DNA POLYMERASE AND IN VIVO ANALYSIS OF REPLICATION

TEXTBOOKS/REVIEWS


PRIMARY LITERATURE


The prediction, based on the just-discovered structure of the double helix, that complementary base pairing underlies the replication process.


The demonstration–using sedimentation of density-labeled DNA to equilibrium in CsCl density gradients--that DNA replication is semi-conservative, as predicted by Watson and Crick. If you thought you knew this paper, check out the discussion questions.


First direct evidence of a moving locus of DNA synthesis at the replication fork.

Nearest-neighbor analysis to show that DNA synthesized by DNA Polymerase I was instructed by the template. Also gave first evidence that two strands are antiparallel. Analysis was rather arcane, but this is what you had to do before DNA sequencing was invented.


Demonstration that E. coli mutants that lack the DNA Pol I polymerase activity are still viable. I recommend this paper because Cairns is remarkable precise about the assumptions in his analysis and what he can and cannot conclude from his data. For example, near the end of his paper he says he cannot conclude that DNA Pol I is unnecessary for DNA replication, and that proves to be prescient, because although it is not the replicative polymerase, it is important for okazaki fragment maturation.


Discovery of Okazaki fragments by the Okazakis. But does the evidence make sense?


Independent evidence from electron microscopy convinces the doubters that the replication fork is asymmetric.


The first convincing evidence that short RNA oligonucleotides (RNA primers) are linked to the 5' end of Okazaki fragments.


Demonstration that 3' to 5' exonuclease activity could help DNA polymerase I proofread mistakes at the 3' end of the primer.


Minireview of the structure of DNA polymerases.


Overview of how polymerases maintain replication fidelity.
SOME QUESTIONS TO THINK ABOUT THE LECTURE

1) Formally DNA polymerases could have added nucleotides onto either (A) the 3’ OH end or (B) the 5’-phosphate end of a growing DNA strand. You are Arthur Kornberg, with no structural information about the enzyme. What type of modified nucleotide analog can you feed to the enzyme to distinguish between these two possibilities? How would you monitor the fate of this nucleotide? How would you determine whether any further normal nucleotides could be incorporated? What results would you expect for either possibility A or B?

2) How would you draw the free energy reaction diagram for DNA polymerization to illustrate the importance of pyrophosphate release in driving the reaction forward? What would you add to your polymerase reaction tube if you wanted to reverse the polymerization (assume the exonuclease activity has been destroyed by mutation)?

3) The basic principle of purifying a biochemical activity is straightforward: for each fractionation step account for where all the input activity ended up and where all the input protein ended up. Hopefully you will be able to identify fractions that have increased activity per protein (i.e. specific activity) relative to the input. Such accounting requires accurate quantitative assays for both activity and protein. Specifically you want to make sure that if your assay tube has 2-, 4-, or 10-fold more activity, that your assay readout increases by 2-, 4- or 10-fold, respectively, i.e. you are in the “linear range” of the assay for your activity. For an enzyme like DNA polymerase what should the assay concentration ranges be for its substrates (primer-template and nucleotides) relative to their Km, and why?

4) John Cairns started his search for DNA Pol I mutants with a suspicion that Pol I was not responsible for the bulk of replicative DNA synthesis. Suppose he had suspected the opposite? What type of mutants would he have needed to obtain? How would you show that you had mutated DNA Pol I activity? How would you show that this activity was essential for viability? Is that sufficient to establish that DNA Pol I activity is essential for cellular DNA replication, or do you need to do something further?

5) In the polymerization reaction cycle, nucleotide binding is considered a rapid reversible step relative to later irreversible catalytic steps. Hence, although there is not a true equilibrium between the nucleotide free and nucleotide bound states of the DNA polymerase, we can approximate their relationship by a pseudo dissociation constant. The nucleotide specificity of the DNA polymerase arises in part from the ratio between the dissociation constants for the correct versus the incorrect nucleotides. Why is this primarily determined by the ratios of the off rate constants for the correct versus incorrect nucleotides?

6) A central requirement for kinetic proofreading is that the discard pathway be irreversible. This requires some free energy input into the system, either coupled to the discard pathway itself or stored up from some earlier step. For DNA polymerase proofreading, where is the energy input that makes the exonuclease step irreversible?
7) How does the requirement for DNA polymerase to have primers contribute to its fidelity? How do you explain its contribution in terms of the kinetic parameters that determine fidelity?

8) The idea that the exonuclease activity of replicative DNA polymerases is important for replication fidelity was inferred from kinetic analysis of purified DNA polymerases. What experimental perturbations and readouts would be needed to show that this is indeed the case in cells?

9) Autoradiographic analysis of intact replicating E. coli DNA revealed large circular DNA with single bubbles of various sizes as shown in the lecture handout. Based on that type of data, what can you conclude about the following: (1) number of initiation sites per replication event, (2) position of initiation sites, (3) number of moving replication forks per replication event. What additional information would allow you to derive more conclusions about these issues?

10) Okazaki concluded that newly synthesized DNA that was small became converted into much larger DNA, i.e. that small DNA fragments (later named Okazaki fragments) were an intermediate in the DNA replication pathway. Why is this conclusion not really justified by his experimental design? How might you improve on his design?

11) Okazaki’s experiment is usually cited as one of the key evidence for semidiscontinuous DNA replication. But Okazaki himself concluded that DNA synthesis is fully discontinuous. Why did he conclude this from his results? Can you speculate why he got this result?

**POINTS RAISED IN LECTURE**

1) **Genetics and biochemistry** are both powerful entry points into the study of a biological problem. In the case of DNA replication, biochemistry dominated the early days. (In other cases—e.g. the development of the operon model for transcriptional regulation—genetics led the way.) As the replication field matured, the two approaches were used together to great advantage. For example, genetics indicated that DNA polymerase I was unlikely to be the primary replicative polymerase, and provided the mutants that allowed biochemists (specifically, our own Tom Kornberg) to uncover and purify DNA polymerase III, the true replicative polymerase. Biochemistry allows you to delineate the exact mechanism by which a biological task is performed. Genetics allows you to establish the functional relevance of a mechanism to the biological task of interest.

2) **Function is determined by phenotypes and assays.** The basic currency of genetics is the mutant phenotype; it allows one to define, observe, and follow mutations. The inability of a mutant to perform a certain task suggests that the wild-type version of the mutant protein normally carries out this task. The linchpin of biochemistry is the assay; it allows one to define, observe, and follow biochemical activities. Understanding the exact nature of the assay is important, since slight changes in the assay may change the nature of the activity being followed, and result in purification of different proteins or complexes of proteins. In well-studied systems this somewhat artificial distinction between genetic phenotypes and biochemical assays becomes blurred. For example, mutant cells can be phenotypically characterized by biochemical assays of extracts made from the cells. Another example is the precise mutagenesis of a specific activity of a protein to determine the mechanistic role of this activity in the protein’s function.
3) **Divide and Conquer.** The reductionist approach at the core of our scientific process is to break things down into their component parts, figure out how the parts work individually, then determine how the parts work together. This approach is difficult to pursue biochemically when fundamental functions are carried out by complex assemblies of subunits and are not readily assigned to separable individual subunits. Replication has been particularly amenable to this approach because specific functions can be attributed to individual proteins, even though those functions are often greatly enhanced in the presence of the correct neighbors or setting. In contrast, the role of individual protein or RNA molecules in the ribosome is currently best studied in the context of the full ribosome, using mutations that perturb specific proteins or RNAs.

4) **Structure informs but is not sufficient to establish function.** Based on the Watson-Crick structural model of DNA, it was predicted that DNA would replicate semi-conservatively. But actual proof required the Meselson-Stahl experiment. Also, the W-C model did not anticipate the idea of a replication fork, nor did it predict any features of the replication machinery. Later, EM studies on replication intermediates provided further structural insight leading to the idea of replication forks. However, it was Kornberg’s unswerving faith (yes, faith has a role in science) that proteins must carry out virtually all chemical reactions in the cell that paved the way for biochemical identification of replication proteins and eventual understanding of how those forks function.

5) **Crude extract assays may be quite complex.** Original DNA polymerization assay was not just assaying the polymerization reaction, but also the conversion of DNA to nucleosides, the conversion of nucleosides to nucleotides (i.e. phosphorylation by kinases), and the conversion of DNA to primer-template junctions. Fractionation of extract led to separation of these multiple activities and eventual discovery of the true substrate requirements for the polymerization reaction itself. Thus even though the initial assay had the wrong substrates and required multiple activities, having some detectable signal allowed Kornberg to tease apart the complexity and eventually focus on the activity of interest.

6) "**Purity is in the eye of the beholder**"-Nick Cozzarelli. There is no absolute standard for knowing that your activity is "pure". The 5% impurity in your 95% pure protein could be responsible for the activity you are following or could be associated with an additional activity you don’t yet realize you are assaying. The classic approach is to show that a protein(s) repeatedly and tightly cofractionates with your activity over many separation steps, giving you increasing (though never absolute) confidence that the protein is responsible for the activity.

Molecular biology has provided several powerful new ways to help associate an activity with a gene product(s), but does not provide a better method for assessing purity. An example where sophisticated molecular biology could not rescue a bad conclusion based on poor purification is provided by a high profile Science paper stating that yeast Trf4, which shows some sequence homology to DNA polymerases, has very weak DNA polymerase activity. Affinity purified Trf4 from E. coli had the activity, whereas affinity purified Trf4 mutated in conserved polymerase residues did not. Seems like a slam dunk, but couldn’t be reproduce and was later refuted. Presumably, the prep of affinity purified wild-type Trf4 was contaminated with a tiny amount of DNA polymerase, whereas the mutant prep was not. Hence, the classic approach to demonstrating purification is still highly relevant in this era of high-powered molecular biology.
Holoenzymes allow a core activity to be used in various settings and processes. Although some complex assemblies such as ribosomes are held tightly together, others—as exemplified by the proteins at the T4 replication fork—only loosely associate. Often many different degrees of association exist among proteins involved in the same process. For example, the catalytic core of DNA Pol III contains three tightly associated subunits. The gamma complex (clamp loader), which itself contains five tightly associated subunits, is bound less tightly to the catalytic core and often disassociates during purification.

The concept of a "holoenzyme" was initially introduced to incorporate the idea that large complexes containing loosely associated subunits that modify or regulate the core catalytic activity can be reproducibly purified (although usually under "gentle" conditions). These holoenzymes are thought to work as functional units in vivo. Thus, the complex containing two catalytic Pol III cores tethered to each other to one gamma complex via a tau protein dimer is called the Pol III holoenzyme and is thought to be the form of polymerase that works at the replication fork. Given the continuous range of association strengths found among proteins, the definition of a holoenzyme is somewhat arbitrary and now includes some complexes that cannot be purified together but are still thought to function as a unit (compare the concept of holoenzyme for eukaryotic RNA polymerases to that of DNA polymerases).

Importantly, the loose association of components in holoenzymes provide a modular way for these components to assemble into more than one type of holoenzyme so as to perform different functions. For example, only a small minority of the Pol III core in a cell is part of the Pol III holoenzyme and participates in DNA replication. The bulk of the Pol III core is likely associated with other proteins in holoenzymes involved in DNA repair.

Limitations of loss of function (LOF) genetic analyses. Loss of function genetic analysis is a powerful way to determine the function of genes and proteins. It uses the general scientific strategy of asking whether “A is required for B” by disrupting A and determining whether B is affected. If a wild-type function is impaired when a gene is disrupted, one can usually conclude that the gene is required in some manner for that function. However, things are not as straightforward when a function is not measurably affected, as there are several possible reasons for “false negative” results. For example, initial genetic analysis using the polA1 mutation may have missed a replicative role for DNA Pol I in okazaki fragment maturation because of three possible limitations of the analysis: (1) incomplete disruption of Pol I activity (2) weak redundant function provided by other polymerases (3) indirect and insensitive readout of cell viability instead of direct analysis of okazaki fragment maturation.

Replicative DNA polymerases are designed for extremely high fidelity. All replicative polymerases polymerize 5’ to 3’, are instructed by the template, require a primer, and use deoxynucleotide triphosphates as precursors. Almost all also possess a 3’ to 5’ exonuclease activity, which allow them to proofread their own mistakes. The requirement for a properly based paired primer in effect allows the polymerase to monitor whether an incorporation mistake has recently been made. The 5’ to 3’ polymerization direction ensures that the high energy bond for each polymerization step is contributed by the incoming monomer and not by the polymer (“tail” extension; contrast to “head” extension during translation). Hence, exo “deletion” of any polymerization mistake does not compromise the ability to continue polymerization (contrast this to protein translation).

Fundamentals of fidelity. Fidelity requires first the ability to distinguish between correct and incorrect reaction settings. Secondly, fidelity requires having a molecular choice to either advance a reaction (forward reaction) or abort the reaction (discard reaction). Third, fidelity requires that ability to shift the choice of reaction based on the setting that is sensed. Correct settings have to be preferentially
shunted to the forward reaction, and incorrect setting have to be preferentially shunted to the discard reaction.

For replicative DNA polymerases the important distinction in reaction setting is between having a correct or incorrect nucleotide opposite a specific template nucleotide. At the core of the ability to make this distinction is the polymerase’s induced fit requirement for catalysis. For example, only proper base pairing of the incoming nucleotide allows the conformational change that “fits” the polymerase snugly around the nucleotide-primer-template. This induced fit also couples the base pairing of the incoming nucleotide identity to the choice of reaction, because the fit allowed by the correct base pairing positions catalytic groups properly to promote polymerization. In contrast, an if the incoming nucleotide is incorrectly base-paired, the induced fit will be much harder, catalysis will be much slower, and the competing discard reaction of the nucleotide leaving the active site will be favored.

Fidelity is not always a desirable thing, since highly faithful polymerases will stall if they face damaged or unusual nucleotides in the template strand. In this setting no incoming nucleotide will be “correctly” base-paired with the template, as defined by the ability to satisfy a proper induced fit. In the past ten years a large class of nonreplicative polymerases with low fidelity have been identified. Many of them allow the cell to contend with damaged template nucleotides or promote mutagenesis when genetic diversity might be needed. These enzymes do not have a tight induced fit and/or have little or no 3’ to 5’ exonuclease activity.

11) **Kinetic manipulation of molecular choice**: The fact that the induced fit also requires proper base pairing of the primer nucleotides provides a second opportunity to test base-pairing, this time of the recently incorporated nucleotide. However, there is no intrinsic discard pathway, such as the diffusion of a free nucleotide away from the active site. So for replicative DNA polymerases which must maintain high fidelity, nature has designed a special discard pathway that aborts the extension from an incorrectly base-paired primer without aborting the entire polymerization process. This discard pathway involves the excision of the incorrectly base-paired primer nucleotide with a 3’>5’ exonuclease activity. In principle, both correctly and incorrectly incorporated nucleotides can be removed by this discard pathway or be extended further in a polymerization reaction. The key is to change the relative kinetics of these two reactions depending on whether the incorporated nucleotide is correct or incorrect. This relative kinetics determines whether the substrate preferentially partitions to the forward pathway of polymerization (correctly incorporated nucleotide) or the discard pathway of exonucleolytic cleavage (incorrectly incorporated nucleotide). Hence the term kinetic partitioning.

The separate and distant exonuclease catalytic site places the exonuclease reaction at an inherent disadvantage when the correct nucleotide is incorporated. Moreover, the proper primer-template base-pairing resulting from this incorporation allows the induced fit to facilitate rapid catalysis of the forward polymerization reaction. Thus, almost all the molecules in this setting are partitioned toward polymerization.

However, when an incorrect nucleotide is incorporated, the disrupted base pairing of the primer-template inhibits the induced fit of the polymerase and greatly slows the polymerization reaction. The mispairing at the primer-template junction also slightly increases the rate of exonuclease reaction. The net result is that the kinetic competition is shifted in favor of the discard pathway when an incorrect nucleotide is incorporated.

12) **Structural analysis of in vivo replication intermediates** established that: (1) replication was localized to moving replication forks; (2) synthesis at forks occurs in a semidiscontinuous manner; (3) discontinuous synthesis involves synthesis of short 100-2000 bp okazaki fragments that are ligated to
form the daughter strands on one side of the fork. These observations suggest the presence of a number of activities needed at a replication fork in addition to the replicative DNA polymerase.
Lecture 1: DNA Polymerase and the Replication Fork

Use of biochemistry (assays) and genetics (mutant phenotypes) to define function

Fidelity/Specificity: bioregulation through substrate control of molecular choice

Breaking down complex processes by structural analysis of intermediates
Using both Biochemistry and Genetics to understand function

**Biochemistry**
- **Assay** → **Activity (Function)**
  - Purify activity
  - Protein
  - Protein Interactions
  - Proteomics

**Genetics**
- **Mutant** → **Phenotype (Function)**
  - Reverse genetics (targeted gene disruption by HR, ZFN, TALEN, CRISPR/Cas9)
  - Expression interference (RNAi, siRNA, CRISPRi)
  - Gene
  - Genetic Interactions
  - Genomics
  - Homologs

Expression interference (RNAi, siRNA, CRISPRi)

Complementation or positional cloning

Expression or hybridization cloning

Protein expression and purification

Purify activity
DNA Replication: The Task and Challenge

**Speed:** very rapid duplication of every nucleotide
(ex: $6 \times 10^9$ bp in 8 hrs in humans)

**Fidelity:** extremely low error rate
($\sim 1/10^9$ nucleotide error rate)

**Count:** exactly two copies of every sequence per cell cycle

**Regulation:** coordination with other chromosomal events
(eg. mitosis, repair, recombination, transcription, chromatin packaging)
Enzymology of DNA Synthesis: DNA Polymerases

- **dNTP precursor**
  - pyrophosphate release provides energy

- Instructed by single-stranded template
  - senses complementarity of new nucleotide

- **Primer requirement**
  - senses complementarity of primer

- **5’ > 3’ polymerization off primer**
  - extension off 3’ hydroxyl
  - moving 3’ > 5’ on template

*enhances fidelity by allowing error correction*
In principle: Monitor incorporation of radioactive nucleotide precursors (◯) into acid insoluble form (physically separate product from precursor)

In practice: Can be difficult to devise the right assay conditions when you do not know the precise nature of the activity

Initial conditions used were really assaying a complex mixture of activities:

*E. coli* extract - source of polymerase activity but also kinase and nuclease activity

³H Thymidine - converted to thymidine triphosphate by kinases in extract

DNA - intended as nuclease decoy but nucleases convert to primer-template and source of A,G,C nucleotides

First Experiment: 50 out of 1 million cpm insoluble

Ten Years Later: purify DNA Polymerase I, show it is template-directed and figure out enzyme requirements
DNA Polymerase Structure and Catalysis

Crystal structure of bacteriophage T7 DNA Polymerase complexed with primer-template and dNTP

Structure resembles a right hand

Two Mg$^{++}$ ions positioned by conserved acidic residues catalyze reaction

Rest of enzyme positions primer-template and dNTP and ensures catalysis only occurs with proper “fit”
DNA Pol I has 3’ > 5’ Exonuclease Activity

Careful quantitative analysis of biochemical activity can suggest biological function

Exo Assay: 5’ T T T T* 3’
3’ AAAAAAAA A 5’

DNA Pol I no dTTP

5’ T T T T 3’
3’ AAAAAAAA A 5’

*exo activity is slow relative to pol activity

*exo activity is enhanced by stalling pol activity
or making 3’ end single-stranded
3’ mismatch generates both conditions

Proofread Assay:

5’ T T T T* 3’
3’ AAAAAAAA A 5’

DNA Pol I + dTTP

5’ T T T T T T T T T 3’
3’ AAAAAAAA A 5’

5’ T T T C* 3’
3’ AAAAAAAA A 5’

DNA Pol I + dTTP

5’ T T T T T T T T T 3’
3’ AAAAAAAA A 5’

C*

mismatch specific exo activity under normal pol conditions
both pol and exo activities are sensing primer-template pairing
The Polymerase and Exonuclease Activities of Replicative DNA Polymerases Reside in Distinct Domains

Movement between P and E sites requires primer-template unwinding translocation of 3’ end
DNA Pol I is not the replicative DNA polymerase in E. coli
Illustrates importance of genetics for establishing functional relevance in cell

Use biochemical assay to screen for mutants lacking DNA polymerase activity

Purification of residual polymerase activity from polA1 yields DNA Pol II and Pol III
Genetics and biochemistry later show:
- DNA Pol III is the replicative polymerase
- DNA Pol I is important for okazaki fragment maturation
Purification of DNA Pol III: Different Template, Different Assay, Different Activity

Introducing the concept of holoenzymes and modular enzyme subassemblies
## Fidelity Overview

### Fidelity Comparisons

<table>
<thead>
<tr>
<th>DNA Replication</th>
<th>Speed</th>
<th>Error Rate</th>
<th>Product Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Prokaryotes)</td>
<td>500 bp/sec</td>
<td>$10^{-9} - 10^{-10}$</td>
<td>$5 \times 10^6$ (E. coli)</td>
</tr>
<tr>
<td>(Eukaryotes)</td>
<td>50 bp/sec</td>
<td>$6 \times 10^9$ (humans)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$1 \times 10^{11}$ (lily)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RNA Transcription</th>
<th>Speed</th>
<th>Error Rate</th>
<th>Product Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30 bp/sec</td>
<td>$10^{-4}$</td>
<td>$10^3 - 10^6$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Protein Translation</th>
<th>Speed</th>
<th>Error Rate</th>
<th>Product Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20 aa/sec</td>
<td>$10^{-4}$</td>
<td>$10^2 - 10^3$</td>
</tr>
</tbody>
</table>

### Contributions to E coli DNA Replication Fidelity

<table>
<thead>
<tr>
<th>Fidelity</th>
<th>Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intrinsic Fidelity (polym)</td>
<td>$10^{-3} - 10^{-4}$  (sensing dNTP complementarity to template)</td>
</tr>
<tr>
<td>Exonuclease Proofreading (polym)</td>
<td>$10^{-2} - 10^{-3}$  (sensing primer complementarity to template)</td>
</tr>
<tr>
<td>Mismatch Repair (post polym)</td>
<td>$10^{-2}$  (sensing complementarity of two strands) (distinguishing parental and daughter strands)</td>
</tr>
<tr>
<td>Overall Replication Fidelity</td>
<td>$10^{-8} - 10^{-9}$</td>
</tr>
</tbody>
</table>
How to Distinguish Mismatch versus Correct Base Pair

Geometry From Crystal Structure

Sources of Polymerase Discrimination

H-bonding (binding energetics)

Outside the active site, unpaired nucleotides are H-bonded to H2O. Inside the active site these H-bonds can be replaced by WC base pairing but only incompletely replaced by mismatch pairing.

Mismatch H bonding can also exacerbate steric and stacking clashes (see below)

Steric Constraints (structure/geometry)

Imposed by enzyme’s “induced fit”, which can test for precise base pair geometry, proper base stacking, and correct primer-template fit.
**Intrinsic Fidelity:** Potential Base Pair Discrimination for dNTP at Three Stages Of the DNA Polymerization Reaction Cycle

Example: T7 DNA polymerase (arrow thickness roughly corresponds to rate constants)
Other polymerases discriminate differently at each stage

Rapid dNTP Binding
Pseudo-equilibrium

Slow Conformational Change
"Induced Fit"

Polymerization Reaction

\[
\frac{K_D^{c}}{K_D^{i}} \approx \frac{> 8 \text{mM}}{20 \mu \text{M}} \approx 400x
\]

\[
\frac{k_{conf}^{c}}{k_{conf}^{i}} \approx \frac{300 \text{ s}^{-1}}{0.3 \text{ s}^{-1}} \approx 1000x
\]

Rapid and Not Measured
**Error Correction:** Exonuclease activity allows the polymerase’s kinetic discrimination to lead to different primer fates.

Arrow thickness roughly corresponds to rate constant.

Fast reaction pathways for correct primer with correct nucleotide:

1. $E \cdot DNA^C_N$
2. $E \cdot DNA^C_{N+1} \cdot dNTP^C$
3. $E^* \cdot DNA^C_N \cdot dNTP^C$
   - $k_{pol}^c$
4. $E \cdot DNA^C_{N+2} \cdot PP_i$

Slow reaction pathways for incorrect primer with correct nucleotide:

1. $E \cdot DNA^I_{N+1}$
2. $E \cdot DNA^I_{N+1} \cdot dNTP^C$
3. $E^* \cdot DNA^I_{N+1} \cdot dNTP^C$
   - $k_{pol}^c$
4. $E \cdot DNA^I_{N+2} \cdot PP_i$
**Error Correction**: Kinetic manipulation of molecular choice based on complementarity of primer

Arrow thickness roughly corresponds to rate constant

When a **correct** nucleotide is incorporated, 3’>5’ exonuclease activity is much slower than 5’>3’ polymerase activity. Addition of the next nucleotide is kinetically favored.

When an **incorrect** nucleotide is incorporated, disruption of the primer greatly slows 5’>3’ polymerase activity for the next nucleotide (and slightly increases 3’>5’ exonuclease activity). Excision of the incorrect nucleotide is kinetically favored.
**Error Correction:** Kinetic manipulation of molecular choice based on complementarity of primer

Black arrow thickness roughly corresponds to relative rate constant
Light blue arrow thickness roughly corresponds to relative flux

Example:
T7 DNA Polymerase

\[
\begin{align*}
\frac{k_{\text{T7pol}}}{k_{\text{T7exo}}} & \sim 1500 \\
\frac{k_{\text{C primer}}}{k_{\text{T7pol}}} & \sim \frac{1}{230}
\end{align*}
\]
A Discard Strategy for Fidelity: Kinetic manipulation of molecular choice between irreversible forward and discard pathways

The choice is ultimately determined by the relative flux of molecules that proceed down the two competing pathways (light blue arrow).

Pathway irreversibility usually requires some chemical energy expenditure (e.g. dNTP hydrolysis), which could be coupled to either pathway or to a reaction step preceding these pathways.

For each substrate, the molecular flux (and hence molecular choice) is determined by the ratio of the forward to discard rate constants (black arrows) for that substrate. For cognate substrates this ratio should favor the forward reaction. For noncognate substrates, the ratio should “flip” to favor the discard pathway.

In principal, just one or both pathways could discriminate between cognate and noncognate substrates, i.e. change rate constants with substrate. In practice, nature often discriminates with both.
Intrinsic Fidelity + Error Correction

- $5'\text{-TTC} \ 3'\text{-AAGCTCA}$
- $5'\text{-TTC} \ 3'\text{-AAGCTCA}$

**Symbols:**
- $K_D$: pseudoequilibrium constant for nucleotide binding
- $k_{pol}$: composite rate constant for polymerization reaction
- $k_{exo}$: composite rate constant for exonuclease reaction

**DNA polymerase:**
- bound dNTP
- primer-template

**Intrinsic Fidelity + Error Correction**

- $k_{pol}$
- $K_D$
- $k_{pol}$
- $k_{pol}$
- $K_D$
- $k_{pol}$
- $k_{pol}$
- $K_D$
- $k_{pol}$
- $k_{pol}$
- $K_D$
How DNA Polymerases Check for Proper Base Pairing Geometry

Crystal structure suggestion for “Induced Fit”

DNA Polymerase contacts minor groove of primer-template

Base pair fit is “tested” before polymerization

Only W-C base pairs allow proper stacking
Induced fit positions nucleotide, primer 3’, metal ions

Base pair fit is still “tested” after polymerization

Purple: Interaction surface with DNA polymerase
Green: Universal H-bond acceptors
H-bonding with DNA polymerase
Most DNA polymerases in the cell have NON-replicative roles

Prokaryotic DNA Polymerases (E. coli)

<table>
<thead>
<tr>
<th>DNA Polymerase</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pol I</td>
<td>DNA Replication (RNA primer removal); DNA repair</td>
</tr>
<tr>
<td>Pol II (Din A)</td>
<td>DNA repair</td>
</tr>
<tr>
<td>Pol III holoenzyme</td>
<td>DNA Replication</td>
</tr>
<tr>
<td>Pol IV (Din B)</td>
<td>DNA repair; TLS; adaptive mutagenesis</td>
</tr>
<tr>
<td>Pol V (UmuC, UmuD') ₂C</td>
<td>TLS (translesion synthesis)</td>
</tr>
</tbody>
</table>

Eukaryotic DNA Polymerases

<table>
<thead>
<tr>
<th>DNA Polymerase</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pol α</td>
<td>DNA Replication (Primer Synthesis)</td>
</tr>
<tr>
<td>Pol β</td>
<td>Base excision repair</td>
</tr>
<tr>
<td>Pol γ</td>
<td>Mitochondrial DNA replication/repair</td>
</tr>
<tr>
<td>Pol δ</td>
<td>DNA Replication; nucleotide and base excision repair</td>
</tr>
<tr>
<td>Pol ε</td>
<td>DNA Replication; nucleotide and base excision repair</td>
</tr>
<tr>
<td>Pol θ</td>
<td>DNA crosslink repair</td>
</tr>
<tr>
<td>Pol ζ</td>
<td>TLS</td>
</tr>
<tr>
<td>Pol λ</td>
<td>Meiosis-associated DNA repair</td>
</tr>
<tr>
<td>Pol μ</td>
<td>Somatic hypermutation</td>
</tr>
<tr>
<td>Pol κ</td>
<td>TLS</td>
</tr>
<tr>
<td>Pol η</td>
<td>Error-free TLS past cyclobutane dimers</td>
</tr>
<tr>
<td>Pol ι</td>
<td>TLS, somatic hypermutation</td>
</tr>
<tr>
<td>Rev1</td>
<td>TLS</td>
</tr>
</tbody>
</table>

Many of the nonreplicative polymerases have *low fidelity* and are *error-prone* because they tolerate non-WC bp and lack 3’>5’ exo activity.

Low fidelity is needed to bypass template lesions that are stalling replication. Low fidelity may be used to increase genetic variation in special circumstances.
Semi-Discontinuous DNA Synthesis at Replication Forks

**Leading** daughter strand: polymerase moves continuously in same direction as replication fork

**Lagging** daughter strand: polymerase moves discontinuously in opposite direction as replication fork

Synthesis with discontinuous lagging strand pieces (okazaki fragments) requires repeated:

- **A.** priming
- **B.** replacement of primed sequence
- **C.** ligation

Okazaki fragment length: prokaryotes 1000 - 2000 nt; eukaryotes 100-200 nt
Detection and Analysis of In Vivo DNA Replication Intermediates

Replication is localized to forks (Autoradiograph E. coli DNA)

New DNA synthesis is small (pulse label then size)

Semi-discontinuous synthesis (EM replicating λ DNA)

EM can distinguish SS from DS DNA

SS DNA (lagging)

DS DNA (leading)

SS regions in trans, some interrupted by small DS segments consistent with discontinuous okazaki fragment synthesis only on strand of specific polarity (lagging)

Short (5-10 sec) pulse thought to represent initial nucleotide incorporation

Infer H (labeled) and L (unlabeled) from film grain density

New DNA is small (pulse label then size)

Infer H (labeled) and L (unlabeled) from film grain density

SS DNA (lagging)

DS DNA (leading)

SS DNA (lagging)

DS DNA (leading)

consistent with discontinuous okazaki fragment synthesis only on strand of specific polarity (lagging)
Replication Fork Tasks and the Activities That Perform Them

<table>
<thead>
<tr>
<th>Task</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>synthesize DNA</td>
<td>polymerase</td>
</tr>
<tr>
<td>separate parental strands</td>
<td>helicase</td>
</tr>
<tr>
<td>begin DNA synthesis</td>
<td>primase</td>
</tr>
<tr>
<td>replace primer</td>
<td>nuclease/polymerase</td>
</tr>
<tr>
<td>connect okazaki fragments</td>
<td>ligase</td>
</tr>
<tr>
<td>stabilize SS DNA</td>
<td>SSBP</td>
</tr>
<tr>
<td>ensure processivity</td>
<td>clamp loader/clamp</td>
</tr>
<tr>
<td>unlink parental strands</td>
<td>topoisomerase</td>
</tr>
</tbody>
</table>
APPENDIX

Bioreg 2015 Lecture 1 Replication
Using Biochemical Assays to Define Biochemical Functions

Assay must distinguish or physically separate products from substrates

- Polymerization: conversion of radioactive nucleotide from acid soluble to acid precipitable
- Nuclease: conversion of incorporated radioactive nucleotide from acid precipitable to acid soluble

Quantification is important for inferring biological relevance

- 3’ > 5’ exonuclease negates the polymerization reaction, but is generally much slower
- DNA Pol I’s poor polymerization raised the possibility that it was not the replicative helicase

Small differences in assay conditions can define different activities

- Gapped/nicked template defines Pol III core activity
- Primed single-stranded template defines Pol III holoenzyme activity

Complications of assaying crude extracts
(beware of wasting clean thoughts on dirty enzymes)

- Can be detecting multiple types of activities and be affected by multiple competing activities
- Can be detecting multiple similar activities
The Awesome Challenges of Genetics

*polA* mutant revisited

**Lecture 1:** *polA1* mutant with <1% assayable DNA Pol1 activity have relatively normal replication

Cairns concludes DNA Pol1 is not important for DNA replication

**Lecture 2:** DNA Pol1 plays a role in okazaki fragment maturation, an important part of replication

What happened to the awesome power of genetics?

### Caveats about gene analysis

<table>
<thead>
<tr>
<th>Caveats applied to <em>polA1</em> mutants</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Limitations in Phenotypic Analysis</strong></td>
</tr>
<tr>
<td><strong>Excess Activity/Leaky Allele</strong></td>
</tr>
<tr>
<td><strong>Redundancy</strong></td>
</tr>
</tbody>
</table>

With all these caveat, what is the evidence that DNA Pol1 is important for OF maturation and DNA replication?

*polA12* ts mutant accumulates increased OFs at restrictive temp (similar to the ts *lig4* mutant)

*polA12 lig4* double mutant not only accumulates OFs but rapidly ceases DNA synthesis at restrictive temp
General Strategies for Isolating DNA Synthesis Mutants

1) Screen: assay macromolecular synthesis \textit{in vivo} in ts-lethal mutants soon after shift to restrictive temp

- DNA synthesis \( - \) (\(^3\)H Thymine incorp.)
- RNA synthesis \(+\)
- Protein synthesis \(+\)
- Nucleotide synthesis \(+\)

2) Selective enrichment: ts mutants that fail to incorporate poisonous nucleotide analog during transient shift to restrictive temp

Can recycle survivors to further enrich

WT incorporate 5-bromouracil (5-BU) - UV sensitive
Replication mutants poorly incorporate 5-BU - survive

\(32^\circ\rightarrow 42^\circ\rightarrow 42^\circ\rightarrow 32^\circ\)

\(\text{wash out 5-BU}\)

\(\text{UV}\)
Mismatch Repair: Correction of Replication Errors

E. coli mismatch repair

MutS - recognizes mispaired bp by susceptibility to kinking
MutH - recognizes nearby GATC
MutL - association with MutS and MutH stimulates MutH to nick unmethylated daughter strand (basis of strand bias)

Exonuclease and helicase II, directed by MutS and MutL excise daughter strand from nick to mispaired bp
DNA polIII, clamp, clamp-loader, and SSB synthesize replacement DNA

MutS bound to mispaired DNA

DNA- both parental strands methylated at GATC sites
daughter strand transiently unmethylated after replication

Dam methylase eventually fully methylates GATC sites so both strands are marked as parental for next round of replication
Demonstrating 5′ → 3′ Chain Growth

Test: Use dideoxy nucleotide to ask if 3′ OH required for monomer incorporation?

Result: Dideoxy is incorporated but subsequent polymerization is blocked
Polymerization via Head Growth vs Tail Growth

Head Growth: activated high energy bond end of polymer drives polymerization

Tail Growth: activated high energy bond end of monomer drives polymerization

Figure 3-5
Mechanisms of tail growth and head growth in the biosynthesis of macromolecules. X and Y = sugar moieties; AA = amino acids; tRNA = transfer RNA of the indicated amino acid; ACP = acyl carrier protein.