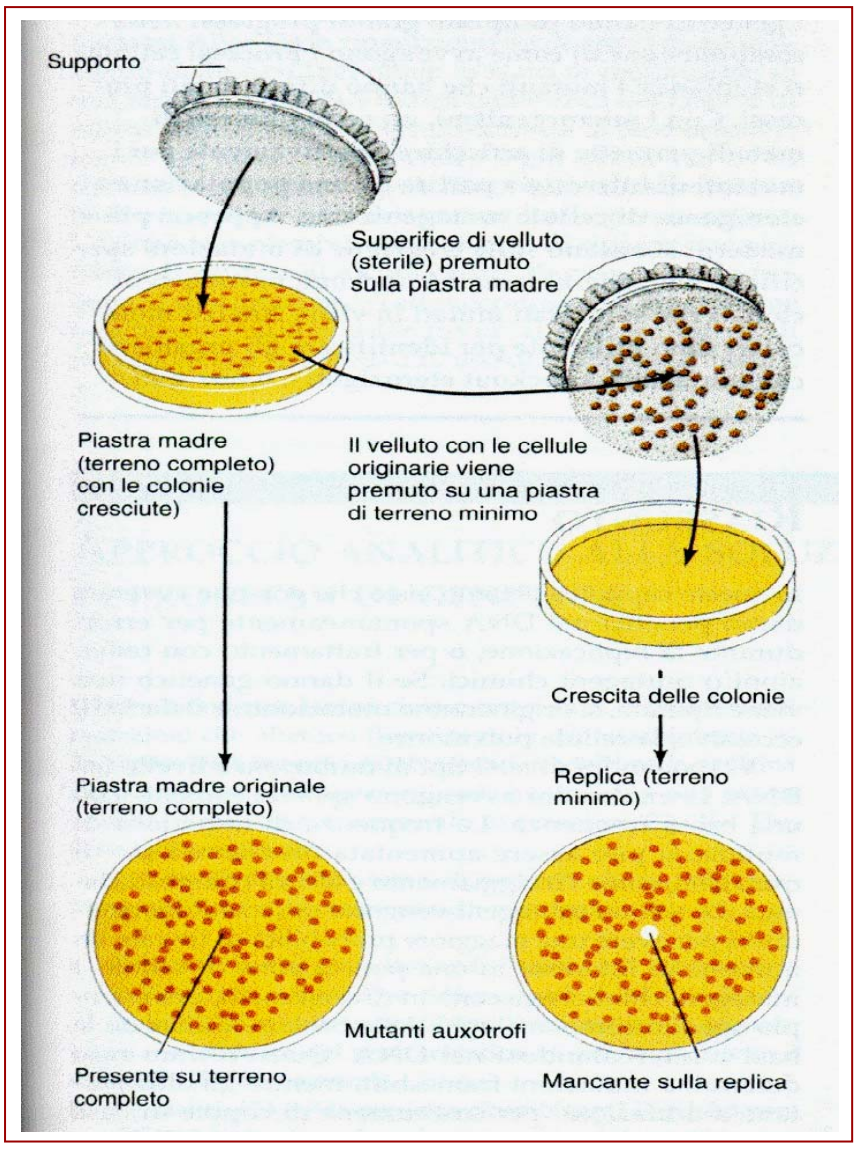


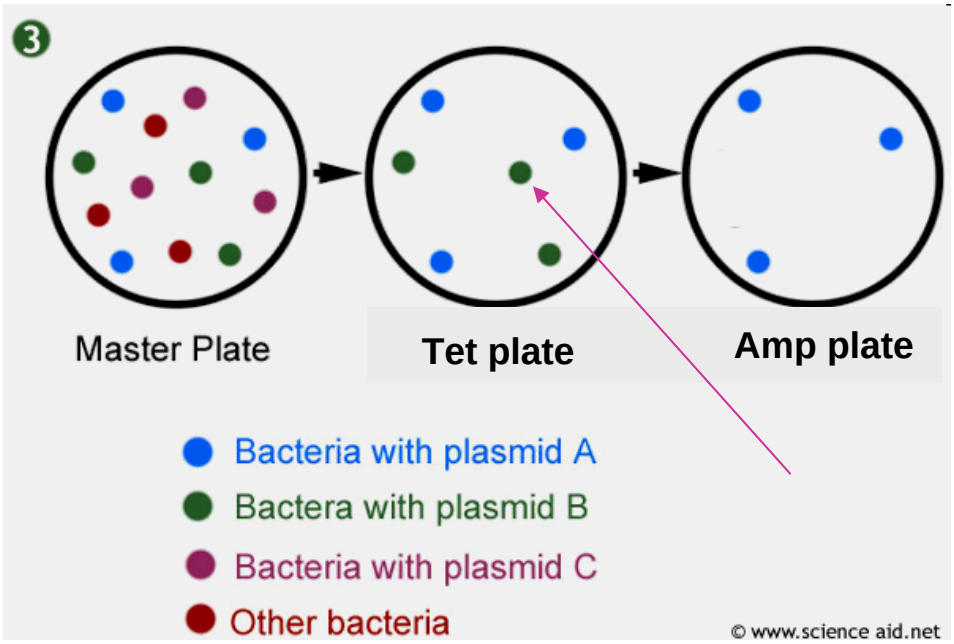



**2) Replica plating on plates containing tet + amp**

**3) Cells on the replica plate are allowed to grow into colonies. Any colony present on the master plate but missing from the replica plate carries a recombinant pBR322 with DNA inserted within amp$^r$ gene**




Supporto
Superifice di velluto (sterile) premuto sulla piastra madre
Piastra madre (terreno completo) con le colonie cresciute)
Il velluto con le cellule originarie viene premuto su una piastra di terreno minimo
Crescita delle colonie
Replica (terreno minimo)
Piastra madre originale (terreno completo)
Mutanti auxotrofi
Presente su terreno completo
Mancante sulla replica


3
Master Plate
Tet plate
Amp plate
Bacteria with plasmid A
Bactera with plasmid B
Bacteria with plasmid C
Other bacteria
© www.science aid.net

# DNA Vectors used for molecular cloning

| Vector | Approximate Size Limit of Inserted DNA (in kb) |
|---|---|
| Plasmids | 10 |
| Lambda and its derivatives | 23 |
| Cosmids | 46 |
| Bacterial artificial chromosomes (BACs) | 300 |
| Yeast artificial chromosomes (YACs) | 500 |

**Unit definition**

One **U**nit is defined as the amount of enzyme required to produce a complete digest of 1 μg of DNA in 60 min. in a reaction volume of 0.05 ml.

## Frequencies of restriction sites in DNA molecules

R.E. with shorter recognition sequences (N = 4) cut DNA more frequently than those with longer recognition sequences (N = 6 or 8).

**$4^4 = 256$** **while $4^6 = 4096$**

*e.g.* *Sau* 3A (**GATC**) cuts (¼)(¼)(¼)(¼) = once every 256 bp

*Kpn*I **G G T A C C**

1/4 1/4 1/4 1/4 1/4 1/4 = **1/4096**

*Hinc*II (**GTPyPuAC**) cuts (¼)(¼)(½)(½)(¼)(¼) = once every ~1Kb

This calculation is based on the assumption that DNA composition is **random** → 50% G::C and 50% A::T

However eukaryotic DNA has a low content of CpG dinucleotides; the sequence recognized by HpaII (**CCGG**) is represented once in **SV40 DNA** (5.2 Kbp), but there are 26 sites in plasmid **pBR322** (4.3 Kbp).

In bacteria the GC content varies from about 25% to 75% between different species → (see next slides; a short tour into Genomics)

BACTERIAL GENOMES

**Table 1**

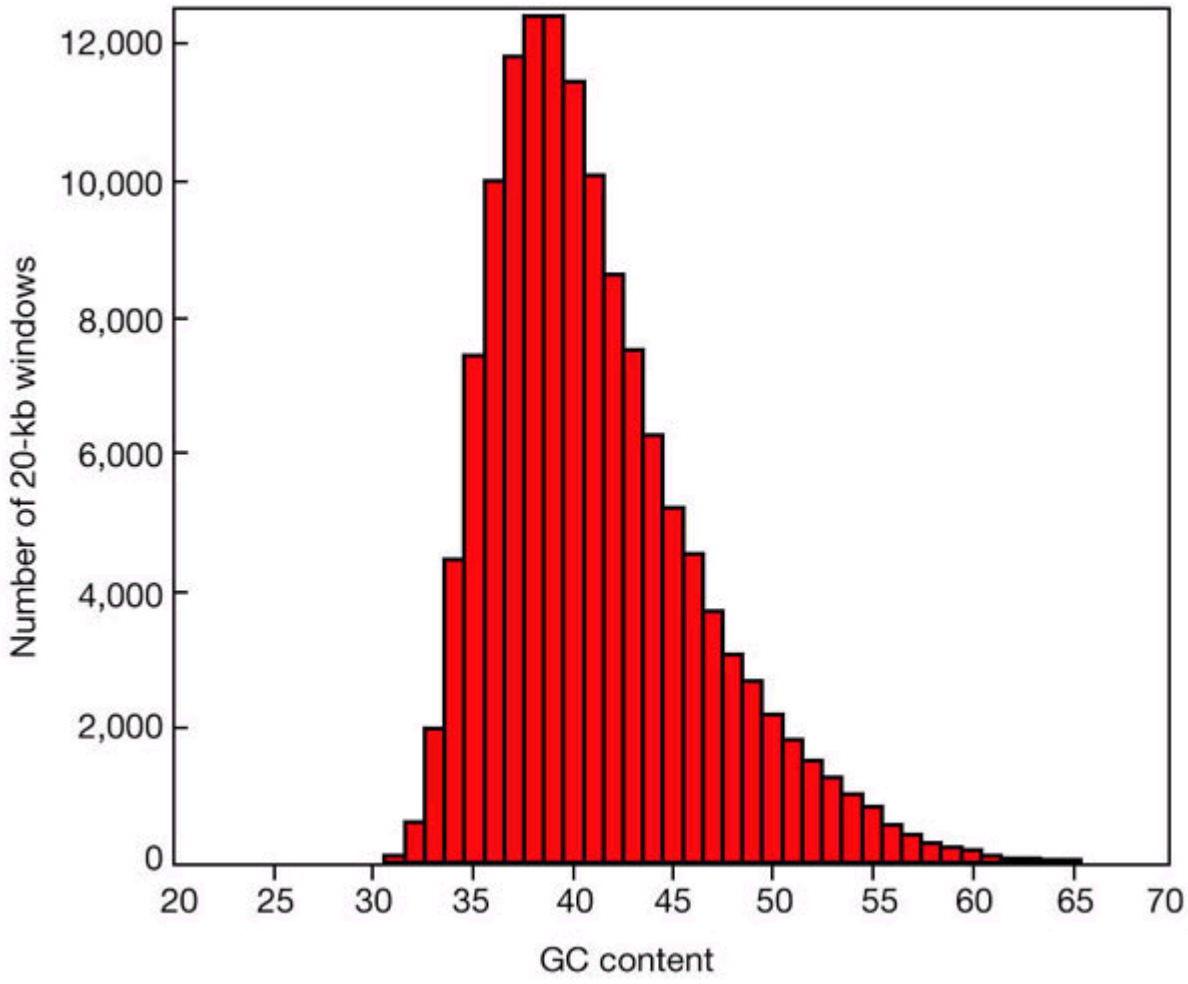
**Accuracy of RescueNet in 15 bacterial genomes.**

| Organism | GC % | Number of Genes Annotated | Training Set Size | Sn. (%) | Sn. >225 bp (%) | Sn. Conserved (%) | Sp. (%) |
|---|---|---|---|---|---|---|---|
| ***Buchnera*** | 26.2 | 564 | 292 | **88.65** | **91.24** | **89.97** | **96.18** |
| ***B. burgdorferi*** | 28.6 | 857 | 403 | **90.54** | **96.39** | **95.66** | **98.02** |
| ***C. jejuni*** | 30.6 | 1654 | 673 | **90.14** | **95.08** | **92.14** | **99.23** |
| ***M. jannaschii*** | 31.4 | 1715 | 692 | **88.39** | **91.82** | **91.02** | **96.50** |
| ***M. genitalium*** | 31.7 | 483 | 301 | **89.44** | **91.52** | **89.89** | **92.32** |
| ***H. influenzae*** | 38.0 | 1754 | 885 | **91.56** | **96.34** | **93.10** | **98.01** |
| ***H. pylori*** | 38.9 | 1593 | 712 | **91.39** | **96.80** | **95.70** | **95.49** |
| ***A. aeolicus*** | 43.3 | 1517 | 723 | **95.78** | **96.54** | **95.57** | **87.80** |
| ***B. subtilis*** | 43.5 | 4220 | 1832 | **87.93** | **94.95** | **89.86** | **89.47** |
| ***Synechocystis*** | 47.6 | 3169 | 954 | **93.18** | **96.53** | **91.55** | **90.95** |
| ***Y. pestis*** | 47.6 | 4043 | 1640 | **91.04** | **94.84** | **93.66** | **88.29** |
| ***E. coli*** | 50.8 | 4290 | 1983 | **89.39** | **92.85** | **92.54** | **89.04** |
| ***D. radiodurans*** | 67.0 | 2622 | 1436 | **84.28** | **85.65** | **92.61** | **95.50** |
| ***R. solanacearum*** | 67.0 | 3442 | 1748 | **84.74** | **88.60** | **89.82** | **93.20** |
| ***S. coelicolor*** | 72.1 | 7851 | 956 | **88.35** | **91.55** | **91.55** | **90.10** |

The genomes are listed according to ascending G+C content. For each genome, the table shows: Genome GC content (GC %), the number of genes annotated in GenBank for that genome, the number of genes in the RescueNet training set, overall RescueNet sensitivity (Sn.), the sensitivity of RescueNet in finding genes longer than the 225 bp minimum prediction size (Sn. >225 bp), the sensitivity of RescueNet in finding genes that have been confirmed by homology with other genes in GenBank (Sn. Conserved), and finally, overall RescueNet specificity (Sp.)

HUMAN GENOME

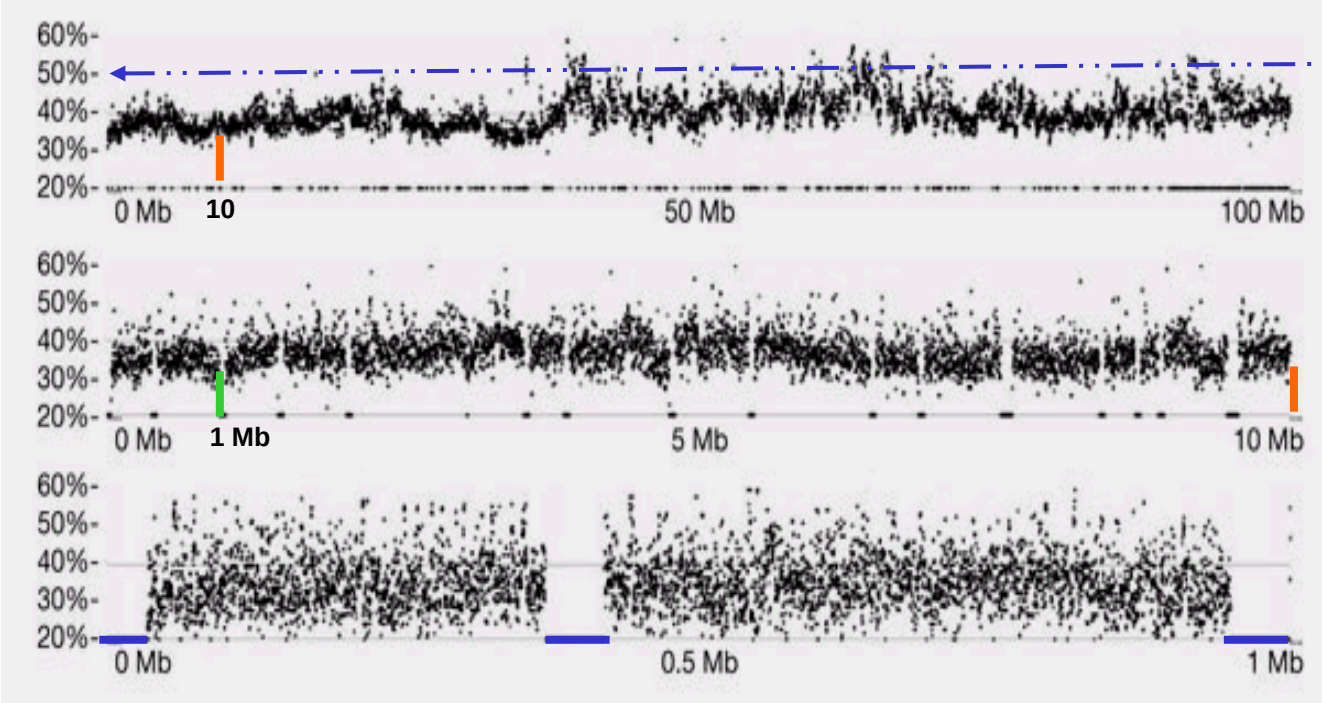
The local GC content undergoes substantial long-range excursions from its genome-wide average of 41%.




Histogram of distribution of average GC content in 20-kb windows across the draft genome sequence.

# GC content of human chromosome 1

Long-range variation in GC content is evident not just from extreme outliers, but throughout the human genome.




Variation in GC content at various scales. The GC content in subregions of a 100-Mb region of chromosome 1 is plotted. This region is AT-rich overall.

Top, the GC content of the entire 100-Mb region analysed in non-overlapping 20-kb windows.

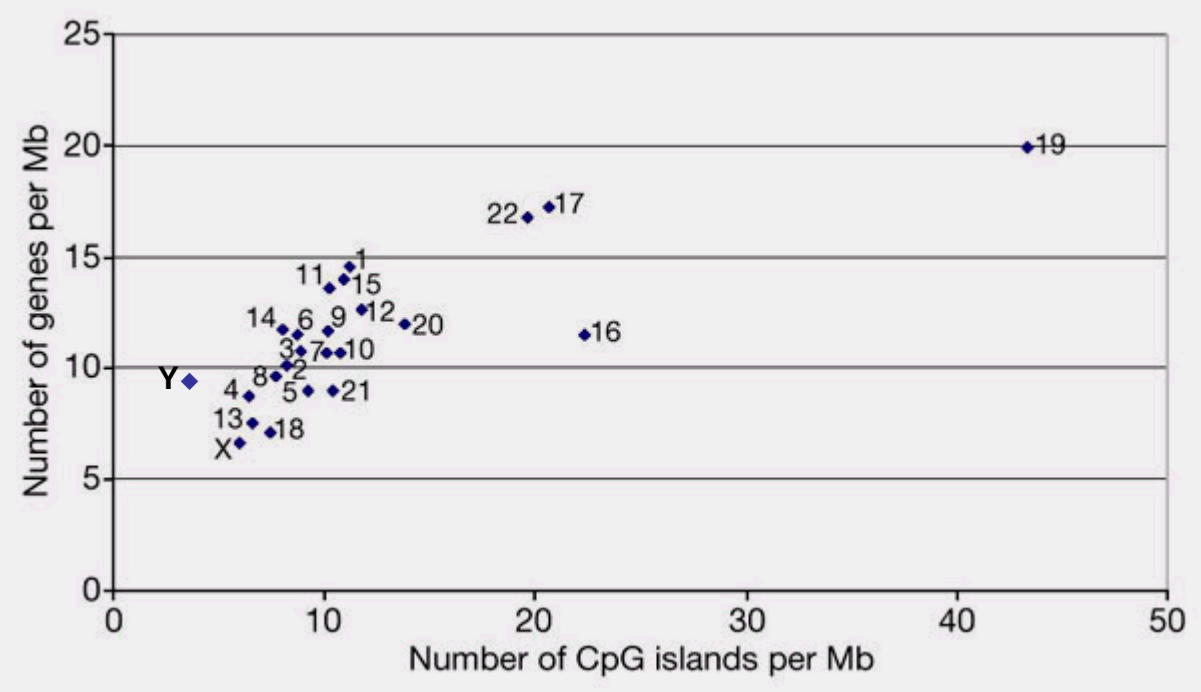
Middle, GC content of the first 10 Mb, analysed in 2-kb windows.

Bottom, GC content of the first 1 Mb, analysed in 200-bp windows. At this scale, gaps in the sequence (—) can be seen.

**CpG islands:**

- approx. 60-70% GC content
- short sequences (95% are less than 1800 bp; 75% are less than 850 bp)
- ≅ 29.000 CpG regions were found in the genome (draft version); they typically occur at or near the transcription start site of genes ;
- the density of CpG islands varies substantially among some of the human chromosomes; most chromosomes have 5–15 islands per Mb, with a mean of 10.5 islands per Mb → **less than 2%.** However, chromosome Y has an unusually low 2.9 islands per Mb, and chromosomes 16, 17 and 22 have 19–22 islands per Mb.

The extreme outlier is chromosome 19, with 43 islands per Mb.




Number of CpG islands per Mb for each chromosome, plotted against the number of genes per Mb.

Chromosomes 16, 17, 22 and particularly 19 are clear outliers, with a density of CpG islands that is even greater than would be expected from the high gene counts for these four chromosomes.

Type II R.E. interact with sequences that are **inverted repeats**, and hence symmetric (two copies of an identical sequence, which are present in a reverse orientation).

Such sequences are said to be ***palindromic***.

| Bacterial source | Enzyme abbreviation | Sequence 5′ → 3′ 3′ ← 5′ |
|---|---|---|
| *Haemophilus aegyptius* | *Hae*III | GG\|CC<br>CC\|GG |
| *Staphylococcus aureus* 3A | *Sau*3AI | \|GATC<br>CTAG\| |
| *Bacillus amyloliquefaciens* H | *Bam*HI | G\|GATC C<br>C CTAG\|G |
| *Escherichia coli* RY13 | *Eco*RI | G\|AATT C<br>C TTAA\|G |
| *Haemophilus influenzae* Rd | *Hin*dII | GTPy\|PuAC<br>CTPu\|PyTG |
| | *Hin*dIII | A\|AGCT T<br>T TCGA\|A |
| *Providencia stuartii* | *Pst*I | C TGCA\|G<br>G\|ACGT C |
| *Serratia marcescens* | *Sma*I | CCC\|GGG<br>GGG\|CCC |
| *Xanthomonas malvacearum* | *Xma*I | C\|CCGG G<br>G GGCC\|C |
| *Moraxella bovis* | *Mbo*II | GAAGAN$_8$\|<br>CTTCTN$_7$\| |

Produces blunt ends

Produce cohesive ends, with 5’ single-stranded overhang

**Pu** is any purine (A or G),
**Py** is any pyrimidine( C or T)

Produce cohesive ends, with 3’ single-stranded overhang

isoschizomers

Type II$_S$ **asymmetric sequence**

# Restriction Endonucleases - Type II enzymes

## Dra I
#129

5′ . . . T T T▼A A A . . . 3′
3′ . . . A A A▲T T T . . . 5′

2,000 units
10,000 units

**Description:** Isolated from *Deinococcus radiophilus* ATCC 27603. This enzyme is an isoschizomer of *Aha* III.

**Assay Conditions:** 10 mM NaCl, 10 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, and DNA.

**Ligation and Recutting:** After ten-fold overdigestion with *Dra* I, greater than 90% of the fragments can be ligated and recut with this enzyme.

**Concentration and Shipping:** 10,000 to 40,000 units/ml. Supplied in 50 mM KCl, 10 mM Tris-HCl (pH 7.4), 0.1 mM EDTA, 1 mM dithiothreitol, 200 µg/ml bovine serum albumin, and 50% glycerol. Store at −20°.

Continuous sequence

**Blunt-end**

---

## Dra III
#510

5′ . . . C A C N N N▼G T G . . . 3′
3′ . . . G T G▲N N N C A C . . . 5′

150 units
750 units

**Description:** Isolated from *Deinococcus radiophilus* ATCC 27603.

**Assay Conditions:** 50 mM NaCl, 10 mM Tris-HCl (pH 8.2), 10 mM $MgCl_2$, 100 µg/ml bovine serum albumin, and DNA.

**Ligation and Recutting:** After ten-fold overdigestion with *Dra* III, greater than 90% of the fragments can be ligated and recut with the enzyme.

**Concentration and Shipping:** 2,000 to 10,000 units/ml. Supplied in 50 mM KCl, 10 mM Tris-HCl (pH 7.4), 0.1 mM EDTA, 1 mM dithiothreitol, 200 µg/ml bovine serum albumin, and 50% glycerol. Store at −20°.

Interrupted sequence

**3'- OH** protruding

---

## Eae I
#508

5′ . . . Py▼G G C C Pu . . . 3′
3′ . . . Pu C C G G▲Py . . . 5′

80 units
400 units

**Description:** Isolated from *Enterobacter aerogenes* (kindly supplied by N.L. Brown).

**Assay Conditions:** 50 mM KCl, 10 mM Tris-HCl (pH 7.4), 10 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 100 µg/ml bovine serum albumin, and DNA.

**Ligation and Recutting:** After ten-fold overdigestion with *Eae* I, 100% of the DNA fragments can be ligated and recut with this enzyme.

**Concentration and Shipping:** 1,000 to 5,000 units/ml. Supplied in 50 mM KCl, 10 mM Tris-HCl (pH 7.4), 0.1 mM EDTA, 1 mM dithiothreitol, 200 µg/ml bovine serum albumin, and 50% glycerol.

→ design a palyndrome

---

**Replaces *Xma* III**

## Eag I
#505

5′ . . . C▼G G C C G . . . 3′
3′ . . . G C C G G▲C . . . 5′

100 units
500 units

**Description:** Isolated from *Enterobacter agglomerans* (NEB #368). This enzyme is an isoschizomer of *Xma* III.

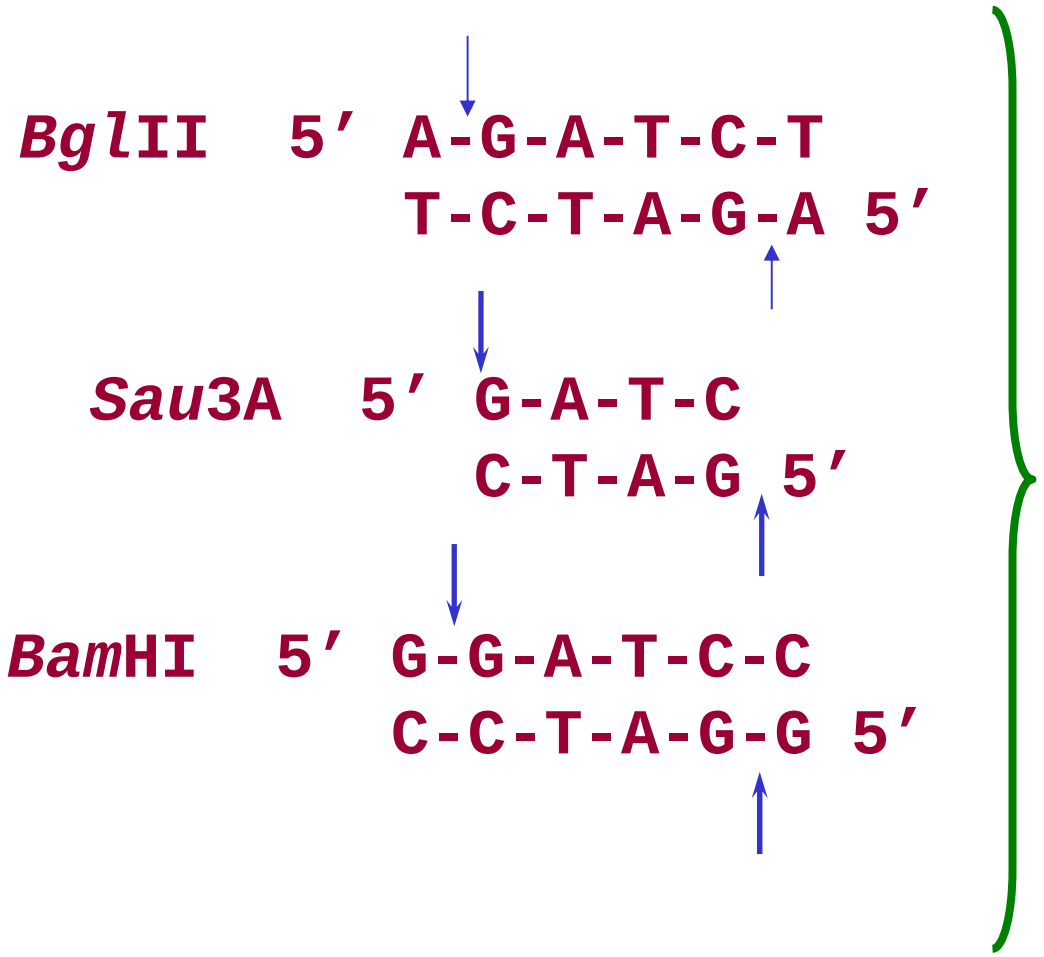
**Assay Conditions:** 150 mM NaCl, 10 mM Tris-HCl (pH 8.2), 10 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 100 µg/ml bovine serum albumin, and DNA.

**Ligation and Recutting:** After several-fold overdigestion with *Eag* I, greater than 90% of the DNA fragments can be ligated and then recut with this enzyme.

**Concentration and Shipping:** 1,000 to 4,000 units/ml. Supplied in 100 mM KCl, 20 mM Tris-HCl (pH 7.4), 0.1 mM EDTA, 1 mM dithiothreitol, 200 µg/ml bovine serum albumin, and 50% glycerol. Store at −20°.

**?** protruding

**All these sticky ends are compatible**




This feature can be very useful for DNA manipulation

## Modification of restriction sites by DNA methylation

Methylases can be used to alter the apparent cleavage specificity of certain restriction enzymes.

These alterations are accomplished *in vitro* by methylation of a subset of the sequences recognized by certain restriction enzymes.

**A)** HincII recognizes the degenerate sequence **G-T-Py-Pu-A-C** and will therefore cleave the four sequences:

**G T C G A C (1)**

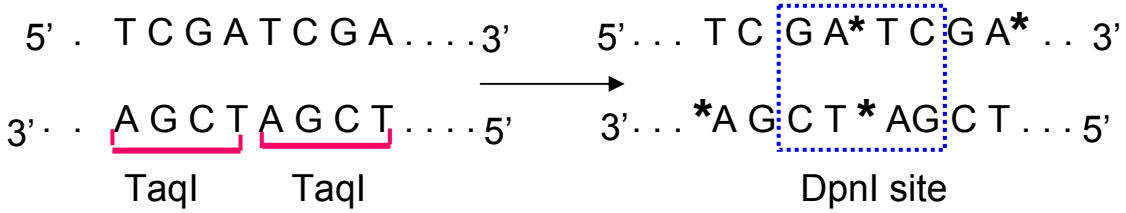
**G T C A A C (2)**

**G T T G A C (3)**

**G T T A A C (4)**

The M. *TaqI* methylase recognizes only the sequence **T C G A**. The sequences of the subset (2) (3) and (4) will remain sensitive to HincII, while those sequences containing the internal T C G A will be resistant to cleavage after methylation with M. *TaqI*.

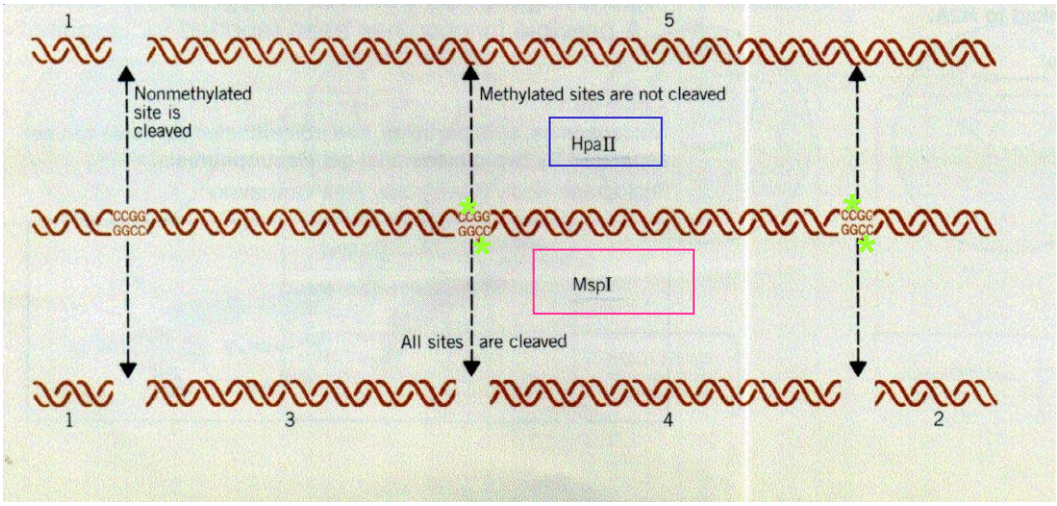
**B)** Use of M. *Taq*I adenine methylase in conjunction with the methylation dependent restriction enzyme *Dpn*I to produce highly specific cleavages




This procedure creates a site in DNA that otherwise would not be cleaved by *Dpn*I. This highly specific cleavage site is **8 bp** long.

The distribution of methyl groups can be examined by taking advantage of restriction enzymes that cleave target sites containing the C-G doublet.

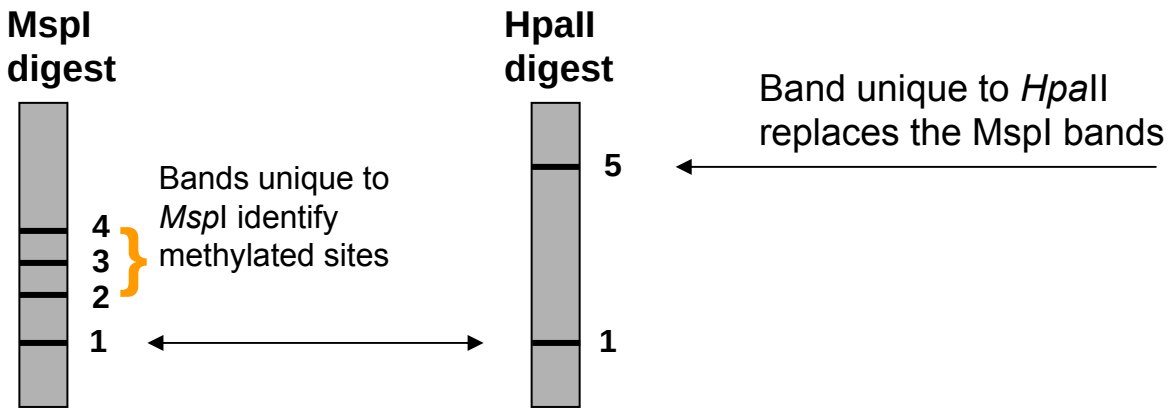
*Msp*I and *Hpa*II are isoschizomers that cleave the same target sequence in DNA ( 5' **CCGG** 3'), but have a different response to its state of methylation




With non-methylated DNA the two enzymes would generate the same restriction bands, but the positions in DNA that are methylated (*) at the sequence 5' CC*GG 3' are not cleaved by *Hpa*II.

[*Moraxella sp*. methylates its own DNA at the 5' C of the sequence 5' C*CGG 3']

**The results of *Msp*I and *Hpa*II cleavage are compared by gel electrophoresis**




Band at same position identifies non-methylated site (1)

## Type III

These enzymes are relatively rare and do not provide endonucleases for gene manipulation. They act as complexes of two subunits (M subunit and R subunit). The recognition sites are **asymmetric** and cleavage occurs by nicking **one** strand at a measured distance to one side of the recognition sequence. Therefore **two sites** in opposite orientations are necessary to break the DNA duplex (e.g. *Eco*P1 and *Eco*P15I).

EcoP15I recognizes the non-symmetric DNA sequence 5'-CAGCAG

(this enzyme is used in SAGE procedure).

Other restriction systems are known which fall outside the Type I-II-III classification: **Mcr systems** (***m***odified ***c***ytosine ***r***estriction), and Mrr systems. These systems are sequence specific and attack DNA **only** when it is methylated at specific positions (e.g. do not attack DNA modified at *dcm* sites)

***Mcr*A** is encoded by a prophage-like element ($^{m5}$CG);
***Mcr*BC** is encoded by two genes, *mcr*B and *mcr*C.
Recognition site = R $^{m5}$C ($N_{40-80}$) R $^{m5}$C where R= A or G.
The cleavage requires GTP and occurs at multiple sites in both strands between the methyl-cytosines.
***Mrr*** restricts DNA modified by a variety of adenine methylases.

For cloning in *E.coli* it is therefore wise to use a strain that lacks the three familiar restriction systems (EcoK, EcoB and EcoP1) and also the Mcr and Mrr systems.

Most strains of *E.coli* contain two site specific **DNA methylases**:

**dam** → transfers a methyl group from SAM to the $N_6$ position of the **adenine** residue in the sequence 5' **G A T C 3'.** This system distinguishes the strands of newly replicated DNA by methylating adenines. It is involved in control of replication and in marking DNA strands for repair.

**dcm** → methylates the internal **cytosine** in the sequence
5' **C C A G G** 3' or 5' **C C T G G** 3'

Some **R.E**. will not cleave DNA methylated by either dam or dcm: ClaI G**ATCGAT** MboI **GATC**
XbaI **TCTAGA**TC ScrFI **CC(AT)GG**
StuI **AGGCCT**GG

Other **R.E**. whose recognition sequence are identical to or overlap the dam or dcm methylase sites are refractory:
BamHI **GGATCC** Bgl II **AGATCT**
BstNI **CC(AT)GG** Sau3AI **GATC**

## Homing Endonucleases

Double stranded DNases that have large, **asymmetric** recognition sites (12-40 bp) and coding sequences that are usually embedded in either introns or inteins.

Recognition sites are extremely rare; however, unlike standard restriction endonucleases, homing endonucleases tolerate some sequence degeneracy.

I-SceI

(Intron-encoded endonuclease present in the mitochondria of *Saccharomyces cerevisiae*.

The homing site (18nt) is:

5'….TAGGGATAA▼CAGGGTAAT ….3'

3' ….ATCCC▲TATTGTCCCATTA …..5'

This sequence is one site that is known to be recognized and cleaved.

## Nicking Endonucleases

These enzymes are altered restriction enzymes that hydrolyze only one strand of the DNA duplex, to produce molecules that are "nicked", rather than cleaved.

These nicks can serve as initiation points for a variety of enzymatic reactions (replacement DNA synthesis, exonucleolytic degradation or the creation of small gaps).

N. BbvC IA (it nicks by virtue of its **inability to form dimers**)

5'…GC▼TGAGG…3'

3'…CGACTCC….5'

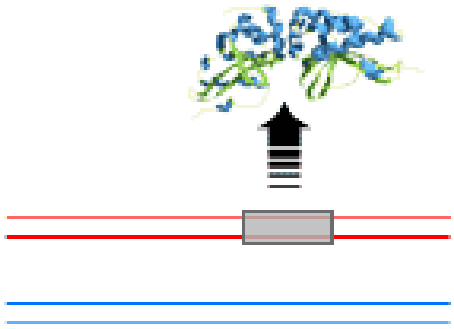
Their activities are monitored by conversion of supercoiled plasmid DNA to open circles.

## Homing endonucleases spreading [Mobile Genetic Elements]

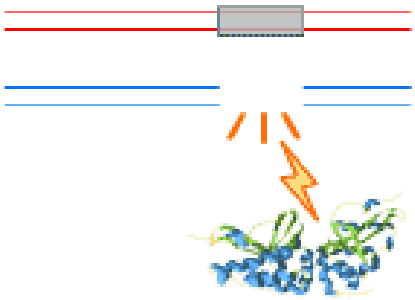
When an intron or intein containing gene meets an intron or intein-free copy of the same gene.

Intron or intein (**grey box**) encodes an homing endonuclease

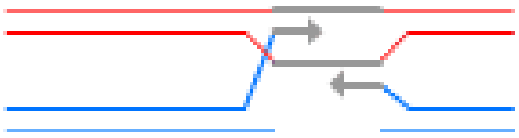
Intron or intein-free gene



Homing endonuclease cuts the intron or intein-free gene



Homologous recombination with intron or intein-containing gene occurs.
This process is referred to as a "gene conversion" event.



Outcome : the intron or intein is now found in both genes.

# R F L P

## Restriction Fragment Length Polymorphysm

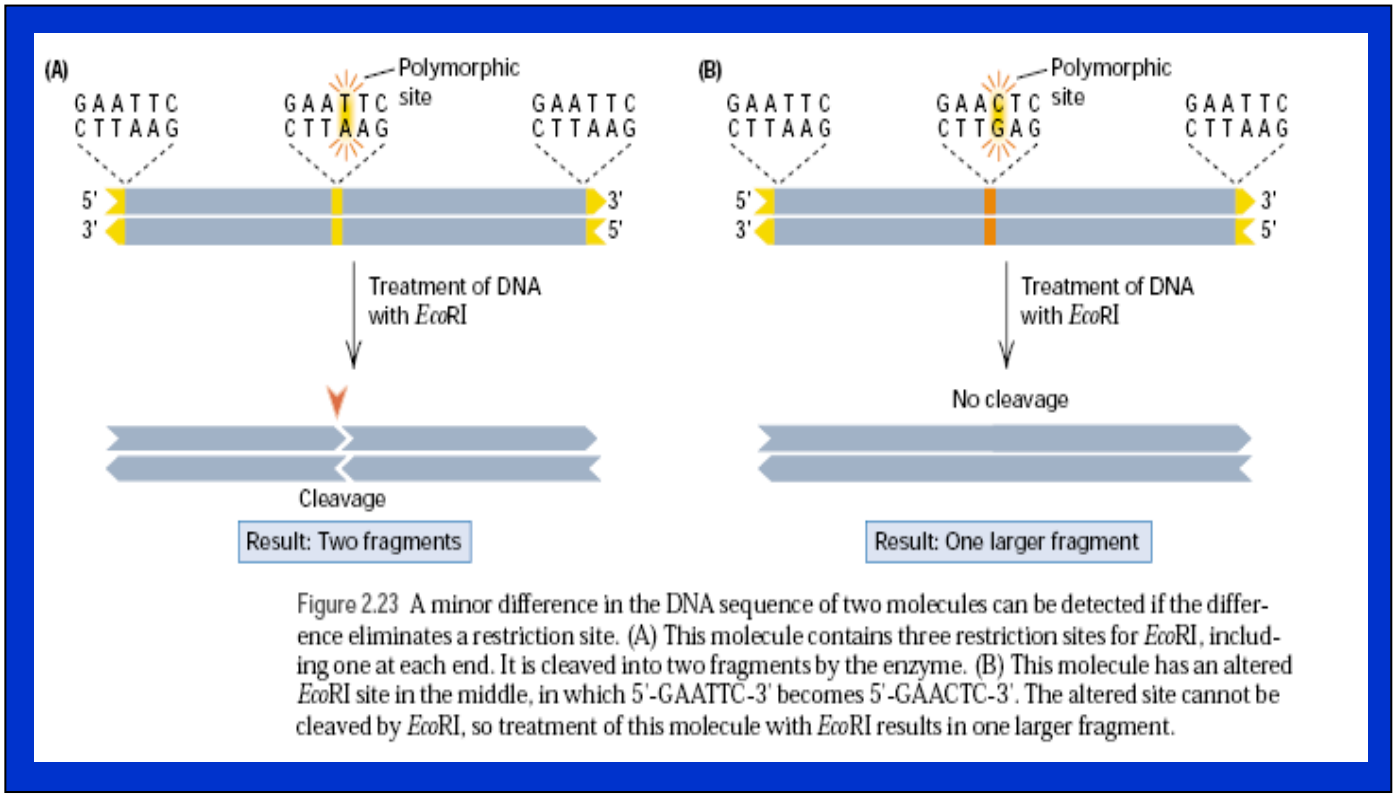
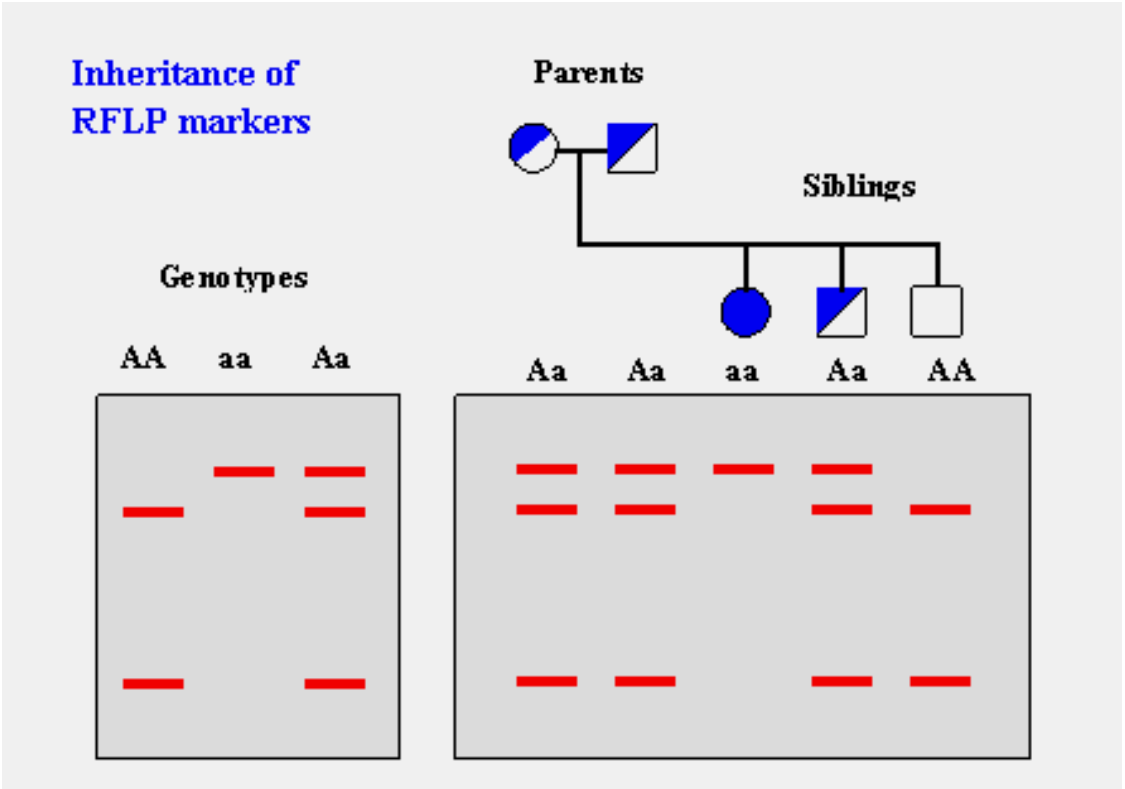
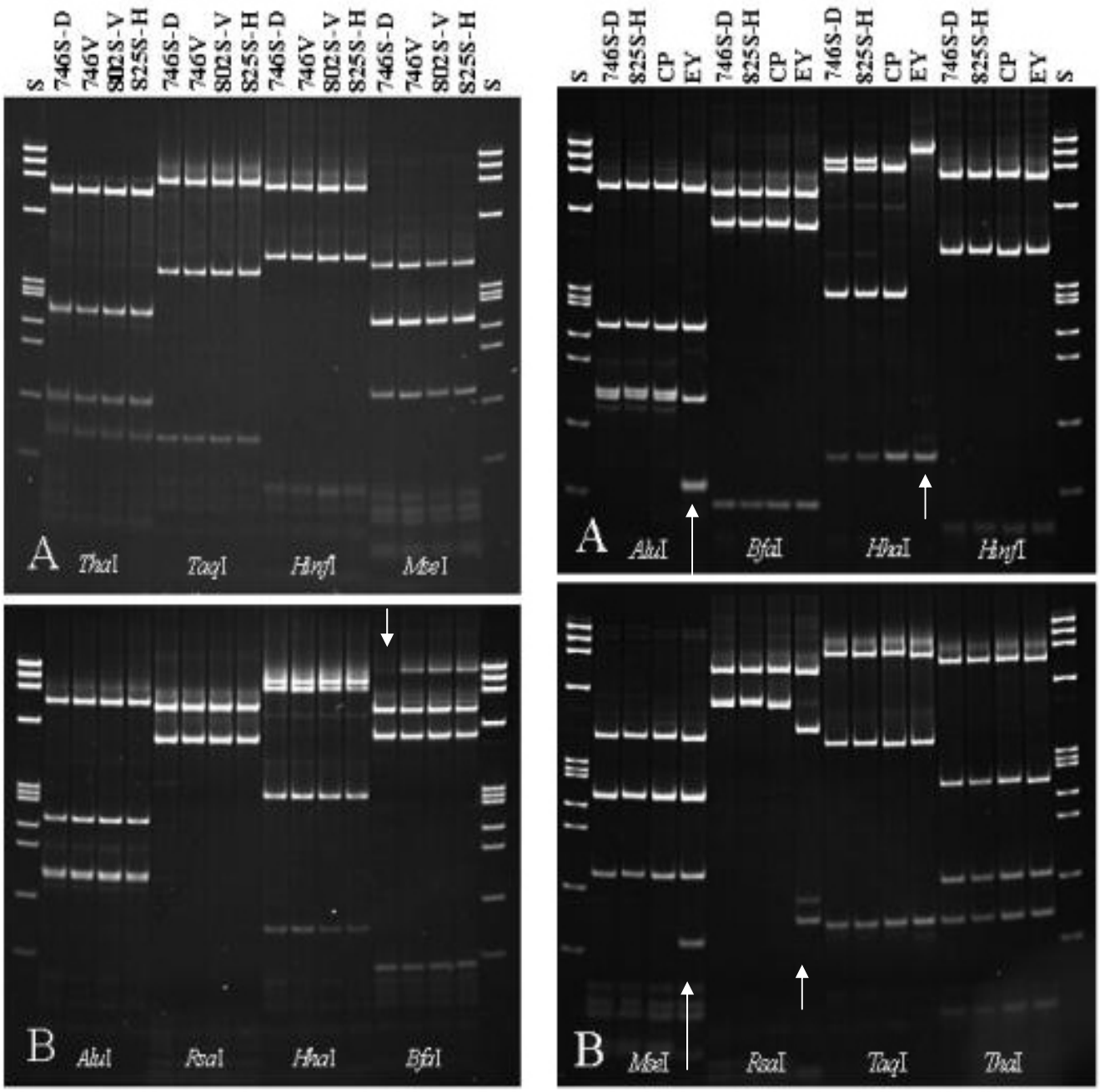



Figure 2.23 A minor difference in the DNA sequence of two molecules can be detected if the difference eliminates a restriction site. (A) This molecule contains three restriction sites for *Eco*RI, including one at each end. It is cleaved into two fragments by the enzyme. (B) This molecule has an altered *Eco*RI site in the middle, in which 5'-GAATTC-3' becomes 5'-GAACTC-3'. The altered site cannot be cleaved by *Eco*RI, so treatment of this molecule with *Eco*RI results in one larger fragment.






**RFLP (restriction fragment length polymorphism)** analysis of PCR-amplified 16S rDNA sequences, digested with restriction enzymes.

Different phytoplasma strains causing diseases in plants were analysed by this technique.