Molecular Biology and Genetics

Prof. Mohammad El-Khateeb
Dr. Mamoun Ahram
Curriculum
(Part I: molecular biology)

- DNA and RNA structures
- Restriction endonucleases and their applications (RFLP and cloning)
- DNA replication
- DNA mutations and chromosomal anomalies
- Mechanisms of DNA repair
- DNA sequencing and PCR
- Transcription
- Translation
- Molecular biology of cancer
Curriculum

(Part II: genetics)

• Genetic Variations
• Chromosomes and Cell division
• Chromosomal Disorders
• Patterns of inheritance and genetic disorders
• Biochemical Genetics
• Genetics of cancer
• Multifactorial Inheritance
• Population genetics
• Population Screening for Genetic Diseases
• Prevention and Treatment of Genetic Disease
References

• Lectures

• Part I: Molecular biology
  – Molecular Cell Biology (http://bc.swhfreeman.com/lodish5e/)
  – The Medical Biochemistry Page: (http://themedicalbiochemistrypage.org/)

• Part II: Genetics
  – Medical Genetics, Authors: Jorde, Carey, Bamshad, White, Published by: Mosby.
  – Emery’s Elements of Medical Genetics, Authors: Peter Turnpenny, Sian Ellard, Published by Churchill Livingstone.
  – Medical Genetics, Authors: Thompson, Mcieans, Willard.
“Dr. Ahram may I be excused? My brain is full.”
Introduction & Nucleic acid structure
Molecular biology

The biochemistry of genetics
NUCLEIC ACID STRUCTURES
Resources

• This lecture
• Campbell and Farrell’s Biochemistry, pp. 227-242, 355-357, 381-382
DNA and RNA are polymers

- They are linear polymers composed of monomers called nucleotides
Chemical composition and bonds

• All nucleotides have a common structure:
  – a phosphate group linked by a phosphoester bond to a pentose
  – The pentose is linked to a nitrogenous base via a glycosidic bond
DNA vs. RNA

- In RNA, the pentose is a ribose; in DNA, it is a deoxyribose.
Nitrogenous bases

• DNA and RNA each consists of only four different nucleotides of two classes: purines and pyrimidines
  – Purines are adenine and guanine
  – Pyrimidines are cytosine, thymine, and uracil

• The bases are often abbreviated A, G, C, T, and U, respectively
Nucleotides are acidic
Nucleotides vs. Nucleosides

- Nucleoside
- Nucleoside monophosphate
- Nucleoside diphosphate
- Nucleoside triphosphate

pentose

Base

glycosidic bond

OH = ribose
H = deoxyribose
<table>
<thead>
<tr>
<th>Base</th>
<th>Nucleoside ( = base + pentose)</th>
<th>Nucleotide ( = nucleoside + phosphate)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ribonucleoside</td>
<td>Deoxyribonucleoside</td>
</tr>
<tr>
<td>Purines</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adenine</td>
<td>Adenosine</td>
<td>Deoxyadenosine</td>
</tr>
<tr>
<td>Guanine</td>
<td>Guanosine</td>
<td>Deoxyguanosine</td>
</tr>
<tr>
<td>Pyrimidines</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cytosine</td>
<td>Cytidine</td>
<td>Deoxyctydine</td>
</tr>
<tr>
<td>Thymine</td>
<td>Thymidine</td>
<td>Deoxythymidine</td>
</tr>
<tr>
<td>Uracil</td>
<td>Uridine</td>
<td>Deoxyuridine</td>
</tr>
<tr>
<td>Base</td>
<td>Nucleoside</td>
<td>Nucleotide</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------------------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td><strong>Purines</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adenine</td>
<td>Adenosine</td>
<td>Adenylate</td>
</tr>
<tr>
<td></td>
<td>Deoxyadenosine</td>
<td>Deoxyadenylate</td>
</tr>
<tr>
<td>Guanine</td>
<td>Guanosine</td>
<td>Guanylate</td>
</tr>
<tr>
<td></td>
<td>Deoxyguanosine</td>
<td>Deoxyguanylate</td>
</tr>
<tr>
<td><strong>Pyrimidines</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cytosine</td>
<td>Cytidine</td>
<td>Cytidylate</td>
</tr>
<tr>
<td></td>
<td>Deoxycytidine</td>
<td>Deoxycytidylate</td>
</tr>
<tr>
<td>Thymine</td>
<td>Thymidine or deoxythymidine</td>
<td>Thymidylate or deoxythymidylate</td>
</tr>
<tr>
<td>Uracil</td>
<td>Uridine</td>
<td>Uridylate</td>
</tr>
</tbody>
</table>
(a) Deoxyribonucleotides

Nucleotide:  
- Deoxyadenylate (deoxyadenosine 5'-monophosphate)  
- Deoxyguanylate (deoxyguanosine 5'-monophosphate)  
- Deoxythymidylate (deoxythymidine 5'-monophosphate)  
- Deoxycytidylate (deoxycytidine 5'-monophosphate)

Symbols:  
- A, dA, dAMP  
- G, dG, dGMP  
- T, dT, dTMP  
- C, dC, dCMP

Nucleoside:  
- Deoxyadenosine  
- Deoxyguanosine  
- Deoxythymidine  
- Deoxycytidine
(b) Ribonucleotides

<table>
<thead>
<tr>
<th>Nucleotide:</th>
<th>Adenylate (adenosine 5′-monophosphate)</th>
<th>Guanylate (guanosine 5′-monophosphate)</th>
<th>Uridylate (uridine 5′-monophosphate)</th>
<th>Cytidylate (cytidine 5′-monophosphate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symbols:</td>
<td>A, AMP</td>
<td>G, GMP</td>
<td>U, UMP</td>
<td>C, CMP</td>
</tr>
<tr>
<td>Nucleoside:</td>
<td>Adenosine</td>
<td>Guanosine</td>
<td>Uridine</td>
<td>Cytidine</td>
</tr>
</tbody>
</table>

20
Nucleic acid polymer

- The hydroxyl group attached to the 3’ carbon of a sugar of one nucleotide forms a phosphodiester bond to the phosphate attached to 5’ carbon of another nucleotide

- Directionality
DNA structure

- A double helix
  - Complementary strands
- Antiparallel
- Backbone vs. side chains
- Groovings
- Specific base-pairing
  - Chargaff's rules
- Stable
- Flexible
Some distances and numbers
DNA-protein interaction
Chargaff's rules

- In addition, Erwin Chargaff established certain rules about the amounts of each component of DNA:
  - Pyrimidines (T + C) always equal purines (A + G)
  - T always equals A
  - C always equals G
  - A + T is not necessarily equal to G + C
Flexibility
DNA forms

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helix sense</td>
<td>Right handed</td>
<td>Right-handed</td>
<td>Left handed</td>
</tr>
<tr>
<td>Mean bp/turn</td>
<td>10.7</td>
<td>10.0</td>
<td>12</td>
</tr>
<tr>
<td>Rise/bp along axis</td>
<td>2.3Å</td>
<td>3.32Å</td>
<td>3.8Å</td>
</tr>
<tr>
<td>Pitch/turn of helix</td>
<td>24.6Å</td>
<td>33.2Å</td>
<td>45.6Å</td>
</tr>
<tr>
<td>Diameter</td>
<td>26Å</td>
<td>20Å</td>
<td>18Å</td>
</tr>
</tbody>
</table>
B-form

- Most common DNA conformation in vivo
- Narrower, more elongated helix than A
- Wide major groove easily accessible to proteins
- Narrow minor groove
- Favored conformation at high water concentrations
- Base pairs nearly perpendicular to helix axis
A-DNA

• Most RNA and RNA-DNA duplex
• Shorter, wider helix than B
• Deep, narrow major groove not easily accessible to proteins
• Wide, shallow minor groove accessible to proteins
• Favored at low water concentrations base pairs tilted to helix axis
Z-DNA

- Left-handed
- Zigzag backbone
- Narrower, more elongated helix than A or B.
- Narrow minor groove
- Conformation favored in alcohol or high salt solution
- Requires alternating purine-pyrimidine sequence
RNA structure

- Note: RNA is single-stranded and does not have a specific structure
The order of bases of a DNA fragment
DNA homology
Humans and others

Pig: MPVERMRMRPWLEEQINSNTIEGLKULNKEKKIFQIPWMHAARHGWDVEKDALFRNWAIHTGKHOPGVDKLP
Human: MPVERMRMRPWLEEQINSNTIEGLKULNKEKKIFQIPWMHAARHGWDVEKDALFRNWAIHTGKHOPGVDKLP
Mouse: MPVERMRMRPWLEEQINSNTIEGLKULNKEKKIFQIPWMHAARHGWDVEKDALFRNWAIHTGKHOPGVDKLP
Norway Rat: MPVERMRMRPWLEEQINSNTIEGLKULNKEKKIFQIPWMHAARHGWDVEKDALFRNWAIHTGKHOPGVDKLP
Chinese Perch: MPVERMRMRPWLEEQIDSQIEGLKVNKEKRIFQIPWMHAARHGWDLEKDALFMRWAIHTGKYQPGIDKLP
Rainbow Trout: MPVERMRMRPWLEEQINSSSLIRGLINTINREKRIFQIPWMHAARHGWDLEKDALFMNWAIHTGKYQLGLIDKLP
Rhesus Monkey: MPVERMRMRPWLEEQINSNTIEGLKULNKEKKIFQIPWMHAARHGWDVEKDALFRNWAIHTGKHOPGVDKLP
WHAT IS THE ORIGIN OF MAN?
Theories

Homo sapiens

100,000 years ago

Homo sapiens

1.8 million years ago

Homo neanderthalensis

Hypotheses:
- Multiregional Origin
- Out of Africa

Theories:
- Multiregional Origin
- Out of Africa

Asian Homo erectus

European Homo erectus

Homo erectus

African Homo erectus

100,000 years ago

1.8 million years ago
Transmission of parental-specific DNA
African origin

The map illustrates migration routes of African origin with estimated numbers such as 40,000, 67,000, 20,000, 13,000, 130,000, and a range of 40,000-60,000.
The tree of life
وقد خلقكم أطواراً ٤ ١
Inter-species comparisons
The Arabian Cradle: Mitochondrial Relicts of the First Steps along the Southern Route out of Africa

Verónica Fernandes,1,2 Farida Alshamali,3 Marco Alves,1 Marta D. Costa,1,2 Joana B. Pereira,1,2 Nuno M. Silva,1 Lotfi Cherni,4,5 Nourdin Harich,6 Viktor Cerny,7,8 Pedro Soares,1 Martin B. Richards,2,9,11 and Luísa Pereira1,10,11,*

A major unanswered question regarding the dispersal of modern humans around the world concerns the geographical site of the first human steps outside of Africa. The “southern coastal route” model predicts that the early stages of the dispersal took place when people crossed the Red Sea to southern Arabia, but genetic evidence has hitherto been tenuous. We have addressed this question by analyzing the three minor west-Eurasian haplogroups, N1, N2, and X. These lineages branch directly from the first non-African founder node, the root of haplogroup N, and coalesce to the time of the first successful movement of modern humans out of Africa, ~60 thousand years (ka) ago. We sequenced complete mtDNA genomes from 85 Southwest Asian samples carrying these haplogroups and compared them with a database of 300 European examples. The results show that these minor haplogroups have a relict distribution that suggests an ancient ancestry within the Arabian Peninsula, and they most likely spread from the Gulf Oasis region toward the Near East and Europe during the pluvial period 55–24 ka ago. This pattern suggests that Arabia was indeed the first staging post in the spread of modern humans around the world.
Human Genome Project

• 1989, a project known as the Human Genome Project was launched

• The main purpose of the project is to sequence and determine the 2 billion nucleotides in the human genome

• This project was completed in 2006
Goals of HGP

• Identify the human genes and their position on chromosomes (~20,000-30,000)
• Determine gene/protein function
• Identify disease-related genes
• Determine human genetic variations
• Determine genetic predisposition to disease (biomarkers)
What is a biomarker?

• A substance used as an indicator of a biologic state
In medicine...

- A molecule that allows the detection of a particular cell type
- A fragment of DNA sequence that causes disease or is associated with susceptibility to disease (genetic marker)
- A protein expressed or not expressed as a result of a disease state
- A substance that is introduced in an organism to examine health and organ function
- A substance whose presence indicates a particular disease state (for example, the presence of an antibody may indicate an infection)
- An indicator of exposure to various environmental substances (toxicology)
Denaturation

- The two strands of DNA can be physiologically or experimentally separated (denaturation)
- The two complementary strands can re-associate into a perfect double helix
The melting temperature ($T_m$)

- The melting temperature, $T_m$, is the temperature where 50% of double helical DNA is separated.
Factors influencing $T_m$

- G·C pairs
- Ion concentration
- Salt concentration
- Destabilizing agents (alkaline solutions, formamide, urea)
The graph illustrates the relationship between temperature and the optical density (OD) at 260 nm (OD$_{260}$) for single stranded DNA and double stranded DNA. The graph shows two curves:

- The curve for single stranded DNA (top curve) has a melting temperature ($T_m$) of 75 °C, indicating a relatively low GC content.
- The curve for double stranded DNA (bottom curve) has a $T_m$ of 85 °C, suggesting a relatively high GC content.

The graph is labeled with the temperature axis (Temperature (°C)) ranging from 65 to 95 °C, and the OD$_{260}$ axis ranging from 0 to 1.0.
Hybridization

- DNA from different sources can form double helix as long as their sequences are compatible (hybrid DNA)
Techniques

• Gel electrophoresis
• Hybridization techniques
Gel electrophoresis

- The length and purity of DNA molecules can be accurately determined by the gel electrophoresis.
START WELLS FILLED WITH SAMPLE
Resources

• http://www.personal.psu.edu/pzb4/electrophoresis.swf

• http://www.sumanasinc.com/webcontent/animations/content/gelelectrophoresis.html

• http://www.sumanasinc.com/webcontent/animations/content/gelelectrophoresis.html
Detection

- The DNA molecules of different lengths will run as "bands".

- DNA is stained (that is, colored) with a dye (ethidium bromide) or radioactively labeled ($^{32}\text{P}$).
Hybridization techniques

- Hybridization reactions can occur between any two single-stranded nucleic acid chains provided that they have complementary nucleotide sequences.

- Hybridization reactions are used to detect and characterize specific nucleotide sequences.
Probes

- A probes is a short sequence of single stranded DNA (an oligonucleotide) that is complementary to a small part of a larger DNA sequence

- Hybridization reactions use labeled DNA probes to detect larger DNA fragments
Southern blotting

- This technique is a combination of DNA gel electrophoresis and hybridization

- Used to detect:
  - the presence of a DNA segment complementary to the probe
  - the size of the DNA fragment
DNA

Separate DNA on an Agarose Gel

Transfer or BLOT DNA fragments from GEL to Membrane

Membrane with DNA bands transferred to it

Radiolabeled probe incubated with Membrane

Bound DNA Bands are Exposed on Film
Restriction endonucleases, RFLP, and gene cloning
Resources

- This lecture
- Campbell and Farrell's Biochemistry, pp. 357-367
Endonucleases

- Enzymes that degrade DNA within the molecule rather than from either end (exonucleases)
Restriction endonucleases

- Enzymes that recognize and cut (break) the **phosphodiester bond** between nucleotides at specific sequences (4- to 8-bp **restriction sites**) generating **restriction fragments**
Biological purpose of restriction endonucleases

- Present in bacteria
- Degrade foreign DNA (restricted growth)
Type II restriction endonucleases

• **Always** cleave always at the same place
  – *EcoRI* (isolated from *E. coli*) cuts at 5'-GAATTC-3'

• **Give the same** set of fragments

• **Some enzymes cut DNA at related sites**
  – *HinfI* (from Haemophilus influenzae) recognizes 5'-ANTC-3' ('N' is any nucleotide)
  – Cuts at 5'-AATC-3', 5'-'ATTC-3', 5'-'AGTC-3' and 5'-ACTC-3'
Types of cleavages

• Restriction enzymes cut DNA in two different ways:
  – Blunt
  – Staggered (off-center)
Blunt end

- Cut at the **same position on both strands** giving a blunt ended fragments
Sticky or cohesive ends

- Cut the two DNA strands at different positions
  - The DNA fragments have short single-stranded overhangs at each end
- Called sticky or cohesive ends
5’ vs. 3’ overhangs

(B) 5’ and 3’ overhangs

\[
\begin{align*}
5' \quad \text{GGATCC} \quad \text{CCTAGG} \quad 3' \\
3' \quad \text{GCTAG} \quad 5'
\end{align*}
\]

\[
\begin{align*}
5' \quad \text{CTGCAG} \quad \text{GACGTC} \quad 3' \\
3' \quad \text{GCTAG} \quad 5'
\end{align*}
\]

BamHI

\[
\begin{align*}
5' \quad \text{GATCC} \quad 3' \\
3' \quad \text{GCTAG} \quad 5'
\end{align*}
\]

PstI

\[
\begin{align*}
5' \quad \text{CTGCA} \quad 3' \\
3' \quad \text{ACGTC} \quad 5'
\end{align*}
\]
Palindromic sequence

EcoRI
5' GAATTC 3'
3' CTTAAG 5'

HindIII
5' AAGCTT 3'
3' TTCGAA 5'

SmaI
5' CCCGGG 3'
3' GGGCCC 5'

TaqI
5' TCGA 3'
3' AGCT 5'
DNA ligase

• Covalently joins DNA ends (example, restriction fragments)
• Catalyzing the formation of 3’→ 5’ phosphodiester bonds between the 3-hydroxyl end of one strand and the 5-phosphate end of another strand
Advantage of restriction endonucleases

• Restriction fragment length polymorphism (RFLP)

• Cloning
DNA polymorphisms

- Individual variations in DNA sequence may create or remove restriction-enzyme recognition sites generating different restriction fragments

- Remember: we are diploid (alleles can be homozygous or heterozygous)

- What is an allele?
Polymorphic restriction site

DNA (Allele 1)  \[ \ast \]  DNA (Allele 2)

Add the restriction endonuclease

4 fragments  3 fragments
Restriction fragment length polymorphism

• The presence of different DNA forms in individuals generates a restriction fragment length polymorphism, or RFLP
Detection of RFLP

- Gel electrophoresis
- Southern blotting
Example

**Variant 1**
*EcoRI* does not cut

**Variant 2**
*EcoRI* does cut

```
GCCGCCATTCTA
CGGCCTTAAGAT
```

```
GCCGAATTCTA
CGGCCTTAAGAT
```

1  2-1  2  Phenotype
RFLP in the clinic

- RFLP can be used as diagnostic tools

- For example, if a mutation that results in the development of a disease also causes the generation of distinctive RFLP fragments, then we can tell
  - if the person is diseased as a result of this mutation
  - from which parent this allele is inherited
Disease detection by RFLP

Molecule I

BamHI 5 kb BamHI 4 kb BamHI
GGATCC GGATCC Probe GGATCC

Molecule II

BamHI 9 kb BamHI
GGATCC GGGTCC Probe GGATCC

Digest DNA with BamHI

Analyze by Southern blot hybridization with probe for polymorphic region

9 kb

4 kb
Think!! What would you see in a gel?
Example 1: Disease detection by RFLP (sickle cell anemia)
Example 2: Disease detection by ASO (Cystic fibrosis)

ASO: Allele-specific oligonucleotide

Cystic Fibrosis allele $\Delta S 08$ has 3bp deletion [AGA]

ASO for normal DNA $5'$ CACCAA[AGA]GATATTTC-3'$
ASO for DNA sequence of $\Delta S 08$ mutation $5'$ CACCAATGATATTTC-3'$

<table>
<thead>
<tr>
<th>Normal ASO</th>
<th>DS08 ASO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heterozygous</td>
<td>Heterozygous</td>
</tr>
<tr>
<td>Heterozygous</td>
<td>CF</td>
</tr>
<tr>
<td>Heterozygous</td>
<td>Homozygous normal</td>
</tr>
</tbody>
</table>

A. Negative  B. Positive  C. Negative
Example 3: Paternity testing
Example 4: Forensics
Molecular cloning
Recombinant DNA technology

- Cloning means that you make several copies of one thing
- How?
- insert a DNA fragment of interest into a DNA carrier (called a vector)
- The result is what is known as a recombinant molecule
Using plasmids as vectors

- Bacterial plasmids are considered excellent vectors

- Most plasmid vectors contain at least three essential parts required for DNA cloning:
  - Can replicate
  - Can be selected for/against by an internal drug-resistance gene (selectable marker)
  - Can inset a foreign DNA fragment
Making of recombinant DNA

- Both DNA fragments (the DNA to be cloned and a vector) are cut by the same restriction endonuclease that makes sticky-ended DNA fragments.
- When mixed, they will bind to each other.
**Diagram: Gene Cloning Process**

1. **Plasmid Vector** + **DNA Fragment to be Cloned** → **Recombinant Plasmid**

2. **Mix E. coli cells with plasmids in presence of CaCl₂**
   - Culture on nutrient agar plates containing ampicillin

3. **Bacterial chromosome**
   - **Transformed E. coli cell survives**
   - **Cells that do not take up plasmid die**

   **Independent plasmid replication**
Cell clones

Cloned DNA

Cell multiplication

Colony of cells each containing copies of the same recombinant plasmid
A benefit of cloning

• Production of therapeutic proteins
  – Insulin and growth hormone
  – Vaccines
DNA replication
a general mechanism
Resources

- This lecture
- Campbell and Farrell's Biochemistry, Chapter 10
What is a genome?

• The entire DNA content of the cell is known as genome

• In general, the more complex the organism, the larger its genome is
Chromosomes

• DNA is organized into chromosomes

• Bacterial genome: usually one and circular chromosome

• Eukaryotic genome: multiple, linear chromosomes complexed with proteins known as histones
Transfer of molecular information

DNA replication → Transcription → RNA replication → Translation → PROTEIN

Reverse transcription
Different suggestions on possible mode of DNA replication

Conservative

Semiconservative ✓

Dispersive

New DNA

Original DNA
Bidirectionally...speaking

- Replication moves progressively along the parental DNA double helix bidirectionally
Replication fork

- Because of its Y-shaped structure, this active region is called a replication fork.
New DNA (long vs short)

- A long strand and shorter pieces (Okazaki fragments) of DNA are present at the growing replication fork.
COMPONENTS OF DNA REPLICATION
RNA primer

- In order for the DNA polymerase to initiate replication, it requires a RNA primer, to be added first complementary to the DNA template.
- It is synthesized by a primase.
1. Primase synthesizes short RNA oligonucleotides (primer) copied from DNA.

2. DNA polymerase III elongates RNA primers with new DNA.

3. DNA polymerase I removes RNA at 5’ end of neighboring fragment and fills gap.

4. DNA ligase connects adjacent fragments.
DNA helicases and SSB proteins

• For DNA synthesis to proceed, the DNA double helix must be opened up ahead of the replication fork

• Opening up the DNA is done by two types of protein contribute to this process
  – DNA helicases and
  – single-strand DNA-binding proteins
DNA helicases

- DNA helicases use ATP to open up the double helical DNA as they move along the strands.
Protein complexes

- In prokaryotes, the complex of primase, helicase, and other proteins is known as primosome.
- The complex of primosome and polymerase is known as replisome.
Single-strand DNA-binding (SSB) proteins

- Single-strand DNA-binding (SSB) proteins bind tightly to exposed single-stranded DNA strands without covering the bases, which therefore remain available for templating.

- These proteins:
  - prevent the formation of the short hairpin structures
  - protect single-stranded DNA from being degraded
  - aid helicases by stabilizing the unwound, single-stranded conformation
DNA topoisomerases

• A swivel is formed in the DNA helix by proteins known as DNA topoisomerases

• A DNA topoisomerase breaks then re-forms phosphodiester bonds in a DNA strand
DNA topoisomerase I

- Topoisomerase I produces a transient single-strand break (or nick)

- ATP-independent
DNA topoisomerase II

• Topoisomerase II is responsible for untangling chromosomes
  – also known as gyrase in bacteria

• It makes a transient double-strand break

• The protein uses ATP
Topoisomerase

Replication fork movement

Helicase
Next Okazaki fragment will start here.
RNA primer

Primase

Clamp

Leading strand
DNA polymerase III dimer

RNA primer
Okazaki fragment
Single-strand binding proteins
DNA polymerase I
Ligase
Lagging strand
How accurate is DNA replication?

- The frequency of errors during replication is only one incorrect base per $10^9$ to $10^{10}$ nucleotides incorporated.

- Why is fidelity high?
  - Hydrogen base-pairing is highly stable between G and C and between A and T.
  - The DNA polymerase can catalyze the formation of phosphodiester bonds when the right hydrogen bonding takes place between the correction bases.
Proofreading mechanism

- A 3’→5’ proofreading exonuclease activity
DNA polymerases in prokaryotes

• DNA polymerase III: DNA polymerization at the growing fork in E. coli

• DNA polymerase I:
  – 5′-to-3′ exonuclease activity (removal of RNA primer) of each Okazaki fragment.
  – Fills in the gaps between the lagging-strand fragments.
  – DNA repair

• DNA polymerase II, IV, and V: DNA repair
DNA polymerase III

- The DNA polymerase III is a very large protein composed of 10 different polypeptides

- The core polymerase is composed of three subunits:
  - $\alpha$ subunit contains the active site for nucleotide addition
  - $\varepsilon$ subunit is a 3-to-5 exonuclease that removes incorrectly added (mismatched) nucleotides from the end of the growing chain
  - $\theta$ subunit stimulate the exonuclease activity
β subunit

- The β subunit, forms a clamp around DNA and holds the catalytic core polymerase near the 3 terminus of the growing strand.
Replication Origin in bacteria

- Bacterial replication origin known as OriC is an ≈240-bp DNA segment

- oriC regions contain repetitive 9-bp and AT-rich 13-bp sequences, referred to as 9-mers (dnaA boxes) and 13-mers, respectively
Components of OriC

- 9-mers: binding sites for the DnaA protein

- 13-mers: AT-rich region
  - facilitates separation of the double strand DNA
Possible mechanism

- When DnaA protein binds to 9-mers, it applies stress on the AT-rich region resulting in DNA "melting"
Initiation of replication in E. coli

• The initiator proteins bind to replication origin forming a large protein-DNA complex that opens up the AT-rich region exposing a single-strand DNA

• The DNA helicase binds to the single-strand DNA

• The DNA primase joins the helicase, forming the primosome, which makes an RNA primer that starts the first DNA chain

• This leads to the assembly of the remaining proteins to create two replication forks

• These proteins continue to synthesize DNA
Two replication forks

- The two replication forks proceed in opposite directions until they meet up roughly halfway around the chromosome.
Origins of replication in human genome

• 20,000 to 100,000 origins of replication along the 46 chromosomes
(a) Origin of replication

Growth

(b) Replication beginning at three origins

Chromosome DNA

Sister chromatids DNA replicas (daughter molecules)
Initiation of DNA replication in eukaryotes
DNA polymerase in eukaryotes

- Eukaryotic cells contain 9 DNA polymerases: α, β, γ, δ, ε, ζ, κ, η, ι

<table>
<thead>
<tr>
<th>Table 10.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Biochemical Properties of Eukaryotic DNA Polymerases</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Mass (kDa)</td>
</tr>
<tr>
<td>Native</td>
</tr>
<tr>
<td>Catalytic core</td>
</tr>
<tr>
<td>Other subunits</td>
</tr>
<tr>
<td>Location</td>
</tr>
<tr>
<td>Associated functions</td>
</tr>
<tr>
<td>3' → 5' exonuclease</td>
</tr>
<tr>
<td>Primase</td>
</tr>
<tr>
<td>Properties</td>
</tr>
<tr>
<td>Processivity</td>
</tr>
<tr>
<td>Fidelity</td>
</tr>
<tr>
<td>Replication</td>
</tr>
<tr>
<td>Repair</td>
</tr>
</tbody>
</table>
Roles of different polymerases

• Polymerase β may function primarily in the repair of DNA damage

• ζ, κ, η, and τ DNA polymerases
  – Replication and repair of damaged DNA
  – No 3' to 5' exonuclease activity
Eukaryotic complexity

• There are more protein components in eukaryotic replication machines than there are in the bacterial analogs
  – The eukaryotic single-strand binding (SSB) protein is formed from three subunits, whereas only a single subunit is found in bacteria
More on eukaryotic complexity

• The DNA primase is incorporated into DNA polymerase α

• The polymerase α begins each Okazaki fragment on the lagging strand with RNA and then extends the RNA primer with a short length of DNA, before passing the DNA to DNA polymerase δ

• DNA polymerase δ then synthesizes the remainder of each Okazaki fragment
...and more

• The polymerases do not have a 5’→3’ exonuclease
• The primer is removed by two special enzymes, FEN-1 and RNaseH1
• DNA polymerase δ then fills in the gap
Role of chromatin

• It seems that the timing of replication is related to the packing of the DNA in chromatin

• In eukaryotes, chromosome duplication requires that DNA is freed from histones and that new chromosomal histones be assembled onto the DNA behind each replication fork
Chromatin remodeling

- As a replication fork advances, it must pass through the parental nucleosomes.
- The chromatin-remodeling proteins free DNA from histones in order for enzymes to move along the DNA.
- The addition of histones to the newly synthesized DNA is aided by chromatin assembly factors (CAFs), which are proteins that package the newly synthesized DNA.
Role of telomeres

• As the growing fork approaches the end of a linear chromosome, the lagging-strand template is not completely replicated.

• When the final RNA primer is removed, there is no place onto which DNA polymerase can build to fill the resulting gap.

• This would lead the lagging-strand to be shortened at each cell division.
First round of replication

Replication

End of chromosome

RNA primer removal

Second round of replication

One chromosome is shorter than the other.

RNA primer removal

Primer gap
Telomerase

• The enzyme that prevents this progressive shortening of the lagging strand is a modified enzyme called telomerase, which can elongate the lagging-strand template
The mechanism

• Telomere DNA sequences consist of many repeats of a short sequence
• In humans, this sequence is GGGATT, extending for about 10,000 nucleotides. Telomerase recognizes the telomere DNA repeat sequence and elongates it in the 5-to-3 direction
• The telomerase synthesizes a new copy of the repeat, using an RNA template/primer that is a component of the enzyme itself
Internal RNA template

Telomerase

DNA synthesis

Telomerase translocation

DNA synthesis
Replication of the lagging strand of a linear chromosome encounters a problem at the 3' end.
How do we age?

• As we grow older, the activity of telomerase is reduced

• An inverse relationship between age and telomeric length has been observed

• The gradual shortening of the chromosome ends leads to cell death, and it has even been suggested that life span is determined by the length of telomeres
Elixir of youth
Eukaryotic cell cycle

http://www.themedicalbiochemistrypage.org/cell-cycle.php#check
Checkpoints

- G2 Checkpoint
  - Check for:
    - Cell size
    - DNA replication

- G1 Checkpoint
  - Check for:
    - Cell size
    - Nutrients
    - Growth factors
    - DNA damage

- Spindle Assembly Checkpoint
  - Check for:
    - Chromosome attachment to spindle

- Resting state (G0)
Cyclins

Cyclin Expression Cycle

- Cyclin D
- Cyclin E
- Cyclin A
- Cyclin B

Concentration

G1 Phase  S Phase  G2 Phase  Mitosis
Role of retinoblastoma (Rb) and p53