Basic Principles of Forensic Molecular Biology and Genetics
Nucleic Acids: Structure and Reactivity
DNA Structure and Reactivity

“We are still far from the time when people will understand the curious relationship between one fragment of nature and another, which all the same explain each other and enhance each other.” --Vincent van Gogh
Central Dogma of Molecular Biology

DNA $\rightarrow$ RNA $\rightarrow$ protein

However:
- DNA dependent DNA polymerase (DNA $\rightarrow$ DNA)
- RNA dependent RNA polymerase (RNA $\rightarrow$ RNA)
- Reverse transcriptase (RNA $\rightarrow$ cDNA)
DNA Structure and Reactivity Cont.

Structure-Function

- “It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.”

  Watson and Crick 1953

B-Form DNA
DNA Structure and Reactivity Cont.

- RNA Structure/Function
  - Structure and function relationship not as obvious as DNA
DNA Structure and Reactivity Cont.

DNA Components

- Bases
  - Purines: A, G, H
  - Pyrimidines: T, C, U
  - 5-Methyl C (at CpG and CpXpG)
DNA Components Cont.

- Sugars
  - β-D-ribose
  - β-D-2 deoxyribose
  - C2’ endo versus C3’ endo form
  - Envelope v. twist forms
DNA Components Cont.

- Nucleosides/nucleotides
  - dA, dATP, ddTTP
  - Anti- v. Syn- N glycosidic bond angles
DNA Structure and Reactivity Cont.

- DNA Components Cont.
  - Single polynucleotide chain
    - 5’ and 3’ ends
DNA Structure and Reactivity Cont.

- **Base Pair Interactions**
  - Watson-Crick (W-C) base pairs
    - C·G, A·T
    - H-bonds (~2-3 kcal/mole)
      - normally ~3-7 kcal/mole; c.f. covalent bonds, ~80-100 kcal/mole
  - Stacking energies
  - Orientations in 3D space

---

**Base Pair Stacking Energies**

<table>
<thead>
<tr>
<th>Dinucleotide base pairs</th>
<th>Stacking energies (kcal/mol/stacked pair)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>(GC) · (GC)</td>
<td>-14.59</td>
</tr>
<tr>
<td>(AC) · (GT)</td>
<td>-10.51</td>
</tr>
<tr>
<td>(TC) · (GA)</td>
<td>- 9.81</td>
</tr>
<tr>
<td>(CG) · (CG)</td>
<td>- 9.69</td>
</tr>
<tr>
<td>(GG) · (CC)</td>
<td>- 8.26</td>
</tr>
<tr>
<td>(AT) · (AT)</td>
<td>- 6.57</td>
</tr>
<tr>
<td>(TG) · (CA)</td>
<td>- 6.57</td>
</tr>
<tr>
<td>(AG) · (CT)</td>
<td>- 6.78</td>
</tr>
<tr>
<td>(AA) · (TT)</td>
<td>- 5.37</td>
</tr>
<tr>
<td>(TA) · (TA)</td>
<td>- 3.82</td>
</tr>
</tbody>
</table>
DNA Structure and Reactivity Cont.

- Base Pair Interactions Cont.
B Form DNA: Structural Parameters

- Helix pitch (35.7 Å)
- Helix diameter (20 Å)
- Helical turn (≈ 360° = 10.5 bp per turn)
- Twist angle (rotation per residue) (34.4°)
- Base pair tilt (minus 6°)
- Axial rise (3.4 Å)

B Form DNA: Double Helix

- Molecular model
DNA Structure and Reactivity Cont.

- Non-B Form DNA: A and Z Helices
  - Table
  - Molecular models

### Table: Helix Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>A-DNA</th>
<th>B-DNA</th>
<th>Z-DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Holliday sense</td>
<td>Right</td>
<td>Right</td>
<td>Left</td>
</tr>
<tr>
<td>Residue per turn</td>
<td>11</td>
<td>10 (10.5)</td>
<td>12</td>
</tr>
<tr>
<td>Axial rise (Å)</td>
<td>2.55</td>
<td>3.4</td>
<td>3.7</td>
</tr>
<tr>
<td>Helix pitch (°)</td>
<td>28</td>
<td>34</td>
<td>45</td>
</tr>
<tr>
<td>Base pair tilt (°)</td>
<td>20</td>
<td>−6</td>
<td>7</td>
</tr>
<tr>
<td>Rotation per residue (°)</td>
<td>33</td>
<td>36 (34.3)</td>
<td>−30</td>
</tr>
<tr>
<td>Diameter of helix (Å)</td>
<td>23</td>
<td>20</td>
<td>18</td>
</tr>
<tr>
<td>Glycosidic bond configuration dA, dT, dC</td>
<td>anti</td>
<td>anti</td>
<td>anti</td>
</tr>
<tr>
<td>dG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sugar puckers</td>
<td>dA, dT, dC</td>
<td>C3' endo</td>
<td>C2' endo</td>
</tr>
<tr>
<td>dG</td>
<td>C3' endo</td>
<td>C2' endo</td>
<td>C3' endo</td>
</tr>
<tr>
<td>Intrachain phosphate–phosphate distance (Å)</td>
<td>5.9</td>
<td>7.0</td>
<td>7.0</td>
</tr>
<tr>
<td></td>
<td>5.9</td>
<td>7.0</td>
<td>5.9</td>
</tr>
</tbody>
</table>

*Values in parentheses are the residues per turn and rotation per residue for B-form DNA as it exists in solution of physiological ionic strength. Other values are taken from X-ray diffraction data.
DNA Structure and Reactivity Cont.

- Non-canonical Base Pairing Schemes
  - Reversed W-C
  - Hoogsteen, Reversed Hoogsteen [8]
  - Wobble, Guanine (anti)-Adenine (syn) [8]
DNA Structure and Reactivity Cont.

- Non-canonical Base Pairing Schemes Cont.
  - Tautomerization
    - keto-enol, amine-imine
DNA Structure and Reactivity Cont.

base pairing of tautomers

base pairing of the ionized forms

Thymine • Guanine (enol)
Cytosine • Adenine (imino)
Thymine (enol) • Guanine
Cytosine (imino) • Adenine

Cytosine • Adenine (ionized) Wobble
Thymine (ionized) • Guanine
Cytosine (ionized) • Guanine Hoogsteen
Adenine (ionized) • Guanine (syn)
DNA Structure and Reactivity Cont.

- Accessibility of Functional Groups in DNA
  - Major and minor groove accessibility
DNA Structure and Reactivity Cont.

sites of electrophilic attack

sites of nucleophilic attack
DNA Structure and Reactivity Cont.

Non-Helical Structures

- Bent DNA
  - pyrimidine dimers
  - multiple tracts of A phased by 10 bp
- Cruciform structures
- Supercoiling (L = T + W)
- Triplex DNA
- Four stranded DNA
DNA Structure and Reactivity Cont.

- Methylation of DNA
**DNA Structure and Reactivity Cont.**

**Methylation of DNA**
- Example of epigenetics
  - any gene regulatory activity that doesn’t involve changes to the DNA sequence and that can persist through one or more generations
- Essential for development of mammals (why is unknown)
- 3 DNA cytosine methyltransferases cloned: DNMT1, -3a, -3b
- Alters major groove
- CpG promoter methylation
  - prevents transcription initiation
  - silencing of genes
    - on inactive X
    - parasitic DNA
    - can ‘spread’ (methylation attracts more methylation)
    - imprinted genes
DNA Structure and Reactivity Cont.

Methylation of DNA Cont.

- CpG islands
  - CpG dinucleotides are underrepresented in mammalian DNA (likely due to mutagenic potential of C deamination)
  - Despite above, are found in genes (promoters, but also first exons and 3’ end)
  - >200 bp, high GC content, observed/expected ratio >0.6 (needs revising)

- Why are there varying methylation patterns in eukaryotes? (c.f prokaryotes where all sites are methylated in the presence of the appropriate methyltransferase)?
  - Exclusion of access to methylation sites by DNA bound proteins
    - Removal of SP1 binding sites flanking CpG island ----> de novo methylation during development
  - Methylation targeting mechanism steered by sequence specific binding proteins
    - DNMT1, DNMT3a associate with Rb, E2F1, histone deacetylase (HDACs), RP58 (transcriptional repressor)
DNA Structure and Reactivity Cont.

Methylation of DNA Cont.
- Mechanisms of Transcriptional Silencing
  - Inhibits initiation, not elongation in mammals
DNA Structure and Reactivity Cont.

- **Methylation of DNA Cont.**
  - Methylation and Tissue Specific Gene Expression (TSGE)
    - DNA modifications are tissue specific
    - Link between methylation and TSGE not clear and in some doubt
  - Methylation and Human Disease
    - C methylation is major contributor to disease- causing germline mutations and somatic mutations causing cancer
    - Abnormal methylation of promoters of regulatory genes causes silencing (e.g. p16 tumor suppressor gene)
Methylation of DNA Cont.

- Methylation and Human Disease Cont.
  - Methylation is also important for development after birth
    - ICF (immunodeficiency, centromeric instability, facial anomalies syndrome)
      - Mutation in DNMT3b ---> undermethylation of satellite DNA and specific chromosome decondensation
    - Rett Syndrome
      - Common kind of mental retardation in young girls
      - Mutation in MBD (methylated CpG binding protein)
      - May not be able to interpret methylation signal

- Major Function of Methylation?
  - Long term silencing of noncoding DNA in the genome, which contains a substantial proportion of repetitive elements
DNA Damage and Repair

Primary structure of DNA is dynamic and subject to constant change such as:
transposition, changes in chemistry or sequence of nucleotides due to:

- spontaneous damage
  - replication, recombination, repair
  - inherent instability of specific chemical bonds
  - physical or chemical agents from the environment

- environmental damage
  - physical agents
  - chemical agents
DNA Damage

- DNA-DNA (intra-/inter-)
- DNA-protein
- ssbr
- dsbr
- chain breaks
- cross links
- adducts
- oxidation (ROS)
- hydrolysis (+H₂O)
- UV
- CPD
- (6-4)PP
- ring saturated pyrimidines
- cyclopurines
- formamido-pyrimidines
- thymine glycol
- cytosine glycol
- depurination (depyrimidination)
- deamination
- C→U
- 5mC→T
- (CG→TA)
- AP site
- chain cleavage (β-elimination)
- thymine glycol
- cytosine glycol
DNA Damage and Repair

**Damage**
Damage refers to the situation whereby the primary structure is covalently altered other than as result of epigenetic gene regulation. Damage can be spontaneous or environmental in nature.

**Mismatches**
Difference in free energy between complementary and non-complementary bases is only ~2-3 kcal/mol (a single hydrogen bond) - translates to potential replication error frequency of 1-10 % per base. However, error frequency is 6-9 orders of magnitude less!

<table>
<thead>
<tr>
<th>Event</th>
<th>Cumulative Error Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base pairing</td>
<td>$\sim 10^{-1} - 10^{-2}$</td>
</tr>
<tr>
<td>DNA polymerase actions</td>
<td>$\sim 10^{-5} - 10^{-6}$</td>
</tr>
<tr>
<td>(base selection/proofreading)</td>
<td></td>
</tr>
<tr>
<td>Accessory proteins (e.g. SSB protein)</td>
<td>$\sim 10^{-7}$</td>
</tr>
<tr>
<td>Post-replicative mismatch correction</td>
<td>$\sim 10^{-10}$</td>
</tr>
</tbody>
</table>
DNA Damage and Repair Cont.

**Tautomeric Shifts**

- If a base in a template strand exists in its rare tautomeric form misincorporation in the daughter strand can result.

  Imino:
  - C (imino) will pair with A
  - A (imino) will pair with C
  - T (enol) will pair with G
  - G (enol) will pair with T

**Deamination**

Bases containing exocyclic amino groups can undergo deamination.
Deamination

Cytosine $\rightarrow$ Uracil

Adenine $\rightarrow$ Hypoxanthine

Guanine $\rightarrow$ Xanthine

5-Methylcytosine $\rightarrow$ Thymine
Deamination (cont’d)

C → U

Mutagenic lesion: If not repaired, can result in a G.C → A.T transition.

It is speculated that the reason for the use of T (methylated U) in DNA instead of U (as in RNA) allows for the facile detection of the deamination product of C → Uracil excised rapidly in DNA by a uracil-DNA glycosylase.

Rate of C deamination in ssDNA > 100 fold increased over that in dsDNA.

5-MeC → T

Highly mutagenic due to degree of inefficiency of MMR system: results in G.C → A.T transition.

A, G deamination occurs at a much reduced rate compared to C (< 2%). Rate is 10-4 that of depurination (see below).
Depurination /depyrimidination

Apurinic/apyrimidinic (AP) sites can exist at a level of 50,000-200,000 AP sites per genome in human tissues.

- AP site deoxyribose exists in an equilibrium between the closed furanose and open aldehyde form, the 3’phosphodiester bonds of which are labile. These can be hydrolysed by a β-elimination reaction in which the pentose carbon b to the aldehyde is activated at alkaline pH and elevated temperature.

- AP sites are repaired by 5’ phosphodiester hydrolysis (by AP endonuclease) followed by 3’ phosphate elimination (by dRp-lyase activity of DNA polymerase).
DNA Damage and Repair Cont.

- **Oxidative Damage**
  Reactive oxygen species (ROS) can cause oxidative damage to DNA. ROS include hydroxyl radicals, hydrogen peroxide and singlet oxygen with the hydroxyl radical being the most important.

  \[ RH_2 + \cdot OH \rightarrow RH + H_2O \]

  Superoxide radicals (produced as a side product of mitochondrial respiration) can be dismutated into hydrogen peroxide which is then converted to the hydroxyl radical through a Fenton type reaction.
Oxidative Damage Cont.

OH. -Mediated oxidative damage can be of the following types:

- Oxidized bases
- Abasic sites
- DNA-DNA intra-strand adducts
- DNA strand breaks
- DNA-protein crosslinks

We will concentrate on oxidized bases, particularly:

- Guanine → 8-oxoguanosine
- Thymine → thymidine glycol, hydantoins
Oxidative Damage Cont.

Guanine \(\rightarrow\) 8-oxoquanosine [18]

Guanine is the most susceptible DNA target to OH.-mediated oxidation.

8-oxo-7,8-dihydro-2′-deoxyguanosine (8-oxodGuo) (23) is a mispairing lesion during replication.
DNA Damage and Repair Cont.

- Oxidative Damage Cont.

**Thymine \( \rightarrow \) thymidine glycol, hydantoins**

The predominant products are 5,6-dihydroxy-5,6-dihydro-thymidine (thymidine glycol) (12) and 5-hydroxy-5-methylhydantoin (13). These lesions block replication.
Environmental Damage

- **Ionizing radiation**
  - Radiolysis of water
  - \( H_2O \rightarrow H_2O^+ + e^- \)

  **Base damage**
  - Damage is OH\(^-\) mediated, similar to that previously described for oxidative damage. Phenomenon of ‘local multiply damaged sites’ → single energy deposition event can result in several radical reactions in the immediate vicinity.

- **Sugar damage and strand breaks**
  - Damage to sugar residues less frequent than damage to bases (by ~3 fold).
  - Due to modified structures at the 3’ end (e.g. phosphoglycolate \( O_3POCH_2COO^- \)) precludes repair by a simple ligation step.

  - Single and strand breaks are also formed.
  - Controversy exists due to the formation of these lesions at non-physiological doses.
DNA Damage and Repair Cont.

- UV Radiation

UV Radiation Spectrum

UV spectrum divided into UV-A (400-320 nm), UV-B (320-290 nm) and UV-C (290-100 nm). Solar radiation consists of mainly UV-A and UV-B since ozone absorbs at 300 nm.
UV Radiation Cont.

CPD

When DNA is exposed to wavelengths approaching its absorption maximum (260 nm), adjacent pyrimidines become covalently linked by the formation of a cyclobutane ring structure resulting from saturation of their respective 5,6 double bonds → called a cyclobutane pyrimidine dimer (CPD)

- 12 isomeric forms (only 4 found in any quantity)
  - *cis-syn* in B form DNA
  - *trans-syn* in denatured DNA, ssDNA, B-Z DNA junctions
  - extremely stable to pH and temperature extremes
UV Radiation Cont. 

CPD Cont.
- blocks DNA polymerase read through
- bulky, helix distorting lesions
- TT--T >> CT >> C--T, TT, T--C >> CC >> C--C (68:13:16:3)
- CPD yield dependent upon sequence context outside of dimer
- process: Py + Py $\overset{UV}{\leftrightarrow}$ Py $\leftrightarrow$ Py
  - equilibrium exists such that the amount of dimers 7% of total thymine

(6-4) PP

Alkali labile lesions also formed in UV treated DNA, called pyrimidine-pyrimidone (6-4) lesions or photoproducts, (6-4)PPs.
- causes major helical distortions
- blocks DNA polymerase read through
- TC, CC more common than TT (CT not seen)
- (6-4)PP $\overset{313nm}{\longrightarrow}$ Dewar isomer
Spore photoproduction

- uv irradiated spores of B. subtilis
- related to state of hydration (A form in dehydrated DNA)
DNA Damage and Repair Cont.

**Chemical Agents**
- **Alkylating agents**
  - Electrophilic compounds that react with nucleophilic centers in DNA.
  - O6-methylguanine
    - generated endogenously
    - pairs with C or T
    - MGMT enzyme
  - N7-methylguanine
  - N3-methyladenine (MPG glycosylase)

- **Cross Linking Agents**
  - Bifunctional alkylating agents can cause intra- or inter-strand crosslinks
  - Nitrogen and sulfur mustard, mitomycin and *cis*-platinum
  - Also caused by UV and ionizing radiation
<table>
<thead>
<tr>
<th>Response</th>
<th>Mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reversal of DNA damage</td>
<td><em>Enzymatic photoreactivation</em></td>
</tr>
<tr>
<td></td>
<td>Repair of spore photoproduct</td>
</tr>
<tr>
<td></td>
<td>Repair of O(^6)-alkylguanine, O(^4)-alkylthymine,</td>
</tr>
<tr>
<td></td>
<td>and alklyphosphothriesters</td>
</tr>
<tr>
<td></td>
<td><em>Ligation of DNA strand breaks</em></td>
</tr>
<tr>
<td>Excision of DNA damage</td>
<td><em>Base excision repair</em></td>
</tr>
<tr>
<td></td>
<td>Nucleotide excision repair</td>
</tr>
<tr>
<td></td>
<td>Mismatch repair</td>
</tr>
<tr>
<td>Tolerance of DNA damage</td>
<td>Replicative bypass of template damage with gap formation and recombination</td>
</tr>
<tr>
<td></td>
<td><em>Translesion DNA synthesis</em></td>
</tr>
</tbody>
</table>
DNA Damage and Repair Cont.

DNA Repair

Cancer is a genetic disease in the sense that mutations can activate proto-oncogenes or inactivate tumor suppressor genes. However, in addition to these spontaneous mutations, cancer risk is increased due to the ‘mutator’ phenotype caused by inherited or acquired faulty DNA repair systems. The main mammalian DNA repair systems include:

- Direct Reversal
- Replicational Bypass (Translesion Synthesis)
- Nucleotide excision repair (NER)
- Base excision repair (BER)
- Homologous recombination
- End joining
- Mismatch repair
DNA Damage and Repair Cont.
**DNA Damage and Repair Cont.**

<table>
<thead>
<tr>
<th>Syndrome</th>
<th>Affected maintenance mechanism</th>
<th>Main type of genome instability</th>
<th>Major cancer predisposition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xeroderma pigmentosum</td>
<td>NER (+ TCR)</td>
<td>Point mutations</td>
<td>UV-induced skin cancer</td>
</tr>
<tr>
<td>Cockayne syndrome</td>
<td>TCR</td>
<td>Point mutations</td>
<td>None*</td>
</tr>
<tr>
<td>Trichothiodystrophy</td>
<td>NER / TCR</td>
<td>Point mutations</td>
<td>None*</td>
</tr>
<tr>
<td>Ataxia telangiectasia (AT)</td>
<td>DSB response/repair</td>
<td>Chromosome aberrations</td>
<td>Lymphomas</td>
</tr>
<tr>
<td>AT-like disorder</td>
<td>DSB response/repair</td>
<td>Chromosome aberrations</td>
<td>Lymphomas</td>
</tr>
<tr>
<td>Nijmegen breakage syndrome</td>
<td>DSB response/repair</td>
<td>Chromosome aberrations</td>
<td>Lymphomas</td>
</tr>
<tr>
<td>BRCA1/BRCA2</td>
<td>HR</td>
<td>Chromosome aberrations</td>
<td>Breast (ovarian) cancer</td>
</tr>
<tr>
<td>Werner syndrome</td>
<td>HR? / TLS?</td>
<td>Chromosome aberrations</td>
<td>Various cancers</td>
</tr>
<tr>
<td>Bloom syndrome</td>
<td>HR?</td>
<td>Chromosome aberrations (SCE1)</td>
<td>Leukaemia, lymphoma, others</td>
</tr>
<tr>
<td>Rothmund-Thomson syndrome</td>
<td>HR?</td>
<td>Chromosome aberrations</td>
<td>Osteosarcoma</td>
</tr>
<tr>
<td>Ligase IV deficiency†</td>
<td>EJ</td>
<td>Recombination fidelity</td>
<td>Leukaemia (?)</td>
</tr>
<tr>
<td>HNPCG</td>
<td>MMR</td>
<td>Point mutations</td>
<td>Colorectal cancer</td>
</tr>
<tr>
<td>Xeroderma pigmentosum variant</td>
<td>TLS†</td>
<td>Point mutations</td>
<td>UV-induced skin cancer</td>
</tr>
</tbody>
</table>

*Defect in transcription-coupled repair triggers apoptosis, which may protect against UV-induced cancer.
†One patient with leukaemia and radiosensitivity described with active-site mutation in ligase IV.
‡Specific defect in relatively error-free bypass replication of UV-induced cyclobutane pyrimidine dimers.

Abbreviations: BER, base-excision repair; DSB, double-strand break; HNPCG, hereditary non-polyposis colorectal cancer; HR, homologous recombination; MMR, mismatch repair; NER, nucleotidase-excision repair; SCE, sister-chromatid exchange; TCR, transcription-coupled repair; TLS, translation synthesis.
DNA Repair

Direct Reversal

- Enzymatic photoreactivation
- Repair of spore photoproduct
- Repair of $O^6$ alkyl guanine, $O^4$-alkyl thymine and alkylphosphotriesters
- Ligation of single strand breaks
Ligation of single stranded breaks

+ DNA ligase, NAD or ATP

+ AMP + NMN or PP_i
Replicational Bypass

To minimize cell death from replication blockage a process of translesion synthesis (TLS) has evolved

- TLS-, SOS-, mutagenic-, error-prone-, Y-family- DNA polymerases

Bacterial mutatants non-mutable by UV radiation in 1976

- Mutations are not accidental but result from active chemical process

Hypothesized that error prone DNA polymerases exist that could insert random nucleotides in a template independent fashion and that DNA damage induced mutagenesis might be a specific cellular response to damage

In E. coli, SOS response involves >30 inducible genes

- Initially recA and lexA mutants
- Then, umuC and umuD (pol V) and dinB (pol IV)
- Genome screens for homologs resulted in the identification of an extended superfamily (Y family)
DNA Damage and Repair Cont.
Nucleotide Excision Repair (NER)

NER deals with a wide class of helix distorting lesions. Over 25 proteins participate in the NER pathway. Main function is to remove UV photoproducts, crosslinks and other bulky lesions. Two distinct pathways:

- Global genome NER
  - surveys entire genome for distortions
- Transcription coupled repair (TCR)
  - acts on damage that blocks RNA polymerase

Xeroderma Pigmentosa (XP) is a disorder of the NER system which increases a patient's risk for developing skin cancer >2000 fold.
- caused by inability to excise UV photoproducts (e.g. CPDs)
- mutations in one of seven genes (XPA-XPG)

Cockayne’s syndrome (CS) is a TCR specific disorder caused by mutations in the CSA or CSB genes.
- no predisposition to cancer
- CS cells particularly susceptible to apoptosis
  - protects against tumorigenesis
NER

- Damage recognition
- DNA unwinding (discrimination of damaged from non-damaged strand)
- Incision on both sides of the lesion
  - Removal of ~25-30 nucleotides
- Repair synthesis
- Ligation
DNA Damage and Repair Cont.

DNA Repair Cont.

- Base Excision Repair (BER)

BER is the cell’s main protection against ROS, methylation and deamination.

- no disorders caused by inherited deficiencies in BER have been identified
  - partial redundancies between different glycosylases
- however, inactivation of BER core proteins causes embryonic lethality
DNA Damage and Repair Cont.

Apurinic site (aldehyde form) → β-Elimination → Strand break on 3' side of apurinic site
DNA Damage and Repair Cont.
DNA Damage and Repair Cont.

DNA Repair Cont.

- Homologous Recombination (HR) and End Joining (EJ)

Double strand breaks (DSBs) arise due to:
- Ionizing radiation/X rays
- Free radicals

Once DSB is detected, a complex machinery is mobilized to halt cell cycle progression and to recruit repair factors. If a second undamaged copy of template is available -> HR. If not, ER -> more error prone.
DNA Damage and Repair Cont.
Mismatch Repair

MMR removes:
- Nucleotides mispaired during replication
- Insertion/deletion loops due to slipped strand mispairing of repetitive sequences (microsatellites) during replication or recombination

The microsatellite instability (MSI) phenotype of hereditary non-polyposis colorectal cancer (HNPCC) caused mainly by mutations in MMR genes
- especially hMLH1 and hMSH2
DNA Damage and Repair Cont.

<table>
<thead>
<tr>
<th>E. coli</th>
<th>S. cerevisiae</th>
<th>H. sapiens</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Msh6</td>
<td>MSH6</td>
<td></td>
<td>Recognizes the mismatch</td>
</tr>
<tr>
<td>Msh3</td>
<td>MSH3</td>
<td></td>
<td>Forms a complex with Msh2 (MSH2)</td>
</tr>
<tr>
<td>Msh1</td>
<td>Not identified</td>
<td>Functions in mitochondrial MMR</td>
<td></td>
</tr>
<tr>
<td>Msh4</td>
<td>MSH4</td>
<td></td>
<td>Required for mitotic recombination</td>
</tr>
<tr>
<td>Msh5</td>
<td>MSH5</td>
<td></td>
<td>Required for mitotic recombination</td>
</tr>
<tr>
<td>MshL</td>
<td>Mlh1</td>
<td>MLH1</td>
<td>Couples mismatch recognition and subsequent repair</td>
</tr>
<tr>
<td>Pms1</td>
<td>PMS2</td>
<td></td>
<td>Forms a complex with Mlh1 (MLH1)</td>
</tr>
<tr>
<td>Mlh2</td>
<td>PMSF</td>
<td></td>
<td>Forms a complex with Mlh1 (MLH1)</td>
</tr>
<tr>
<td>Mlh3</td>
<td>MLH3</td>
<td></td>
<td>Forms a complex with Mlh1 (MLH1)</td>
</tr>
</tbody>
</table>

**MMR gene** | **Mechanism of biallelic inactivation** | **Microsatellite instability phenotype** |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MSH2</td>
<td>Somatic mutation + LOH or two somatic mutations</td>
<td>Germine mutation + LOH or germline + somatic mutation</td>
</tr>
<tr>
<td>MLH1</td>
<td>Promoter hypermethylation (biallelic) or somatic mutation + LOH or two somatic mutations</td>
<td>Germine mutation + LOH</td>
</tr>
<tr>
<td>MSH6</td>
<td>Frequent frame shift mutations in coding Cx repeat</td>
<td>Germine mutation + somatic (frameshift) mutation or germline mutation + LOH</td>
</tr>
<tr>
<td>MSH6</td>
<td>Frequent frame shift mutations in coding A8 repeat</td>
<td>None identified</td>
</tr>
<tr>
<td>PMS2</td>
<td>Two somatic mutations</td>
<td>Germine + somatic mutation</td>
</tr>
<tr>
<td>PMS2</td>
<td>Data not available</td>
<td>Data not available</td>
</tr>
<tr>
<td>MLH3</td>
<td>Data not available</td>
<td>Data not available</td>
</tr>
</tbody>
</table>
DNA Damage and Repair Cont.
DNA Replication

[Diagram of DNA replication process with labels for polymerase, ligase, leading template, lagging template, RNA primer, Okasaki fragment, DNA polymerase, helicase, parental DNA, single-stranded DNA binding proteins, and new strand.]
DNA replication

- Topoisomerase
- Parental DNA duplex
- Helicase
- Primase
- RNA primer
- Single-strand binding protein
- Leading strand template
- Newly-synthesized leading strand DNA
- Dimeric replicative DNA polymerase III
- Okazaki fragment
- DNA polymerase I
- Lagging strand template
- DNA ligase
- Lagging strand DNA
Functional RNA Molecules

- rRNA
- mRNA
- tRNA
- catalytic RNA
- small nuclear RNA (snRNA) (splicing)
- guide RNA (silencing)
- telomerase RNA (chromosome end replication)
- signal recognition particle (SRP) RNA (protein translocation)
- small nucleolar RNA (snoRNA) (rRNA modification)
- micro-RNA
RNA Structure

The difference in the biology of RNA compared with DNA stems from the presence in RNA of the 2’ OH residue in the sugar ring. This creates a nucleophilic center that changes the reactivity of the molecule and sterically hinders the formation of a ‘B form ‘ double helical’ structure.
RNA Structure

- RNA synthesized as single strands
- folds into stable structure, e.g. tRNA
  - short regions of W-C base pairing interrupted by:
    - non-canonical pairs
    - loops, including hairpins
    - bulged nucleotides
    - more complex structures e.g. pseudoknots, junctions
  - 2D structure predicted by use of thermodynamic secondary structure prediction algorithms
    - includes covariation between two nucleotides that conserve W-C --> evidence of a base pair between these nucleotides
- modified nucleotides and sugars
  - pseudouridine, 5’ methyl cytosine, ribothymidine (tRNA greatest number and percentage of these)
  - 2’-O-methyl substituents
- dsRNA --> A-form
  - not > 10 bp
  - stabilized by H-bonding of the 2’-OH group with neighboring phosphates or other riboses
  - Non W-C pairs common e.g. G-U and A-C
## mRNA

### Life cycle

**Nucleus:**
- transcription
- capping
- methylation
- polyadenylation
- splicing
- editing

**Translocation to cytoplasm**

**Cytoplasm:**
- translation
- degradation
- poly(A) remodeling/localisation

### Structural elements

- 5’ cap
- 5’ UTR
- coding region
- 3’ UTR
- poly(A)
RNA Splicing

Eukaryotic mRNAs are transcribed as precursors (pre-mRNA) containing intervening sequences (introns). These sequences are subsequently removed and the flanking regions containing the exons are spliced together to form mature mRNA.
Splicing

(a) 5' splice site

Branch point

Polyadenylation tract

3' splice site

Commitment complex

Complexes across intron

Exon definition complex

ATP

U2 snRNP

Pre-spliceosome complex

ATP

U5, U4/U6 snRNPs

Spliceosome
Translation Initiation
Peptide Bond Formation
Translation: Peptide Release
rRNA

Proteins

Subunits

Assembled ribosomes

23S (2900 bases) + 5S (120 bases) + (Total: 31) → 23S + 50S

16S (1540 bases) + S1 + S2 + S3 + (Total: 21) → 16S + 30S

5S

L1 L2 L3

30S

70S

50S
Terminally differentiated cells
- have a pattern of gene expression that is unique to each cell type
- manifested by presence and relative abundance of specific mRNAs

Transcriptome
- The unique pattern of gene expression within each cell type

Microarrays
- Used to monitor gene expression in different tissues
"This is not the end. It is not even the beginning of the end. But it is, perhaps, the end of the beginning."