Molecular Biology Through Discovery
DNA Replication

Outline:

A. Overview of DNA replication [Part I]
B. Problems raised by the double helix [Part II]
C. Meselson and Stahl experiment [Part II]
D. Reconciliation of semi-conservative replication with double helix [Part II]

A. Overview of DNA replication

The overall goal of these notes is to help you understand how the general strategy of DNA replication was elucidated, despite the formidable theoretical objections raised against it. In going through the historical events, it may be helpful to do so armed with our present day understanding of DNA replication. This section may serve as an introduction or review of its main features.

You might find one or more of the following animations on DNA replication helpful:

- Crude and simple (23 sec)
  (http://www.uark.edu/campus-resources/mivey/images/lagging%20fast.gif)
- More sophisticated graphics + narration (1 min 33 sec)
  (http://www.youtube.com/watch?v=teV62zrm2P0)
- 3-dimensionally more complex graphics + narration (2 min 5 sec)
  (http://www.youtube.com/watch?v=-mtLXpgjHL0)

DNA replication may be broken down into the following steps:

1. Initiation of replication
   A DNA molecule is not replicated all at once. Instead, replication proceeds from one or more origins of replication (**Fig. 1, inset**). Bacterial chromosomes, bacteriophages, and plasmids generally have a single origin of replication. The much larger chromosomes of eukaryotes generally have many. The origins are determined by the binding of specific proteins to specific DNA sequences.

2. Unwinding of the DNA
   Since DNA replication requires access to the bases (A, C, G, or T) hidden within the double helix, the region that is undergoing replication needs to be unwound. There are two problems that need to be solved: how to do the unwinding, and how to prevent the unwound strands from snapping back to reform the double helix. Unwinding is accomplished by a protein, DNA helicase (an unnamed green blob in **Fig. 1**) that sits at the replication fork (shown in **Fig. 1** as a site where the DNA separates into two dark blue strands) and separates the two strands, using ATP for energy. The second problem is solved by single-strand-binding proteins (shown in **Fig. 1** as gray blobs sitting on the unwound strands near the replication fork). They bind to single strands of DNA exposed by unwinding the helix and so make them unavailable to pair with their complementary strands.

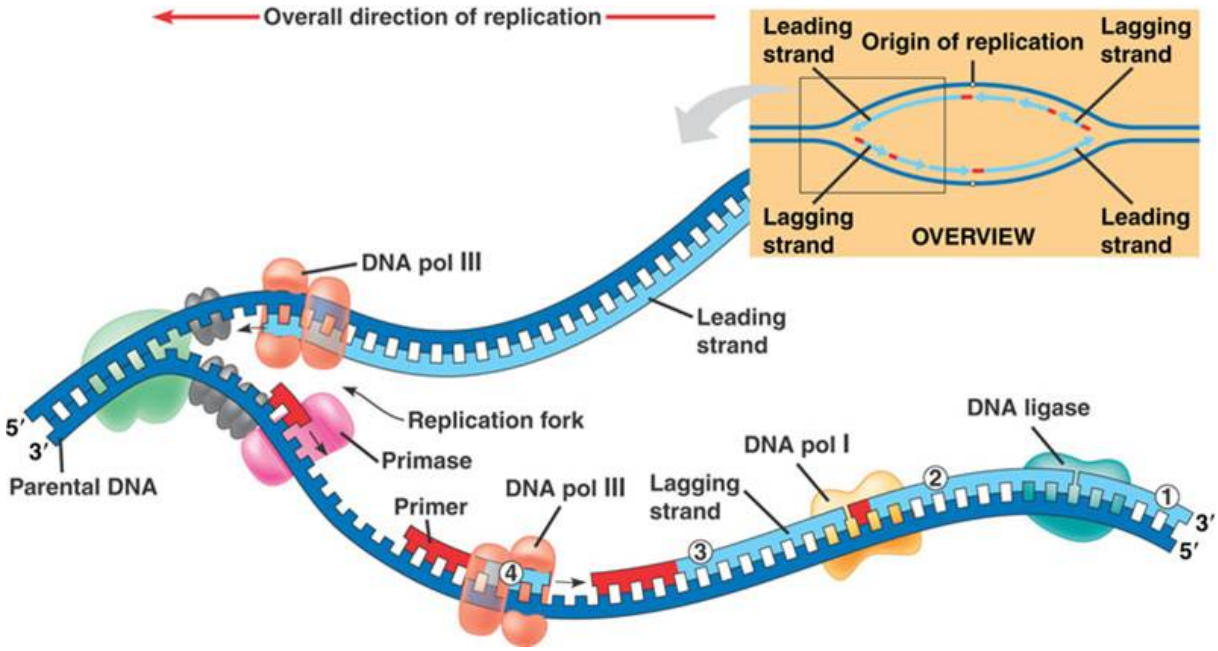

Figure 1: DNA replication. (image from Teresa Fischer, Indian Rivers State College)

SQ1. Why is it necessary for the DNA to unwind in order for replication to take place?

3. Initiation of DNA synthesis

   This step can be understood only by bearing in mind the following:

   Rule #1: Synthesis of DNA always proceeds by adding nucleotides to a pre-existing 3'-OH end on the strand to be synthesized

   This implies that you can't simply open up a double helix and start synthesizing DNA, as at the beginning, there is no pre-existing new strand. DNA synthesis absolutely requires a primer (red fragments in **Fig. 1**). Why this rule exists is discussed later in the Overview.

   If DNA synthesis cannot begin from nothing, then how ***does*** it begin? The solution is to begin a new strand with an ***RNA*** primer, synthesized by the enzyme primase (pink blob in **Fig. 1**). The synthesis of RNA does not follow the same rule as that of DNA. It ***can*** start from nothing.

SQ2. Identify the regions in Fig. 1 that are composed of RNA and those that are composed of DNA. Where is the RNA primer for the strand labeled 'Leading strand'?

4. Extension of DNA

   This step can be understood only by bearing in mind the following:

   Rule #2: DNA is always made 5' to 3'

   Again, the why behind this rule will be discussed later. For now, it is enough to note that in **Fig. 1**, the arrows for both new strands (light blue) are pointing 5' to 3'. New DNA is synthesized by DNA polymerase (labeled 'DNA pol III' in **Fig. 1**) by recruitment of free nucleoside triphosphates (not shown in the figure) to pair with the exposed base on the parental strand (the template) just beyond the 3' end of the new strand (where the arrow is in the figure), as shown in **Fig. 2**).

SQ3. What is the orientation of the upper parental strand (top, dark blue)? What is the orientation of its associated new strand? How about the orientations of the lower parental strand and its associated new strand?

SQ4. Would you describe the overall direction of replication (which in the Fig. 1 is right to left) as 5' to 3'? 3' to 5'?

5. Extension of DNA (leading strand)

   One of the two new strands is synthesized in the same direction as the movement of the replication fork. In **Fig. 1**, this is the new strand labeled 'Leading strand'. The movement of the replication fork exposes additional downstream template to which new nucleotides can be paired and added to the growing new leading strand by DNA polymerase. Extension of the leading strand can proceed for millions of nucleotides.

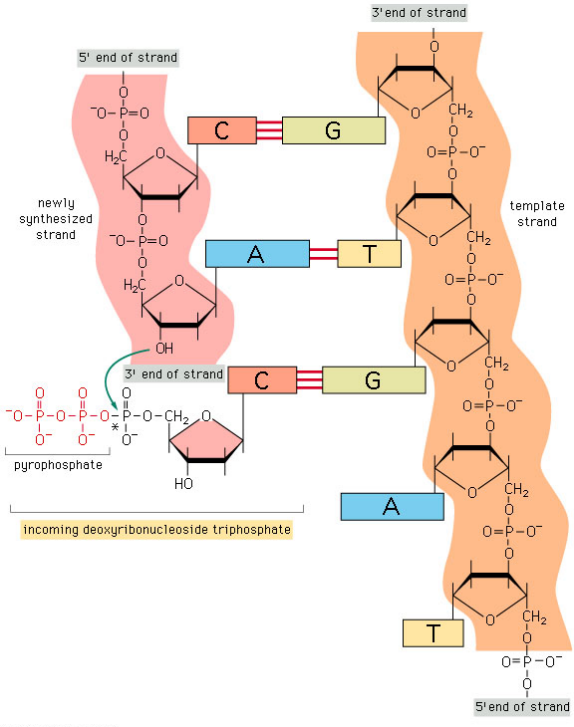

Figure. 2. Recruitment of nucleoside triphosphate to add onto new strand. (from Alberts B et al *Essential Cell Biology*, 1998, Garland Publishing)

6. Extension of DNA (lagging strand)

   The other new strand is called the lagging strand. It is synthesized in the direction opposite to the movement of the replication fork, posing a logistical problem. The movement of the replication fork exposes additional template, but this new template lies ***upstream*** from the new strand. DNA on the lagging strand is therefore synthesized discontinuously, shown in **Fig. 1** as separate small fragments, called Okazaki fragments, hundreds or thousands of nucleotides in length – very small compared to the full DNA molecule.

7. Proofreading

   Replication of DNA depends on faithful pairing of A to T and G to C and not alternative pairings. The canonical ("correct") base pairs are more stable than the alternatives, but all occur with some frequency. In fact, the difference in stability is such that you would expect to find an erroneous basepair at a frequency of one every 100 nucleotides per replication.[1] Yet, the observed rate of mutation in human cells is about 1 error per genome every 5 cell divisions.[2] Since the human genome contains about 3 billion nucleotides, the error rate comes to about one every 10^{10} nucleotides per cell division, i.e. 10^{8} more accuracy than expected from pure basepairing.

   Where does that accuracy come from? Most DNA polymerases have the ability to detect rare cases when a wrong nucleotide has managed to get incorporated into the new strand, and this accounts for some of the increased accuracy. If DNA polymerase senses a change in DNA structure resulting from a mispairing, it backs up and lops off the last nucleotide, making the previous 3'-OH available. Then it tries again with another nucleoside triphosphate. RNA polymerase has an error rate orders of magnitude higher than DNA polymerase.[3]

SQ5. Why is it essential for DNA polymerase to have a very low error rate but not so important for RNA polymerase?

8. Removal of RNA primers
   RNA primers are removed (in prokaryotes) by the action of DNA polymerase I (a yellow blog in **Fig. 1**), which extends the DNA at the end of the leading strand or the end of Okazaki fragments, chewing up RNA ahead of it like a train chugging down a track tossing off a string of cows in front of it.

SQ6. Develop that cow/train metaphor. What would you identify with an Okazaki fragment? What is the cow? What is the train? What is the process?

9. Ligation of lagging strand fragments
   Once the RNA primers are removed, the lagging strand consists of contiguous DNA fragments. This is done by the enzyme DNA ligase (a blue blob in **Fig. 1**), which uses the energy of ATP to form a covalent bond between the 3'-OH of one fragment and the 5'-phosphate of the next.

By the end of the replication process, a single parental molecule consisting two contiguous strands have produced two progeny molecules consisting of two contiguous strands.

The two rules enunciated in Steps 3 and 4 are so important, I'll say them again:

Rule #1: Synthesis of DNA always proceeds by adding nucleotides to a pre-existing 3'-OH end on the strand to be synthesized

Rule #2: DNA is always made 5' to 3'

Where did these rules come from? It may be that they're just the accidental side effect of an arbitrary evolutionary path – maybe not the best solution, but we're stuck with it – but I don't think so. I present below some rationalizations for their existence.

Why does DNA synthesis in nature require an RNA primer?
Recall that base pairing by itself is far from sufficient to ensure faithful DNA replication. DNA polymerase is able to detect errors and correct them in part by the perturbations in the overall double helical structure introduced by mispairing. But this is possible only when there is an already existing double helical structure, and there isn't when DNA replication is beginning. Therefore, the first several nucleotides made are inherently untrustworthy. By making them as RNA, they are marked for later destruction once Okazaki fragments appears behind them. In short, RNA priming is a mechanism to avoid regions of error-prone replication.

Why is DNA replication always 5' to 3'?
Replication would be so much simpler if we didn't have to worry about the lagging strand, that is if replication on one strand proceeded 5' to 3' and on the other 3' to 5'. Why doesn't this happen? I think the reason can by extending the logic seen in **Fig. 2**. There you see that 5' to 3' synthesis is driven by the release of pyrophosphate from the 5' end of the nucleoside triphosphate and attachment of the remaining phosphate to the 3' end of the growing DNA chain. If DNA synthesis could proceed 3' to 5', then the 5' end of the growing DNA chain would have end with triphosphate, and the 3' end of the nucleoside triphosphate would attach itself to that end, splitting off pyrophosphate.

Well, what's wrong with that? If there are never errors, then nothing. But if DNA polymerase detects that an erroneous nucleotide has inserted itself, it will lop it off. In the case of 5'-to-3' replication, the old 3'-OH is exposed once more and synthesis can resume. But in the case of hypothetical 3'-to-5' replication, backing up and removing the offending nucleotide will expose a 5'-phosphate. There's no way to put the pyrophosphate back on. So replication will come to a halt. In short, 3'-to-5' replication is incompatible with proofreading, and that's a disaster.

[End of Part I]

References

1. Loeb LA, Springgate CF, Battula N (1974). Errors in DNA replication as a basis of malignant changes. Cancer Res 34:2311-2321. (http://cancerres.aacrjournals.org/content/34/9/2311.full.pdf+html)
2. Drake JW (1999). The distribution of rates of spontaneous mutation over viruses, prokaryotes, and eukaryotes. Annals NY Acad Sci 870:100-107 (http://onlinelibrary.wiley.com/doi/10.1111/j.1749-6632.1999.tb08870.x/abstract)
3. Sydow JF, Cramer P (2009). RNA polymerase fidelity and transcriptional proofreading. Curr Opin Struct Biol 19:732-739. (http://linkinghub.elsevier.com/retrieve/pii/S0959-440X%2809%2900162-6)