

## DNA Replication in *E. coli*

### References

Voet<sup>2</sup>, pp. 854-857, 1020-37.

Alberts, B. et al., pp. 238-257.

Snyder, L. and Champness, W. "Molecular Genetics of Bacteria", 2nd edition, 2003, ASM Press, Washington DC. Chapter 1, pp.13-45. - *Good introductory coverage.*

Friedberg, E.C. et al. 2001. Error-prone DNA polymerases: novel structures and the benefits of infidelity. *Cell*. 107, 9-12. *A recent review on new polymerases.*

Lusetti, S.L. and Cox, M.M. 2002. The bacterial RecA protein and the recombinational DNA repair of stalled replication forks. *Ann. Rev. Biochem.* 71,71-100. *Current research on the integration of replication, repair and recombination mechanisms.*

Sherratt, D.J. 2003. Bacterial chromosome dynamics. *Nature* 301, 780-785. *Nice review of the current state of the field.*

### A. The *E. coli* Chromosome

The *E. coli* chromosome is a circular, supercoiled structure consisting of 4,639,221 bp predicted to encode 4289 proteins. Unlike eukaryotic chromosomes, there are no histones or introns but there are a number of basic proteins that bind to DNA to form an elaborately folded structure called a nucleoid. The bacterial chromosome is maintained in a negatively supercoiled state by the action of topoisomerases with opposing actions. One of these is DNA gyrase which inserts negative supertwists into covalently closed DNA and is a target for a class of therapeutically effective antibiotics (nalidixic acid and fluoroquinolones).

A "classic" paper in the literature is that of Watson and Crick (1953) on the structure of DNA. The structure suggested that DNA replication is semi-conservative but other modes of replication could not be excluded. In another classic paper, Meselson and Stahl (1958) grew *E. coli* cells in medium with <sup>15</sup>N and then shifted the cells to medium with <sup>14</sup>N. Samples were removed at various times, the DNA extracted, sheared and centrifuged to equilibrium in a CsCl gradient. The results were consistent with a semi-conservative mode of replication.

#### A.1. Initiation of Chromosome Replication

DNA replication initiates at *oriC* and proceeds bidirectionally around the circular molecule. This is referred to as "theta" mode replication. It takes 40 min to replicate the entire chromosome yielding a fork progression rate of about 1,000 nucleotides per second. The frequency of initiation of chromosome replication is dependent on the rate of cell division: the shorter the doubling time, the more frequent initiations occur at *oriC*.

The *oriC* region is quite small - about 250 bp (sequenced in 1978). There are several proteins that bind to the origin (HU, IHF, Fis, DnaA); of these DnaA is the most important as it initiates chromosome duplication by binding to DnaA "boxes". Once the strands of *oriC* are separated, DnaC is bound and recruits the DnaB helicase followed by the DNA polymerase III holoenzyme proteins.

DNA synthesis is regulated at the initiation step. When the initiation stage has been completed, it is important to prevent re-initiation until required later in the cell cycle. In part, this is regulated

by the concentration of DnaA as well as by the SeqA protein which binds to hemi-methylated origins (*E. coli* DNA is methylated at the A in GATC sequences).

## A.2. Elongation

When initiation has been completed, replication of the entire chromosome ensues. This requires the action of many proteins working in concert and these constitute a "replisome" which is stationary relative to the DNA. Replisome proteins include gyrase (a topoisomerase), DnaB helicase, single-strand binding protein (Ssb), DNA polymerase III holoenzyme, primase (DnaG), DNA polymerase I, DNA ligase and a topoisomerase (with the opposite action of gyrase).

All DNA polymerases synthesize in the 5' to 3' direction only. This means that at the replication fork, one strand is synthesized continuously. The other is made discontinuously, in the form of short fragments (about 1000 nt) by repeated cycles of priming by primase, extension by DNA polymerase III, RNA removal and resynthesis by DNA polymerase I, and sealing by DNA ligase.

The processivity of DNA polymerase III is awesome- about 1000 nt per second. A key protein responsible for this is the  $\beta$  subunit, a doughnut shaped protein that clamps the  $\alpha$  (catalytic) subunit onto the DNA. The holoenzyme is an asymmetric dimer and leading and lagging strand replication is coupled. This means that there is a lot of activity associated with the lagging strand holoenzyme monomer- the holoenzyme monomer must disassemble and reassemble for each discontinuous (Okazaki) fragment. Many of the holoenzyme proteins function in loading/unloading the  $\beta$  clamp.

It is essential that replication of the genome be accurate. Part of this fidelity is achieved by a proofreading function built into the holoenzyme via the  $\epsilon$  (epsilon) subunit which has 3' to 5' exonuclease activity. This nuclease can excise the terminally mispaired nucleotide (precisely reversing the synthetic activity) allowing the catalytic subunit another try to put in the correct nucleotide. Although the  $\alpha$  -subunit has catalytic activity, accurate proofreading requires the action of both the  $\alpha$  and  $\epsilon$ -subunits.

Until a couple of years ago, it was assumed that replication complexes formed at *oriC* traversed the entire chromosome. There is increasing evidence, however, that this is not the case. It appears that replication fork progression can be blocked (at sites of endogenous DNA damage?) leading to dissociation of the holoenzyme complex. If this occurs, then a DnaA-independent mechanism of re-initiation of replication ("replication restart") must take place that appears to require the *pri* (primosome) gene products. Replication restart appears to also involve repair and recombination proteins and is a current topic of intensive investigation.

Certain types of DNA damage present a roadblock to fork progression by DNA polymerase III holoenzyme. When this occurs, alternate DNA polymerases come into play, many of which are encoded by genes that are inducible. Unlike DNA polymerase III, DNA polymerases IV (DinB) and V (UmuDC) are able to synthesize across certain types of damaged DNA but these enzymes have a distributive (as distinct from processive) mode of action. Upon completion of their DNA synthesis, they are replaced by DNA polymerase III holoenzyme by a mechanism that is not yet

understood. DNA polymerase II is also involved in replication of imperfect DNA but its function is unclear.

DNA Polymerase III holoenzyme subunits.

Subunit	Function	M (kD)	Gene	Phenotype
$\alpha$	Polymerase	128	<i>dnaE</i>	Ts (temperature sensitive)
$\epsilon$	3' to 5' exo	28	<i>dnaQ</i>	Ts and/or mutator ( <i>mutD</i> )
$\tau$	Dimerizes core	71	<i>dnaX</i>	Ts
$\theta$	?	9	<i>holE</i>	None
$\beta$	Sliding clamp	41	<i>dnaN</i>	Ts
$\gamma$	Clamp loader	48	<i>dnaX</i>	None
$\delta$	Clamp loader	39	<i>holA</i>	Essential
$\delta'$	Clamp loader	37	<i>holB</i>	Essential
$\chi$	Clamp loader	17	<i>holC</i>	Ts
$\psi$	Clamp loader	15	<i>holD</i>	Unknown

*hol* = holoenzyme

Note that the *dnaX* gene encodes two gene products that are produced by frameshifting.

-A AAA AAG AGU GAA-	-AAA AAA GAG UGA	mRNA
K K S E	K K E stop	protein

### A.3. Termination and segregation

There are DNA sequences located about 180° from *oriC* that block fork progression. These sequences are present in multiple copies and half are in the opposite orientation. When Tus protein is bound to these Ter sites, DnaB helicase activity is inhibited and fork progression halted. The duplicated interlocked chromosomes are resolved by at least two topoisomerases and then partitioned into daughter cells but the mechanism of segregation is unknown. This is due mainly to the lack of genetic selections to isolate mutants involved in the process. However, the integral membrane protein FtsK is essential for cell division and interacts with FtsZ, a tubulin that constricts the cell at its middle.

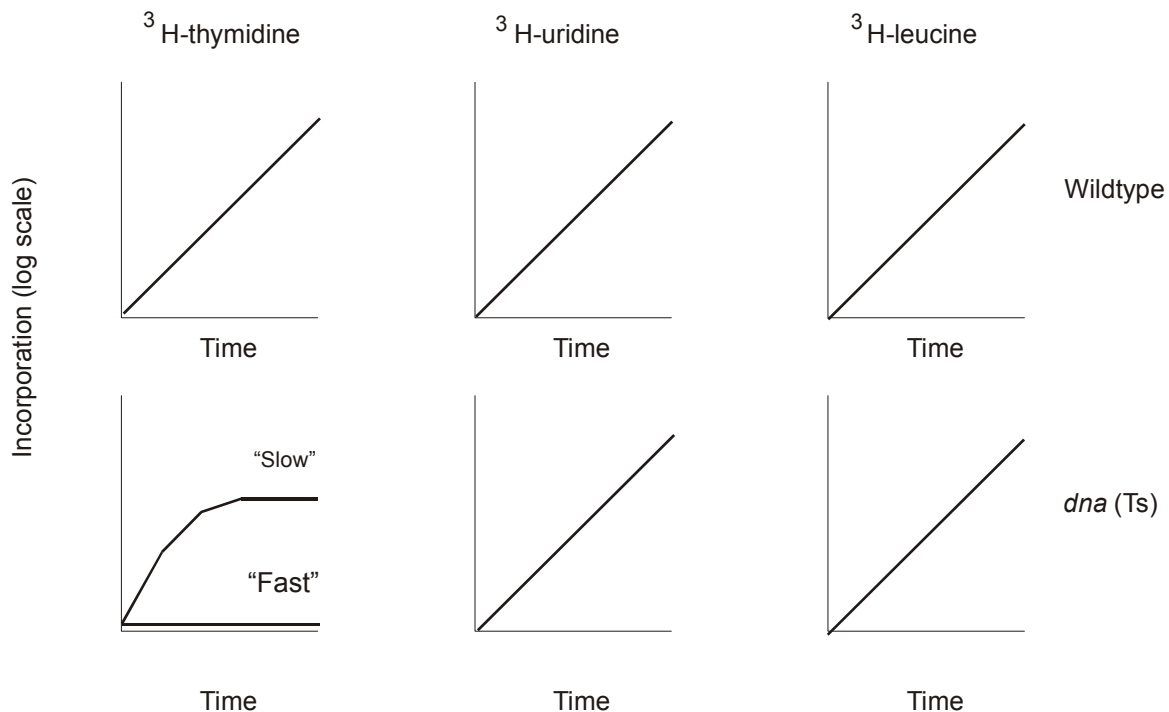
Recombination between replicating chromosomal regions can occur to form chromosome dimers. A site-specific recombination system (XerC) ensures that dimers are converted to monomers before chromosome segregation.

## B. DNA Polymerase I

In the 1950's, Kornberg and colleagues described a DNA polymerase from *E. coli* that was responsible for over 99% of polymerase activity in crude extracts. For many years this activity was known as "DNA polymerase". A proteolytic fragment of it is known as the "Klenow" fragment and lacks the 3' to 5' exonuclease activity.

## C. Genetics of DNA replication

Early genetic studies focused on how mutations affecting essential cellular functions could be characterized. For DNA replication, temperature sensitive conditional lethal mutations were isolated by a "brute force" method and designated *dna* (Ts). The *dna* mutants exhibited reduced incorporation of <sup>3</sup>H-thymidine into DNA at the non-permissive temperature and could be separated into two groups: "fast stop" and "slow stop".



The figure above shows the incorporation of radioactive precursor into DNA, RNA and protein in the wildtype and *dna* (Ts) mutant at the non-permissive temperature (42° C).

None of the *dna* mutants were deficient in DNA polymerase I activity.

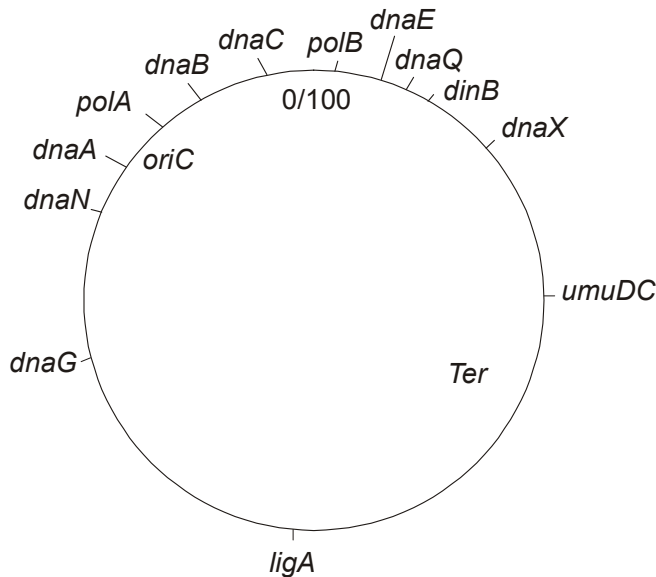
## D. Polymerase Mutants

DeLucia and Cairns (1969) screened extracts from each of 3500 mutagenized survivors for DNA polymerase I activity. One was found with less than 1% of wildtype activity. The mutant strain, designated *polA*, grew just fine (even at 42° C) indicating that polymerase I could not be the enzyme that actually replicates the chromosome. Subsequent characterization showed that the *polA* mutant was sensitive to ultraviolet light and other damaging agents. There was also a decreased rate of incorporation of Okazaki fragments into high molecular weight DNA.

Extracts of the *polA* mutant were used to purify residual polymerase activity. Two distinct activities were detected and designated polymerase II and III. The *polB* mutant showed normal growth. A *dnaE* (Ts) mutant showed temperature activity for DNA polymerase III. Therefore DNA polymerase III is the replicase.

The Table below lists the current number and properties of *E. coli* DNA polymerases.

Polymerase	Gene	Function
I	<i>polA</i>	Repair and replication
II	<i>polB</i>	Repair (inducible)
III	<i>dnaE</i> (= <i>polC</i> )	Catalytic subunit of DNA polymerase III holoenzyme
IV	<i>dinB</i>	Repair (inducible)
V	<i>umuDC</i>	Repair (inducible)



**Location of some DNA replication genes on the genetic map.**

### E. Purification of the Holoenzyme

The *dnaE* gene product turned out to be part of a multi-subunit holoenzyme complex. A number of other subunits were found by their physical association with the DnaE protein (the  $\alpha$  subunit). Other holoenzyme proteins were found by combined genetic/biochemical approaches. In the first approach, extracts were made from each *dna* mutant at 42° C, a primer-template combination added as well as fractions from a wild type extract to find the component that restores DNA synthesizing activity. The second approach was to purify polymerase activity from a *polA polB* double mutant which was as processive as that *in vivo*.

### F. Generalizations about DNA Polymerases

1. They cannot initiate synthesis on double-stranded templates (cf. RNA polymerase).
2. A strict primer-template substrate is required.
3. They work by a nucleophilic attack of the 3'OH of the primer terminus on the  $\alpha$ -phosphate of the deoxyribonucleoside-5'-triphosphate precursors. The direction of synthesis is, therefore, 5' to 3'.
4. Many have integral or associated exonuclease activities.

All known DNA polymerases can be placed in families on the basis of amino acid sequence similarity. The active site configuration of replicases is conserved and the atomic structure of the phage T7 polymerase-template-primer has been solved to 2.2 Å resolution (Doublie et al., 1998 Nature 391:251-8).

### G. Eukaryotic DNA Polymerases

This subject is known in less detail, but as the information comes in, it turns out that major aspects of the mechanism are conserved.

Function	<i>E. coli</i>	Yeast	SV40/Human
Helicase	DnaB	MCM* proteins?	T antigen
Primase	DnaG	Pol $\alpha$	Pol $\alpha$
Polymerase	DnaE	Pol $\delta$ and pol $\epsilon$	Pol $\delta$
Proofreading exo	DnaQ	Subunit of pol $\delta$ and pol $\epsilon$	Subunit of pol $\delta$
Sliding clamp	DnaN	PCNA	PCNA
Clamp loader	$\gamma$ -complex	RF-C	RF-C
Single-strand DNA binding protein	Ssb	RP-A	RP-A

\*  $\gamma$ -complex,  $\gamma$  protein (*dnaX* product), HolA, HolB, HolC and HolD; MCM, minichromosome maintenance; RF-C, replication factor C; RP-A, replication protein A; PCNA, proliferating cell nuclear antigen.