DNA Replication

1. Query: How is the origin opened during the initiation step of replication of E. coli DNA? What is the role of negative supercoiling in this process? How does the DnaA protein help the initiation step?

Negative supercoiling refers to underwinding of DNA, that is, DNA that contains fewer Watson-Crick turns than relaxed DNA. Since DNA has to be unwound (strands separated) for performing replication (or transcription), an already underwound form of DNA (negatively supercoiled DNA) will help the process thermodynamically. If the energy has already been spent to remove DNA turns, the replication machine or the transcription apparatus does not have to invest that energy again to start the process of replication or transcription. Supercoiling is dynamic, that is the unwinding will be distributed dynamically over the whole DNA molecule. That is regions that are unwound will be constantly changing with time. [You may imagine continuous waves of supercoils distributed over a large DNA molecule as the waves performed by a crowd in a large sports arena].

For initiating DNA replication, the underwinding must be localized to the replication origin. The binding of DnaA helps this process. If the binding of DnaA causes positive supercoils (overwinding; additional DNA turns) to be trapped in the bound region, compensatory negative supercoils must be generated in the free DNA portion (or the adjacent DNA must be unwound). Since the DNA strands are not cut or joined, there is no change in the linking number of the molecule. So if you force positive supercoils into it (as DnaA does), the DNA will react by forming an equal number of negative supercoils.

[The situation is analogous to (but in the opposite sense) the case of ethidium bromide binding that we discussed. Ethidium bromide binding unwinds strands or effectively introduces negative supercoils, so the DNA responds by making compensatory positive supercoils.]

The DnaA protein binds ATP, and the ATP-bound form is the active form in opening the origin. There is likely an interesting energy coupling control between the initiation event and the prevention of a second initiation event immediately following the first (thus preventing over-replication of the genome). The energy of ATP hydrolysis may help open the DNA strands at the origin, and at the same time generate the ADP-bound form of DnaA, which is unable to support additional rounds of initiation. The DnaA protein binds ATP, and the ATP-bound form is the active form in opening the origin. There is likely an interesting energy coupling control between the initiation event and the prevention of a second initiation event immediately following the first (thus preventing over-replication of the genome). The energy of ATP hydrolysis may help open the DNA strands at the origin, and at the same time generate the ADP-bound form of DnaA, which is unable to support additional rounds of initiation.

2. Query: After completion of DNA replication, how is a mis-incorporated base corrected? And how is the correction controlled to avoid introducing a mutation?

After DNA synthesis is completed, any mismatch (mis-incorporated) base will be corrected by the mismatch repair system. To avoid mutations, the mismatched base from the new strand should be removed and replaced by the correct base complementary to the parental strand. The parental strand contains methylated bases that serve as signals for the repair enzymes to remove the base from the new strand, which is not yet methylated. There is a lag time before the new strand gets methylated. This methylation is directed by the methyl signal in the parental strand. This is possible because the methylated sequence is symmetric, for example: 5'GATC3'/3'CTAG5'. The A in this sequence is methylated by the DAM methylase enzyme. The methyl signal at A in the parental strand directs the enzyme to methylate the A in the bottom strand.

3. **Query**: Could you explain how DNA replication is adjusted when the cell division times are changed?

The easiest way to think about it (at least for me) is to imagine that the bacterium has a built in cell cycle clock that runs from the firing of the replication origin to the division of the mother cell into two daughter cells. This internal cell cycle clock is fixed, and does not run faster or slower when the cell division time is changed due to availability of nutrients.

Let us say that this clock for E. coli is a one hour clock, that is, the clock starts at 0.00 hr (0 min.) when the origin fires and then measures 1.00 hr (60 min.) to execute the cell division event. The point to note, as we shall see below, is that there are situations when the origin will fire more than once within a single cell, for example twice. There will be a division event 60 min from the firing of the first origin, and a second division event 60 min from the firing of the second origin. Or each origin firing event sets a 60 min. timer (or an alarm clock), which 'goes off' to 'announce' the time for the cell to divide.

The time it takes to replicate the DNA of E. coli is approx 40 min. This time is dependent on the deoxynucleotide triphosphate concnetration in the cell and the rate at which the polymerase enzyme incorporates the deoxynucleotides into the growing DNA chain. Under most conditions, this time is fixed because the nucleotide pool is not limiting and the replication machine has been optimized over evolutionary time to function at its maximum efficiency. Indeed, if the nutrient source is extremely poor, and the cell is starved for deoxynucleotides, the replication machine will slow down, increasing the time of chromosome duplication.

Imagine a cell growing rather slowly under nutrient conditions such that each cell divides every 60 minutes in synchrony with the internal clock. Under this condition immediately after each division the origin fires. 40 min. later chromosome is duplicated. There is a gap period of 20 min (this gives the cell time to complete repair, DNA methylation of the nascent strand, decatenation of linked duplex circles, resolution of dimers etc.), and the cell divides, the origin fires again, and the cycle continues.

I have illustrated this in Fig. 1.

Now imagine a cell growing in rich medium so that the division time is shortened to 40 min. The internal clock remains the same; it still measures sixty min. from origin firing to cell division. Under these conditions, it is as if the cell would have to fire its origin at minus 20 min. so that division can occur at 40 min. (cell cycle length = 40 - (-20) = 60 min.). At zero time, the DNA has already replicated half way (if 40 min. is the time for full replication, half the genome will be replicated in 20 min.). In another 20 min., completion of replication gives two copies of the

genome. According to the clock, cell division can occur in another 20 min, the cell having run through the cell cycle from -20 min to 40 min. The 40 min point (cell division) is also the zero point for the next cell cycle, with another division event to follow in 40 min. So the 20 min. point in the previous cell cycle (when the chromosome has been completely replicated) is also the -20 min. point for the following cell cycle. Right? Because in 20 min. the cell is going to divide, setting the 0 min. point of the next division cycle. So the origin has to fire at -20 min., or midway through each cell cycle. The result is that each of the chromosome in the cell that is about to divide has been half-replicated at the time of division. The same process continues through the rest of the cell cycles as the population grows.

In other words, the cell responds to the shortened division time by firing its origin every 40 min. rather than every 60 min., so that the cell cycle clock of 60 min from origin firing to cell division is maintained constant. Or, the chromosome content of the steady state population of cells is not one copy of the chromosome but one and one half copies per cell. If you look at sequential cell cycles, you will note that there is a cell division event 60 min. following each origin firing event.

I have illustrated this in Fig. 2.

Now you can try the case where the cell is dividing every 30 min. You will have to fire the origin for the first cell cycle at -30 min. at 0 time for the second cell cycle, at 30 min for the third cell cycle and so on. The division event for the firing at -30 min will happen at +30 min (30 min. cell division time, from 0 to 30 min.; and 60 min cell cycle period, from -30 min. to 30 min.); that for the firing at 0 min. will occur at 60 min. (again 30 min. division time, from 30 min to 60 min.; and 60 min. cell cycle period, from 0 min. to 60 min.); and that for the firing at 30 min. will occur at 90 min. (30 min. division time, from 60 min. to 90 min; and 60 min. cell cycle period, from 30 min. to 90 min.).

In other words, the origin is firing at every thirty minute interval. And the average genome content per cell will be one and three quarters equivalent per cell.

Please apply the same logic to cells dividing every 20 min.

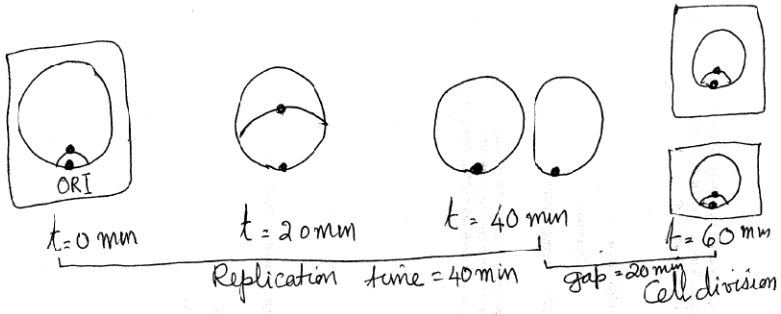

Fig 1. Cell division time = 60 min.
Replication time = 40 mm.
Gap = 20 min.
Cell cycle clock = 60 min.
Interval between origin firings = 60 min

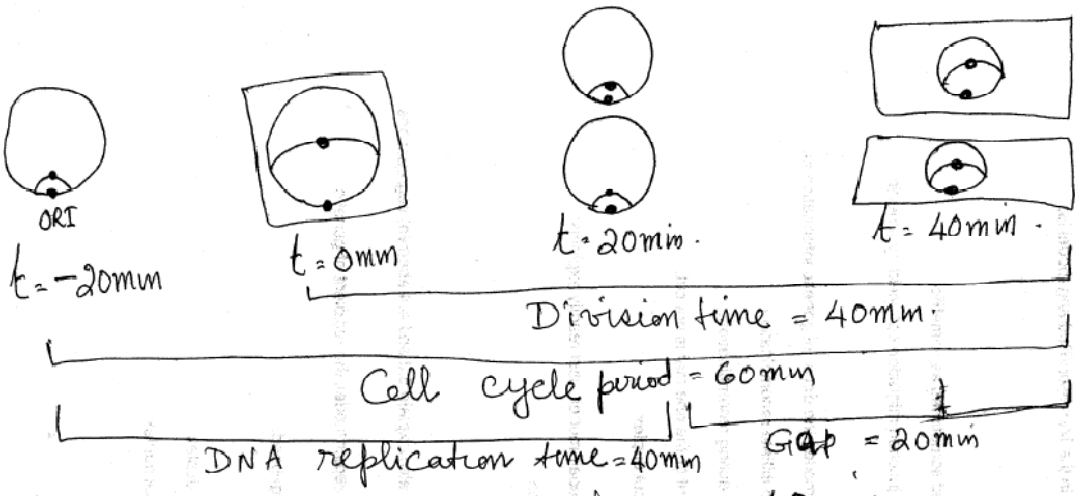

Fig 2.

Cell division time = 40min.
Replication time = 40 min
Gap = 20 min
cell cycle clock = 60min
Genome content per cell = one + one half equivalent
Interval between origin firings = 40min

4. **Query**: How did you calculate the error rate during replication?

The rate of mistakes made by the polymerase during replication is about one in one thousand to one in ten thousand nucleotides. The proof reading activity of the polymerase complex (3' to 5' exonuclease activity to remove the misincorporated nucleotide and re-polymerization in the 5' to 3' direction) reduces the error rate by a factor of one hundred to one thousand. Finally the mismatch correction improves accuracy by a factor one hundred. This is the step where the methylation directs the mismatch correction to avoid mutations. The template strand is methylated at specific positions whereas the nascent strand is not immediately methylated. The mismatch repair system removes the mismatched pair by cutting the nascent strand (which is not yet methylated), and repairing the cut strand using the uncut one as template. Hence the correction disfavors mutations by restoring the same base pair as in the parent DNA molecule.

If you assume the rate of mistake by the polymerase as 10^{-4}, and the reduction of this error rate by factors of 10^{-3} (for proof reading) and 10^{-2} (for mismatch repair), the final error rate is:

$10^{-4} \times 10^{-3} \times 10^{-2} = 10^{-9}$.

So, there is roughly one mistake made (wrong base incorporated) when 10^{9} base pairs are replicated. Since E. coli has 4×10^{6} bp in its DNA, one mistake is made when 250 copies of it are made. One in every 250 cells formed during the growth of the bacterial population will carry

a changed base pair (or mutation). Let us assume that E. coli has 4,000 genes. Since the mutation could be in any one of these 4,000 genes, the chances of a particular gene being mutated is 1/250 x 1/4000 = 1/1000,000 = 10^{-6}. The actual mutation frequency is less, around 10^{-7}. This is not surprising, since all base pair changes will not cause a change in gene function: mutations in non-coding regions, mutations in regulatory regions that are silent, mutations in the third (wobble) position of codons which will not change the amino acid, mutations in certain amino acids that still retain function etc.