Topological Linking Number (Lk): Twist (Tw) + Writhe (Wr)

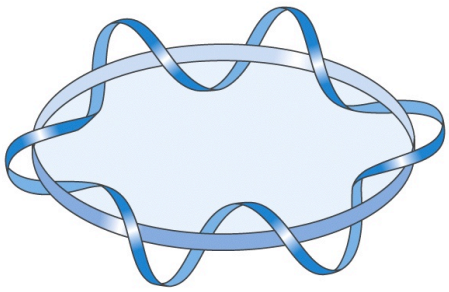
Tw = "the number of times Watson strand crosses Crick"

(the number of turns of B-form DNA: right-handed turns = +)

Wr = "the number of times WC duplex crosses WC duplex"

(supercoiling: compaction of B-form DNA against itself;

right handed supercoils are +, left handed supercoils are -)



Lk = 6

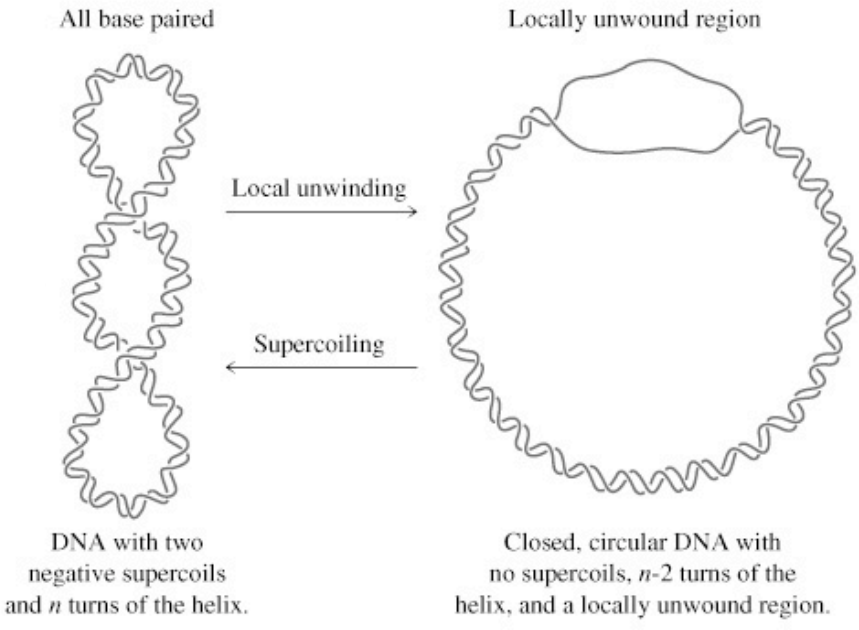
Lk = 6

Tw = 6

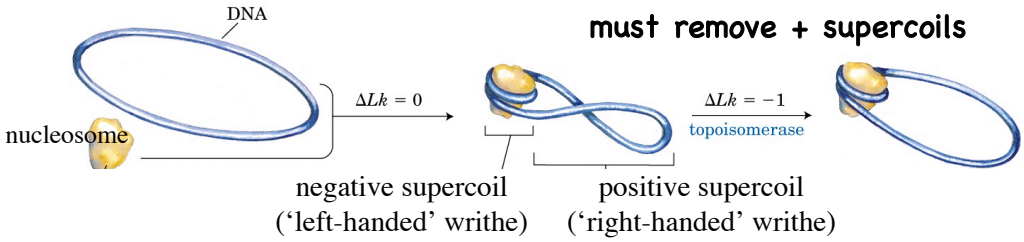
Wr = 0

Lk_o = most favorable Lk = B-form DNA with no supercoils

DNA in cells is negatively supercoiled. Twist and writhe can interconvert, but linking number cannot without breaking the phosphodiester backbone.



DNA packaging introduces a topological problem



Tw = 6	no enzyme is required	**Tw = 6**	**enzyme is required!**	**Tw = 6**
Wr = 0	⇌	**Wr = -1 +1**	→	**Wr = -1**
Lk = 6		**Lk = 6**		**Lk = 5**

"topoisomers" differ in Lk; Lk is altered by "topoisomerase" enzymes

Two kinds of topoisomerases

Type I: cuts one strand

Changes Lk in steps of 1.

Enzyme nicks one strand and rotates it around the other strand.

Type I topoisomerases can act to change Lk only towards Lk_o.

Type II: cuts two strands

Changes Lk in steps of 2.

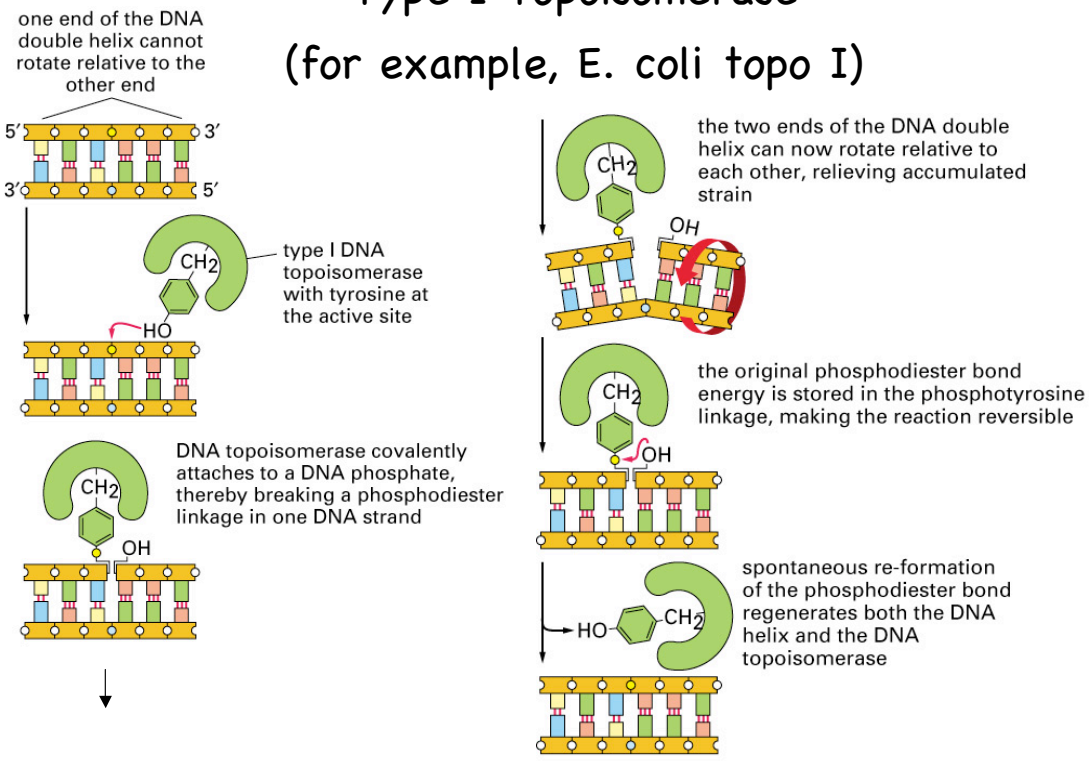
Enzyme breaks one duplex and passes another duplex through the break.

Type II topoisomerases can bias change in Lk in any direction using ATP hydrolysis to drive a conformational change in the enzyme.

BOTH types form a covalent protein-DNA intermediate at a strand break to store phosphodiester bond energy. Unlike restriction enzymes, cleavage is reversible and strands join without DNA ligase.

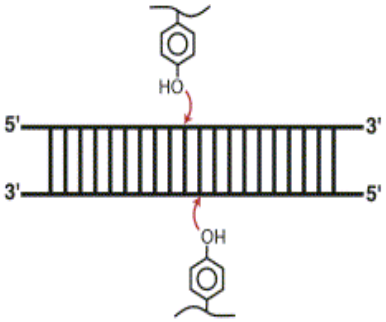
Type I Topoisomerase

(for example, E. coli topo I)

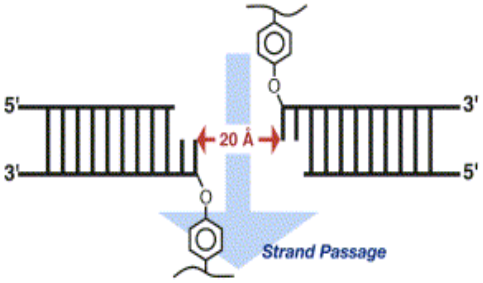


A topoisomerase II dimer makes a DNA break

Tyrosine OH attacks PO4 and forms a covalent intermediate



Structural changes in the protein open the gap by 20 Å!



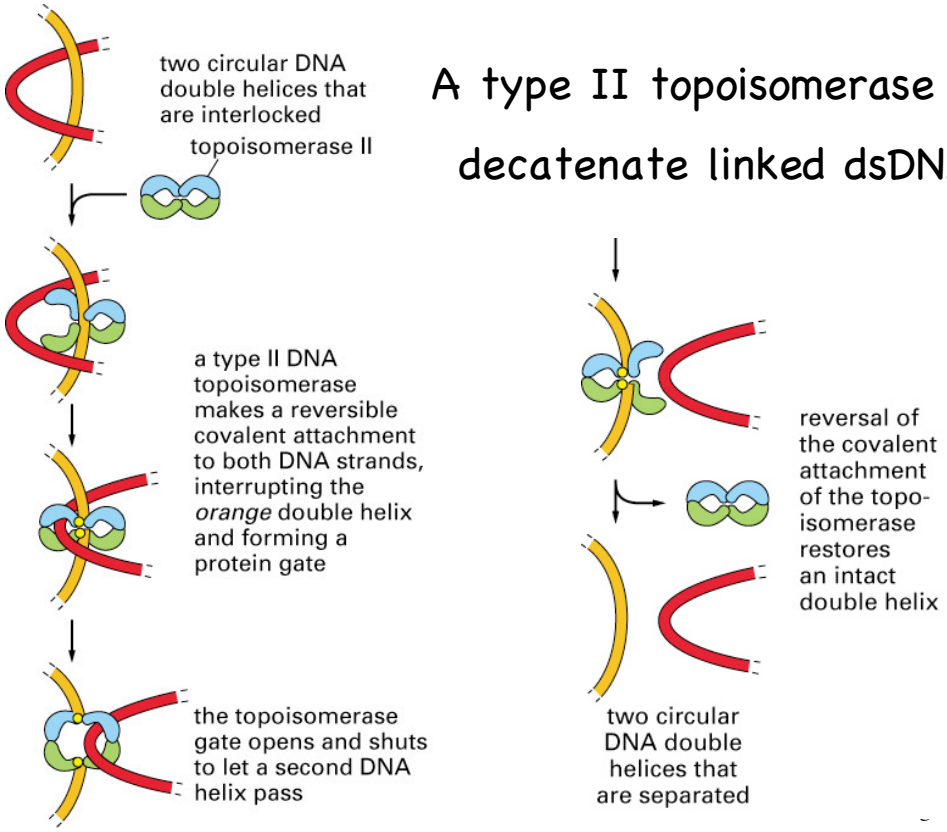
Type II Topoisomerase

(for example, E. coli DNA gyrase)

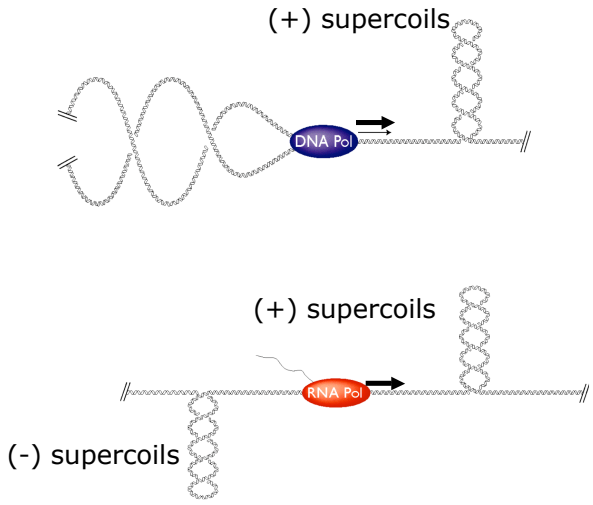
E. coli gyrase changes Lk in steps of -2: it binds to a + supercoil and converts it to - supercoil.

Gyrase is the target of several anti-bacterial drugs

A type II topoisomerase can decatenate linked dsDNAs



DNA topology changes and needs to be changed on a continuous basis

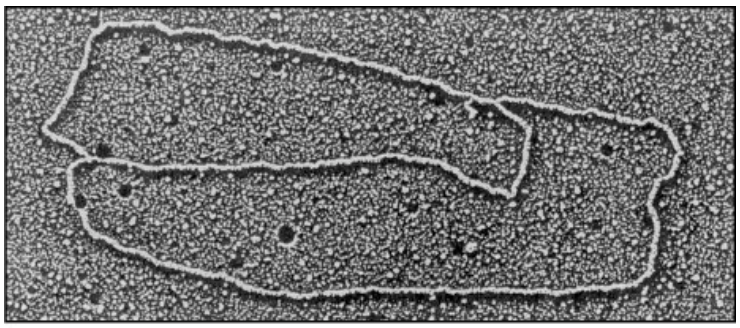


Once genome replication starts, it should finish (otherwise some genes are amplified).

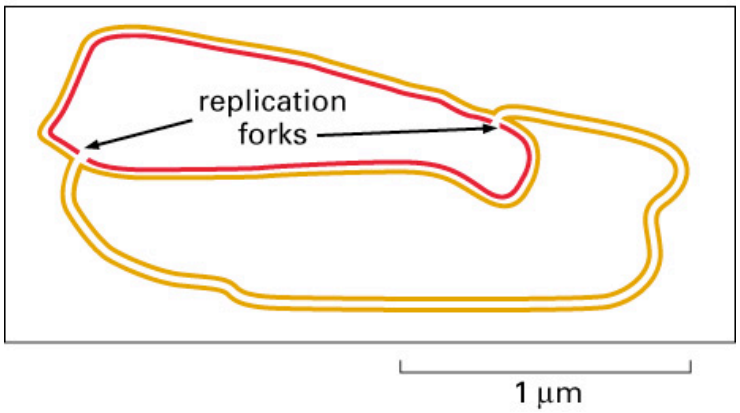
As a result, the **initiation** of DNA replication is highly regulated.

Initiation requires building replication forks.

electron micrograph of replicating DNA circle

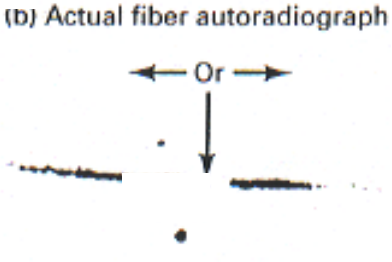


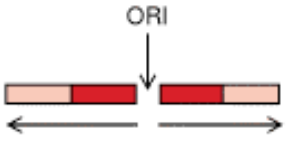
one possible interpretation of events (the correct one)



Bidirectional movement of replication machinery from the site of replication initiation

If replicating cells are exposed to a pulse of [^{3}H]thymidine, the resulting DNA will be heavily labeled ("hot") where early replication occurs and lightly labeled farther away. When such labeled DNA is dried on a microscope slide as long fibers and exposed to a radiation-sensitive emulsion, autoradiographic signals should be produced proportional to the hot-ness of the DNA.

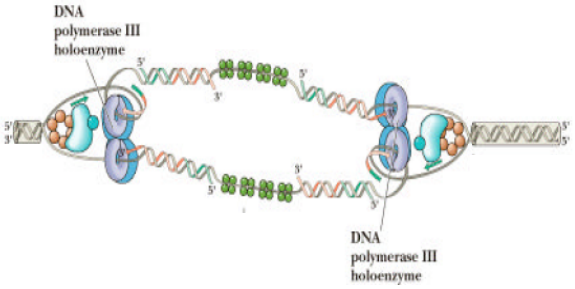


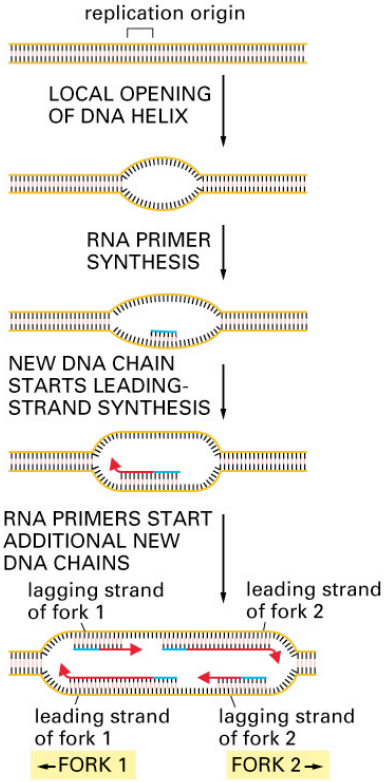


ORI = an origin of replication

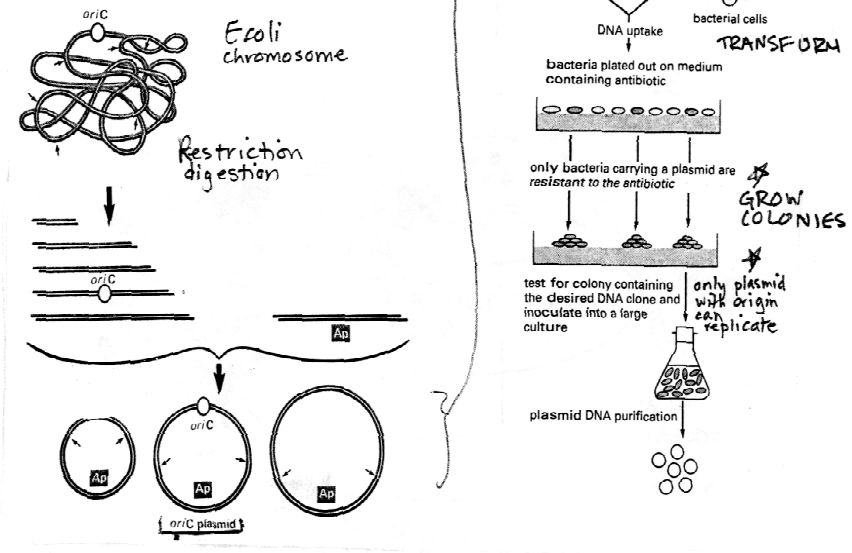
To have bidirectional replication from an origin, leading strand synthesis must be started twice.

The origin "firing" event needs to result in loading of two DnaB hexamers, one on each strand.



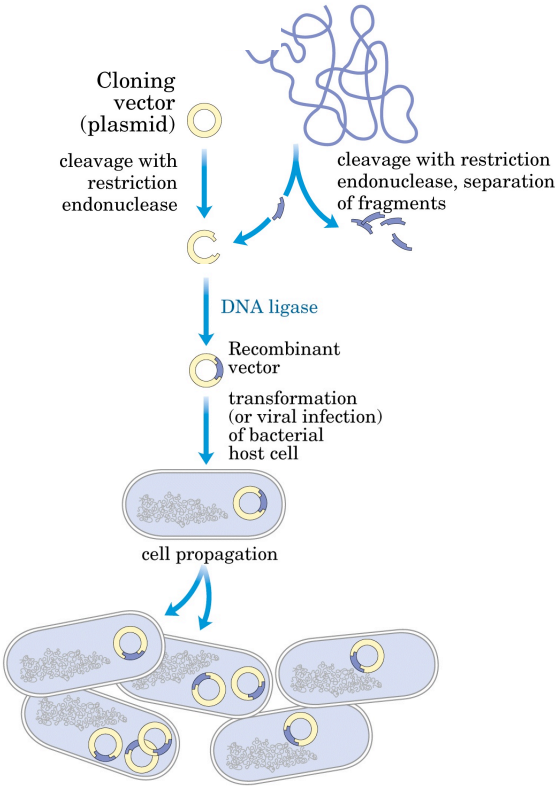


Finding E. coli *oriC*



join each E.coli piece of DNA separately to Amp^R gene

all colonies have the same piece of E.coli DNA: oriC



A replication origin and resistance marker allow DNA cloning

A DNA fragment of interest (or a pool of fragments) is ligated into a autonomously replicating vector, then introduced into a bacterial host.

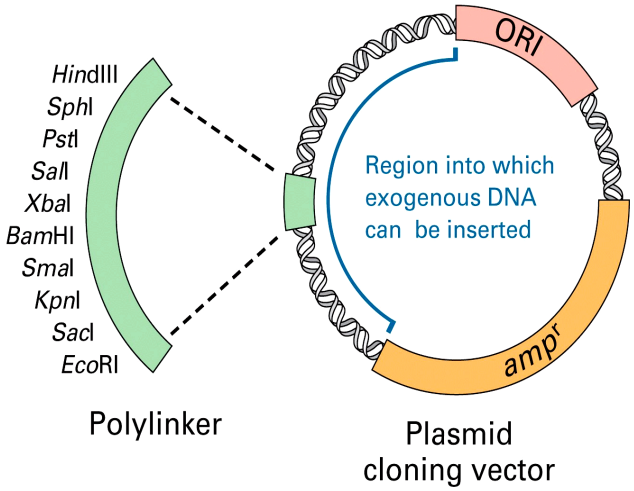
It can be replicated indefinitely.

Plasmids are extrachromosomal and can be easily purified away from the much larger host cell chromosome to isolate the cloned sequence.

Ligation of a pool of insert fragments to the vector molecules creates a library of clones.

A typical cloning vector for E. coli

Origin + selectable marker + DNA insertion site (polylinker of many unique restriction sites)



E. coli *oriC* : the origin of DNA replication

13 bp repeats
AT-rich

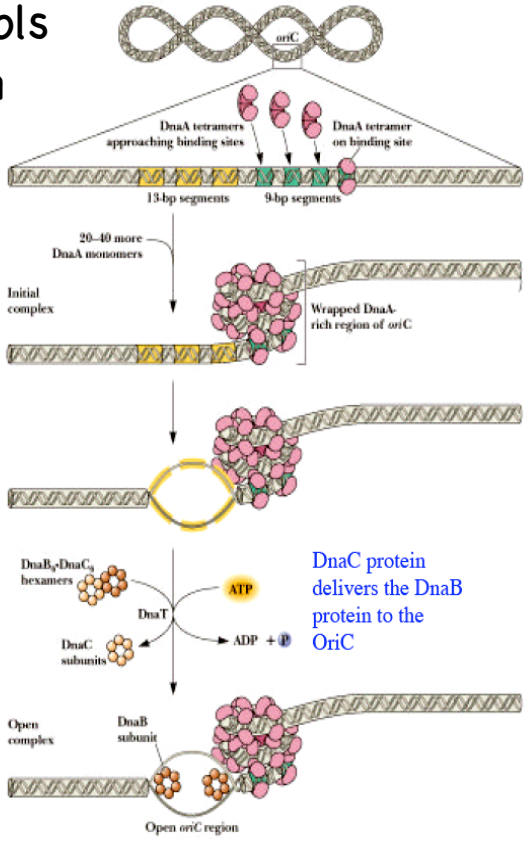
9 bp repeats
Binding sites for DnaA protein

The *oriC* sequence is necessary and sufficient for replication of a circular DNA in E. coli, but the DNA will replicate only once per cell division cycle (like the chromosome).

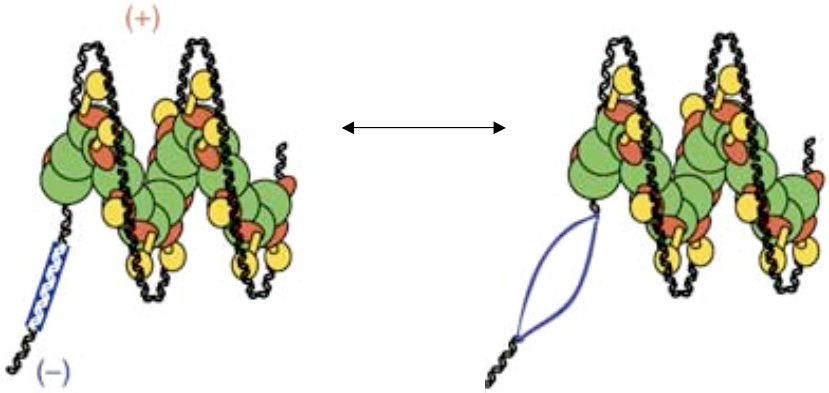
Plasmids with oriC have low copy number. Most plasmids used for cloning have a phage origin and high copy number to get more DNA

E. coli DnaA protein controls origin firing for replication

- ATP binding increases DnaA affinity for DNA
- many (~30) DnaA monomers bind cooperatively to 9 bp repeats and flanking DNA
- Wrapping of DNA around DnaA favors the unpairing of adjacent AT-rich sequence
- DnaA loads DnaB onto each strand
- DnaB hexamer can't get on DNA without opening the hexameric ring; a DnaC hexamer loads DnaB
- DNA binding by DnaB releases DnaC
- DnaA hydrolysis of ATP promotes disassembly



Crystal structure of DnaA reveals the mechanism of DNA wrapping

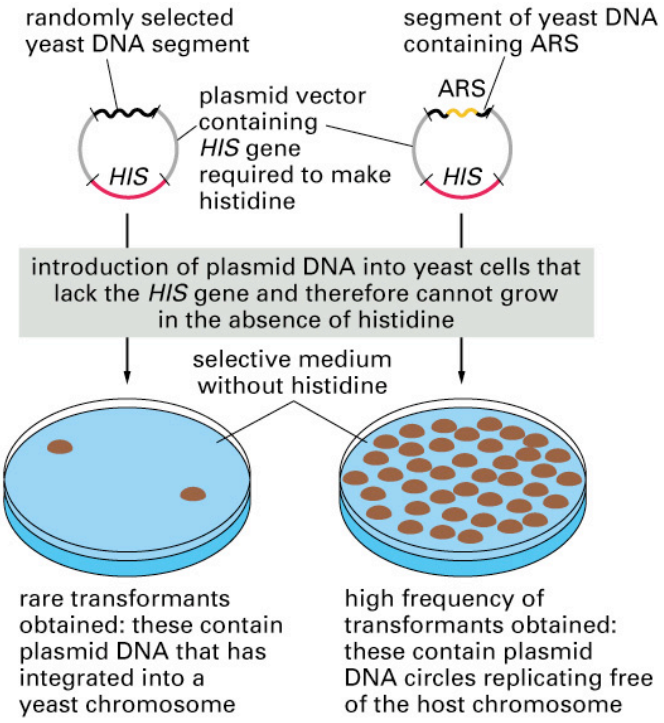


The arrangement of DNA binding sites introduces positive (right handed) supercoils by wrapping DNA on the outside. Compensating strand unpairing in the adjacent AT-rich region creates a replication bubble reading for loading of DnaB.

Eukaryotes need multiple replication origins

Genome	Fork speed	S phase	Origins	Comment
		S phase = DNA Synthesis		
E. coli	4.6 Mbp	30 kb/min	40 min	1	
Yeast	14 Mbp	3 kb/min	20 min	~330	S would last 80 hr if 1 ori
		1 L culture = 4.10^{10} cells --> 400 000 km DNA synthesized (Earth-Moon distance)		
Human	3 Gbp	3 kb/min	7 h	>10 000 ?	S would last 1 year if 1 ori
		2.10^{13} km DNA synthesized (2 light-years) during life time (10^{16} cell divisions)		

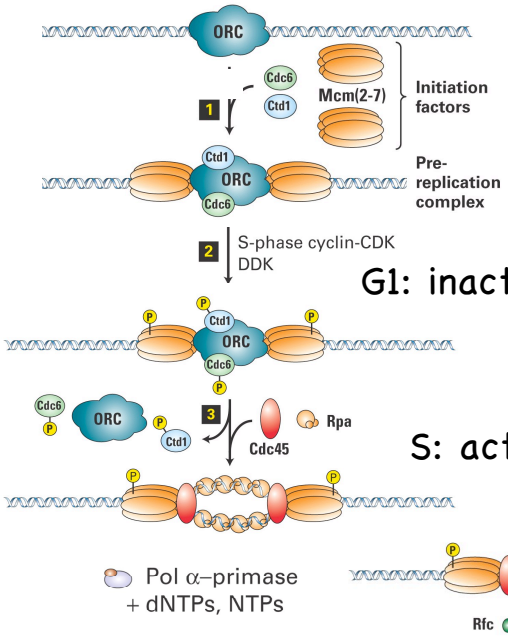
Identification of eukaryotic replication origins



... but not all yeast ARS (autonomously replicating sequences) are used as origins

... and same experiment doesn't find any human DNA pieces able to serve as origins

Eukaryotic DNA replication origins are regulated with the cell cycle: enabled in G1, activated in S, and deactivated by DNA replication.



Phases of the cell cycle:
G1 growth
S genome replication
G2 growth
M cell division

Eukaryotic factors for initiation of DNA replication

- Origin recognition: Origin Recognition Complex (ORC)
- DNA helicase: MCM2-7 (heterohexamer)
- Many other proteins interact with ORC and MCM to regulate the sites and timing of replication initiation.

Model systems: invading genomes

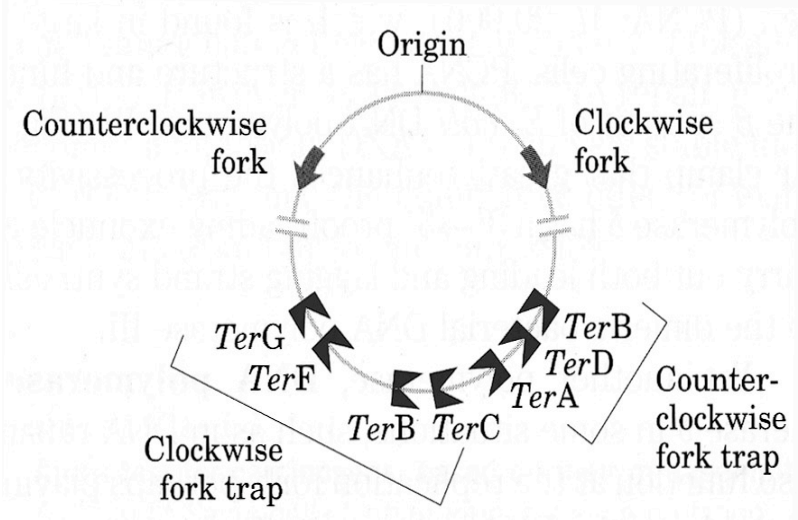
Bacteriophage lambda lO binds origin, lP allows loading of DnaB

SV-40 virus T antigen binds the origin AND is the helicase

to be infective these must overcome host regulation of chromosome replication once per cell cycle

Ter sites coordinate completion of E. coli chromosome replication

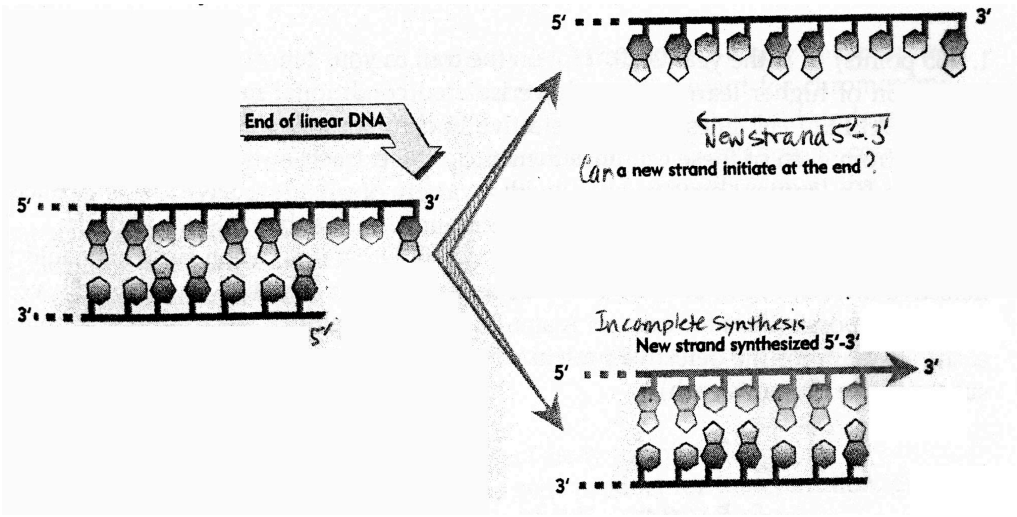
Tus protein binds Ter sites and inhibits the DnaB helicase

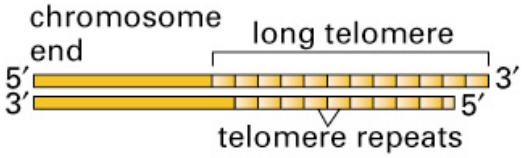


The ter/tus system is not essential

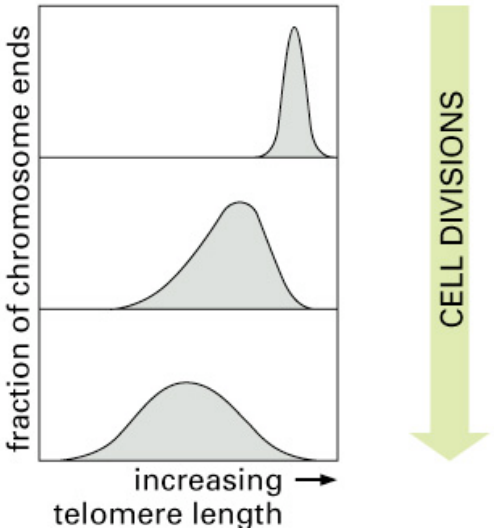
Ends of eukaryotic linear chromosomes incompletely replicated by DNA-dependent DNA polymerases.

The leading strand loses its overhang.
The lagging strand primer is degraded and not replaced.





Chromosome ends (the telomeres) lose some of the end-capping telomeric repeats with each round of genome replication.



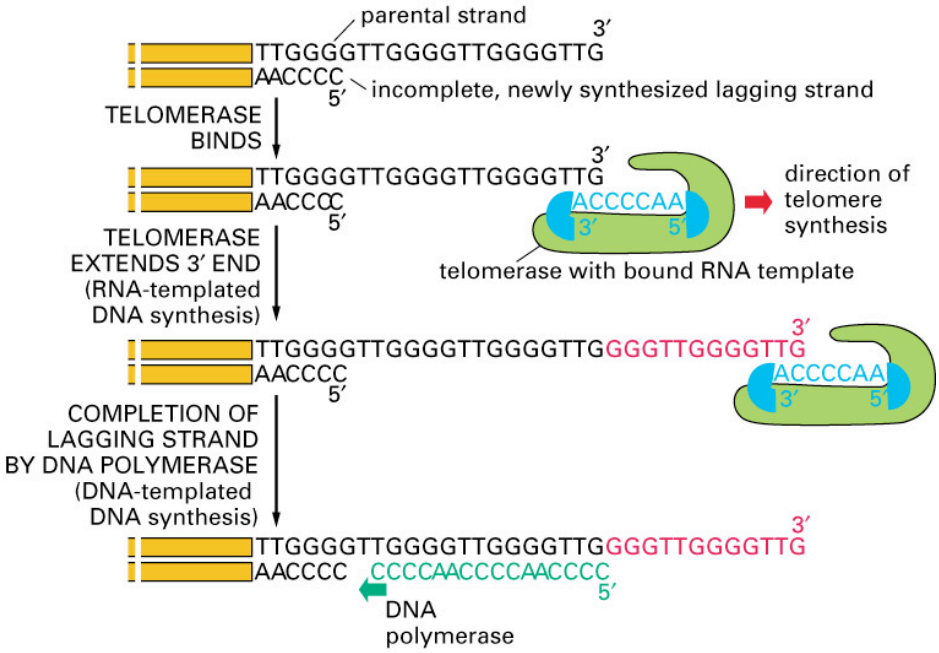
Telomerase adds back telomeric repeats without a DNA template: it is a reverse transcriptase, an RNA-templated DNA polymerase.

Telomerase extends the leading strand telomere overhang. Synthesis is in the 5′-3′ direction, as for all polymerases.

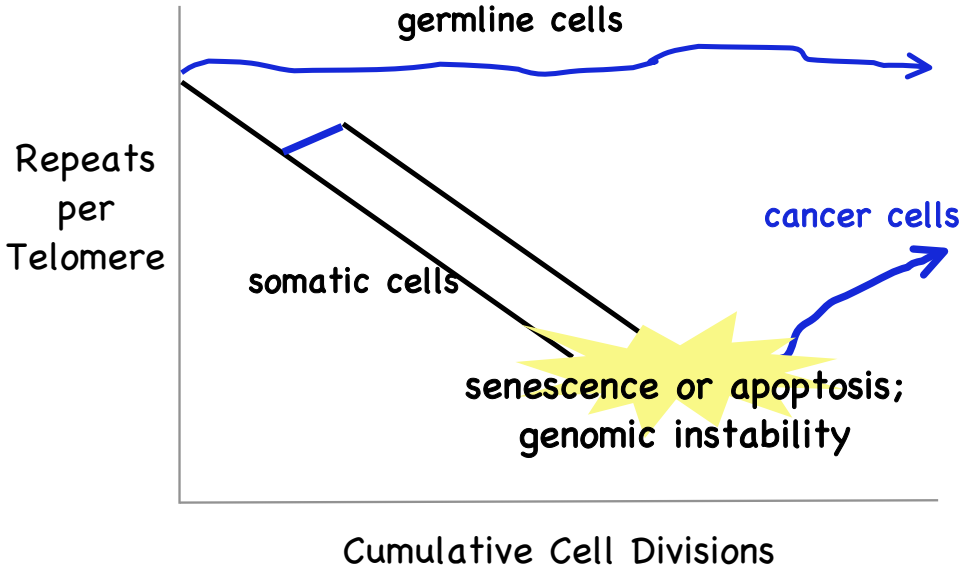
Telomerase is a ribonucleoprotein (RNP). The enzyme contains RNA and protein subunits.

The telomerase RNA subunit contains the template for DNA synthesis. The proteins include the telomerase reverse transcriptase TERT.

Telomerase uses the chromosome end as primer



Human telomere length dynamics



Human genome sequencing suggests that individual humans are 99.9% identical

Sequencing revealed one major allele for most genes in populations

Many variations have not had sufficient evolutionary time to spread throughout populations: subpopulations can have higher incidence of particular traits.

But individuals are 0.1% different

Lots of variation!

3.2×10^9 bp/genome x 0.001 changes/bp =

3.2×10^6 changes/genome

These include base changes, insertions, and deletions.

Sequence variations were first compared across the human population by restriction fragment length polymorphisms: RFLPs

Sequence variations are now compared using single-nucleotide polymorphisms: SNPs

How to find a gene linked with human disease?

Map disease genes by genetic linkage: find SNP or RFLP co-inherited with disease in many members of many families, then look at candidate genes in the disease-linked region of the chromosome.

Like classical genetics, except that the markers are molecular instead of phenotypic, so there can be many more markers per genome and markers in non-coding sequence.

RFLPs

Restriciton Fragment Length Polymorphisms
(Changes of restriction enzyme sites)

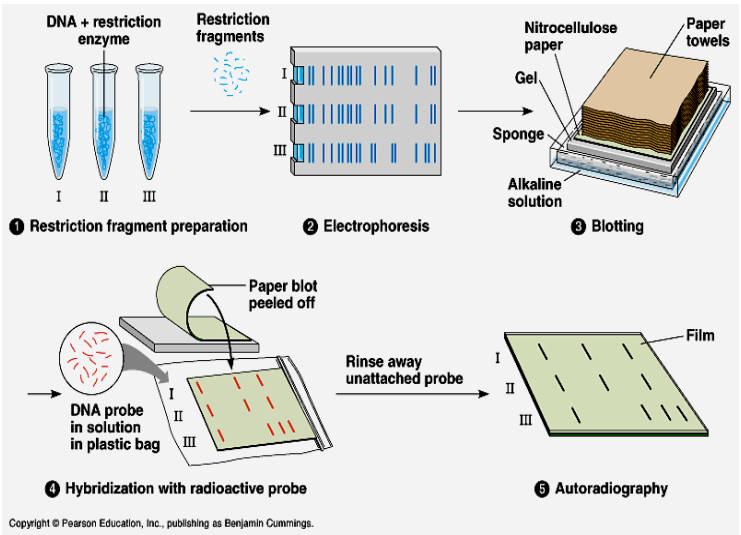
For every random 3 x 10^6 SNPs:

~1/256 will be in 4-base restriction sites
-> ~10^4 RFLPs for EACH four-base cutter!
~1/4096 will be in 6-base restriction sites
-> ~ 7.5 x 10^2 RFLPs for EACH six-base cutter!

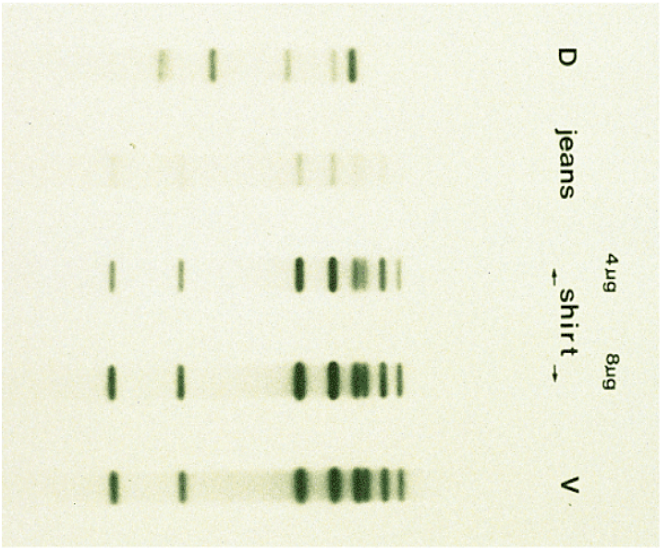
Lots of markers to map genes by linkage to RFLPs

Southern blotting can be used to detect a specific restriction fragment within a total DNA sample

1. Digest total DNA using restriction enzyme(s).
2. Run gel.
3. Transfer DNA from gel to filter paper.
4. Denature total DNA, denature labeled specific DNA probe fragment of interest, hybridize, wash off excess probe.
5. Detect the probe on the paper.



RFLPs as DNA fingerprints in a murder case



Southern blot of DNA samples digested with a restriction enzyme

SNPs

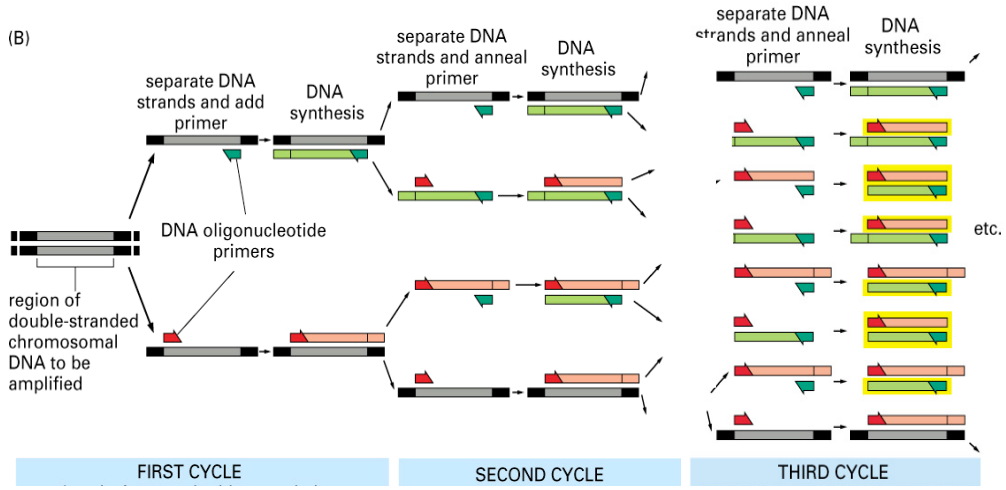
Single Nucleotide Polymorphisms

How to detect these?

Differential hybridization is tricky with only a single mismatch.

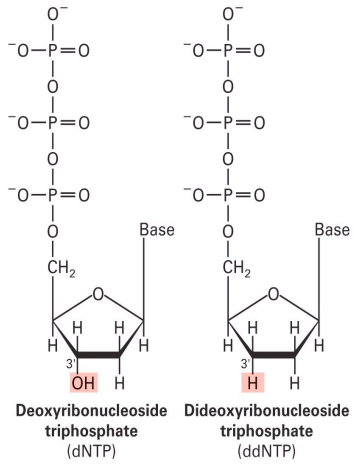
Most detection is by DNA sequencing.

PCR (Polymerase Chain Reaction) can amplify any known DNA sequence

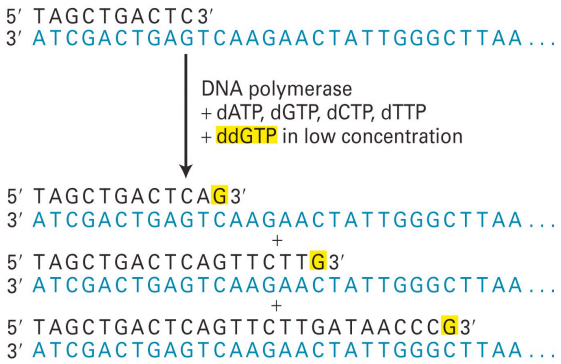


N cycles amplifies the target sequence 2^N-fold

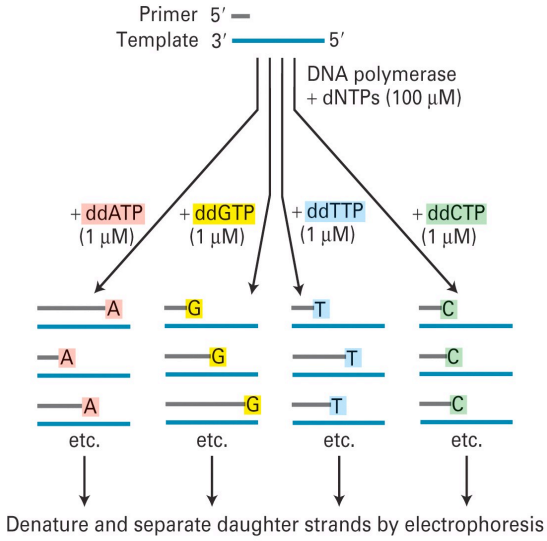
DNA sequencing by chain termination



ddNTPs terminate 5′-3′ synthesis of the polynucleotide chain



A small amount of ddGTP with excess dGTP terminates chains at various Cs in the template



DNA sequencing by chain termination

1. All fragments start at the primer.
2. All fragments ending in a particular base have a different length and a different color tag.
3. Separating the mixture of products by size reveals the sequence.