

DNA REPAIR

Introduction

The integrity of DNA is vital to the survival of the cell. Therefore, efficient mechanisms have evolved to correct mistakes that occur during replication. As we have seen, proofreading functions can increase the fidelity of DNA polymerases up to 100-fold. Nevertheless, replicative errors still occur. Polymerase errors mainly result in misincorporation, but small duplications and deletions can also occur.

DNA is also subject to damage by environmental agents such as radiation, ultraviolet light, and various chemicals. Spontaneous modification of DNA bases can also occur. Spontaneous base changes and environmental agents can cause base modifications and other types of damage that can lead to structural distortions of the double helix. Some distortions can block polymerase activity. In addition, some agents can cause chromosome breaks (double strand breaks).

Most life forms have the ability to respond to these alterations. The responses take one of two general forms. Cells can either repair the damage and restore the DNA to its normal state, or they can tolerate lesions in a way that minimizes their potentially lethal effects.

The first category represents true DNA repair, the major function of which is the prevention of mutations. Tolerance of DNA damage does not necessarily prevent mutation. There are several types of repair systems:

- Direct repair
- Excision repair
 - three types:
 - base excision repair
 - nucleotide excision repair
 - mismatch repair
- Tolerance systems/SOS repair
- Double strand break repair
- Recombination repair (retrieval systems)

As usual, more is known about prokaryotic repair systems than about eukaryotic systems, but the gap is closing rapidly. This is largely because it is now recognized that mutations in oncogenes or tumor suppressor genes can lead to neoplastic transformation of cells.

Thymine dimers

The best studied and most common of all DNA modifications is the thymine dimer. This is an intrastrand dimer formed by two adjacent pyrimidines (usually thymine) as a result of ultraviolet radiation. The dimer is formed by a covalent linkage between two thymines which generates a cyclobutane ring (Fig. 66).

When thymine dimers are present, the double helix is distorted, as the thymines are pulled toward each other. Hydrogen bonding to adenines on the opposite strand, while possible, is weakened. The distortion causes a loss of template information. In addition, it can block an advancing replication fork, or an RNA polymerase in the act of transcription.

Thymine dimers and the discovery of “light” and “dark” repair

Repair was first suggested by observing the effect of UV light on the growth of *E. coli* (UV survival curves). Under standard conditions, the ability of cells to form colonies decreases as a function of UV dose (Fig. 67).

However, survival of UV-irradiated cells is enhanced if they are exposed to visible light before plating. That is, visible light causes the reversal of some of the damage caused by UV light. This is called photoreactivation, or light repair.

Also, it was found that holding cells in non-nutrient medium (in stasis, without light) for several hours after irradiation and before plating also enhanced recovery. That is, holding cells in stasis also reverses some of the damage caused by UV light. This is referred to as dark repair (also called liquid holding recovery).

Proof that these distinct repair systems exist is genetic: *pho* mutants can't photoreactivate but can perform dark repair, while several other types of mutants can photoreactivate but can't carry out dark repair.

Repair Mechanisms

Direct Repair

There are only a few types of direct repair processes, and only a limited group of lesions can be repaired in this way. The defining characteristic of direct repair mechanisms is that lesions are

corrected by reactions that usually require only a single enzyme to directly reverse the damage. No DNA synthesis is required.

Photoreactivation of thymine dimers. Photoreactivation (light repair) is an example of direct repair. Photoreactivation is catalyzed by DNA photolyase. This enzyme carries out a unique reaction in which light energy (365-400 nm) is used to cleave the cyclobutane ring joining the two adjacent pyrimidines. The photolyase protein in *E. coli* has a mass of ~54 kDa. The holoenzyme is associated with two chromophores (FADH₂ and pterin). Light energy absorbed by the chromophores is used for photolysis.

The reaction occurs in two steps.

1. Recognition. This step is light-independent and involves recognition of the DNA distortion and the pyrimidine dimer. Photolyase recognizes and binds to the damaged site.
2. Photolysis. Light energy is used to cleave the cyclobutane C-C bonds.

Photoreactivation is widespread in nature, but is lacking in placental mammals (which lack photolyase). It appears to be especially important in plants.

Reversal of alkylation. Alkylating agents such as methy methane sulfonate (MMS) and nitrosoguanidine transfer methyl groups to several sites of nucleotide bases; the most important product being O⁶-methylguanine. The repair enzyme O⁶-methylguanine methyltransferase can directly reverse the lesion by removing the methyl group. In the process, the methyl group is transferred to a cysteine residue of the enzyme, which results in enzyme inactivation. So the system is easily saturated and overwhelmed by high levels of alkylation.

Excision Repair

Dark repair of thymine dimers is but one of several examples of excision repair. Excision repair is the most common and perhaps the most important class of repair. It is ubiquitous in nature and is used to correct misincorporated and mispaired bases as well as thymine dimers and other damaged or modified bases.

There are three types of excision repair:

base excision repair (BER)

nucleotide excision repair (NER)

mismatch repair (MMR)

These processes share the same general features but differ in mechanistic detail and/or in the type of lesion repaired. The general mechanism of excision repair is as follows:

- Recognition. The damaged site/structure is recognized. This is usually the rate-limiting step.
- Incision/excision step. The damaged structure (thymine dimer, modified base, base mispair etc.) is recognized by an endonuclease that cleaves the damaged strand on one or both sides of the damaged site. Excision involves removal of damaged DNA adjacent to the nick or between two nicks.
- Synthesis step. The gap is filled by new DNA synthesis to replace the excised sequence.
- Ligation step. The nicks are sealed by DNA ligase.

Note that excision repair processes rely on DNA synthesis, using the undamaged strand to code for a replacement. Thus the double-stranded nature of DNA seems designed to optimize repair; the complementary strand serves as a back-up template for any damaged information.

Base excision repair (BER)

Base excision repair (BER) is a process characterized by the excision of nucleic acid base residues in the free form; that is, the bases are excised from the sugar-phosphate backbone. This step is catalyzed by a class of enzymes called DNA glycosylases, which hydrolyze the N-glycosyl bond linking a nitrogenous base to the deoxyribose-phosphate chain, thereby releasing a free purine or pyrimidine base. DNA glycosylases are found in both prokaryotes and eukaryotes. Each removes a limited set of altered bases (Fig. 68).

In BER, nonbulky lesions such as uracil, certain mispaired bases, and modified bases such as N³ methyl adenine are removed from the DNA. (Not thymine dimers.) Repair tracts are generally very small (1-4 nt). Depurination events are also repaired by this pathway.

Uracil N-glycosylase. One of the most important DNA glycosylases is uracil N-glycosylase (ung), which removes uracil (and 5-fluorouracil) from DNA. The ung protein is responsible for removal of uracil in organisms as diverse as humans, plants, and *E. coli*. The enzymes from different organisms share extensive sequence homology.

Uracil in DNA can result from misincorporation by DNA polymerase from small intracellular pools of dUTP. Consequently, some A-U base pairs are generated during replication. (This is usually

kept to a minimum by dUTPase, which converts dUTP to dUMP, which is then converted to TMP by demethylation.)

Uracil also appears in DNA as a result of spontaneous oxidative deamination of cytosine (Fig. 69). This is a greater problem, as it results in a G-U mispair. If this is not repaired, a G-C to A-T transition mutation will result following replication.

(Note: a transition mutation is a base pair substitution which exchanges one purine for another and one pyrimidine for another; as opposed to a transversion, which exchanges purine for pyrimidine.)

G-T mismatch glycosylase. This is another important enzyme that removes the T in a G-T mispair, because the T is most likely incorrect.

The reason for this is that the most common natural modification of eukaryotic DNA is the methylation of cytosine to m5C. However, m5C suffers from spontaneous deamination to thymine at an appreciable frequency (significantly higher than uracil deamination) (Fig. 69). This creates a G-T mispair (G-m5C to G-T). If not corrected, this will result in a transition mutation (G-C to A-T).

To compensate for the high frequency of m5C deamination, a separate BER mechanism (using G-T mismatch glycosylase) has evolved in eukaryotes to remove the T of a G-T mispair. Note that this system is biased; even though both G and T are normal bases, the assumption is that the T is incorrect, which most of the time is the case.

Note: The G-T mispair is an important lesion that is targeted by both the BER and MMR (true mismatch repair) systems.

The GO system. This is an interesting case of multiple BER targeting which concerns 8-oxodG (also called GO). 8-oxodG (8-oxo-7-hydro-deoxyguanosine) is generated by spontaneous oxidation of guanine. The oxidized base frequently mispairs with A, causing a high level of G to T transversions. The MutM protein is a glycosylase that recognizes and removes 8-oxodG in DNA. Another glycosylase, MutY, specifically removes adenine present in G-A mispair. The glycosylases MutM and MutY, together with MutT (which converts 8-oxo-dGTP to the monophosphate), constitute the GO system. One can only conclude that 8-oxodG is a frequently occurring lesion and that its repair is crucial to the cell.

Mechanism of BER

First, a specific glycosylase recognizes a specific type of incorrect base. Glycosylases recognize specific alterations at the nucleotide level. They travel along the DNA and examine nucleotides by "base flipping". The nucleotides are "flipped out" of the helix and into a hole in the enzyme, bringing the target bond to the active site. Presumably only the specific target base (e.g. uracil in the case of ung) will have the correct fit and be cleaved.

Once a target base is recognized the enzyme cleaves the N-glycosyl bond, generating an apurinic or apyrimidinic (AP) site (Fig. 70).

AP sites generated by glycosylase action (or spontaneously, see "depurination" below) are then recognized by AP endonuclease, which cleaves the phosphodiester backbone immediately 5' to the AP site, leaving a 3'-OH and 5'-deoxyribose-phosphate terminus (incision step). AP endonucleases from human (HAP1) and *E. coli* (Xth protein) share extensive sequence homology.

Note that recognition is a two step process in BER, both the glycosylase and the AP endonuclease must recognize a particular lesion.

Completion of BER (synthesis and ligation steps) can occur in several different ways. In one, which occurs in eukaryotes, the single nucleotide gap generated by the glycosylase and AP endonuclease is filled by Pol β . The 5' deoxyribose phosphate is then removed by DNA deoxyribo-phosphodiesterase (dRpase), leaving a nick sealed by DNA ligase III.

A longer BER patch mode also exists in eukaryotes. In this case, DNA synthesis (a short displacement reaction) probably is done by DNA Pol δ or Pol ϵ (and associated factors, including PCNA). Displacement synthesis creates a branched substrate that can be removed by FEN-1 (also called DNase IV, or MF1). Dna2 may also be involved. The nick is finally sealed by DNA ligase I. Note that this process bears a strong resemblance to Okazaki fragment processing. Also recall the polymerase and FEN-1 interact via PCNA.

In *E. coli*, short repair patches (~4 to 6 nt or so) result from nick translation, catalyzed by Pol I. The nick is sealed by DNA ligase. This mechanism is also similar to Okazaki fragment processing.

Depurination repair. This is the most common type of spontaneous damage suffered by DNA; human cells lose an estimated $5-10 \times 10^3$ bases a day through depurination. Depurination leaves an intact deoxyribose sugar lacking a base (an AP site), which is a problem because an AP site

provides no template information. Fortunately, depurination events are efficiently repaired by the BER pathway, starting with AP endonuclease.

Nucleotide excision repair (NER)

Nucleotide excision repair (NER), which has been found in all species tested, excises several types of lesions:

- pyrimidine dimers caused by UV light (T-T is the most common; others are T-C and C-C)
- alkylated nucleotides, including O⁶ methy- or ethylguanine, N⁶ methyladenine
- bases modified by covalent addition of large hydrocarbons, such as cisplatin, cholesterol, the carcinogen benzopyrene.
- bulky mismatches (A-G and G-G)

The substrate range is very broad.

These "bulky lesions" are recognized by a multiprotein complex that scans the DNA for helix distortions. Thus recognition differs from the BER system, which operates at the nucleotide level.

The recognition and removal of damaged DNA is carried out by a complex known as an excision nuclease (exinuclease), which makes a dual incision on each side of the lesion. There are two types of excision nucleases. The prokaryotic type consists of three subunits. The eukaryotic type consists of as many as 16 proteins. The subunits of the prokaryotic exinuclease do not share significant homology with any of the eukaryotic exinuclease subunits. So, unlike the BER systems which are conserved between prokaryotes and eukaryotes, the NER systems are an example of convergent evolution.

NER in *E. coli*

In *E. coli*, Uvr protein-directed repair probably accounts for >95% of all excision repair. Three gene products are required for damage-specific cleavage. The UvrA, UvrB, and UvrC gene products are components of the excision nuclease. A fourth protein, UvrD, is a helicase which removes the incised oligonucleotide (Fig. 71).

UvrA is a damage-specific DNA binding protein with specificity for damage in dsDNA (helix distortion). It is also an ATPase.

UvrB is a weak helicase/ATPase with an affinity for ssDNA.

UvrA-ATP forms a complex with UvrB (A₂B₁ complex). UvrA loads UvrB on ssDNA at the distortion (Fig. 72). The distorted region has a natural propensity to unwind and bend.

UvrB further bends and unwinds (ATP-dependent) a small region (about 5 bp) around the lesion. Once UvrB is loaded, UvrA hydrolyzes its bound ATP and dissociates, leaving a UvrB-DNA complex.

Dissociation of UvrA is essential for UvrC to enter into a complex with UvrB and ssDNA. The UvrB-UvrC complex has endonuclease activity. UvrB makes the 3' incision and UvrC makes the 5' incision. In *E. coli*, the 8th phosphodiester bond 5' and 4th or 5th phosphodiester bond 3' of the lesion are cleaved, resulting in 12-13 nt long repair tracts (short patch repair).

UvrD (DNA helicase II) then enters the complex, displacing UvrC and the ssDNA containing the lesion. DNA Pol I (or Pol II?) then fills the gap from the free 3'-OH terminus, displacing UvrB in the process. (Apparently little or no nick translation takes place in NER.) Finally, the remaining nick is sealed by DNA ligase.

Interestingly, this NER system is also coupled to transcription. Enhanced repair of transcribed DNA is mediated by a protein called transcription-repair coupling factor (TRCF), the product of the *mfd* gene. *mfd* is not an essential gene, but *mfd* mutants show about a threefold increase in spontaneous mutation rate over wild-type cells. This suggests that transcription directed repair is an important mechanism for detecting lesions.

TRCF specifically recognizes RNA polymerase stalled at a lesion. It is capable of interacting with both RNA Pol and UvrA. TRCF releases RNA Pol and the nascent transcript, and recruits the UvrA-UvrB complex to the damaged site.

NER in eukaryotes

Defective NER in humans is associated with several diseases, the best known of which is xeroderma pigmentosum (XP). The disease is hereditary with autosomal recessive inheritance. The frequency of the disease is 10^{-6} in the US but is higher in Japan (10^{-5}). Symptoms of XP range from erythema to xerosis and skin atrophy. Also, nearly 90% of these individuals develop basal and squamous cell carcinomas in their teens, and go on to develop malignant melanoma with high frequency. Most XP individuals also suffer from neurologic disorders including mental retardation. XP individuals are extremely sensitive to sunlight (UV light) and are unable to efficiently repair thymine dimers and other types of DNA damage normally dealt with by NER.

XP individuals are genetically heterogeneous, and seven complementation groups have been identified (XP-A to XP-G). Interestingly, these correspond to biochemical fractions containing different components/complexes of the NER system. In contrast to the *E. coli* NER system, the eukaryotic system is very complex and the excision nuclease (exinuclease) alone consists of at least 16 polypeptides. Remarkably, however, the overall mechanism used by the two systems is quite similar (an interesting case of convergent evolution). The human and yeast systems are the best studied, and much information has come from reconstituted *in vitro* systems and from yeast genetics. We will be concerned only with the overall mechanism and the major activities involved.

The mechanism can be broken down into three steps: damage recognition, dual incision, and repair synthesis. The proteins involved in the first two steps are contained in seven biochemical fractions that are required to reconstitute the excision reaction *in vitro*. Most of the fractions contain several proteins in tight complex (Fig. 73).

Damage recognition is presumed to be carried out by a complex of two proteins, XPA and RPA (Fig. 74). XPA is a dsDNA binding protein with a moderate affinity for damaged DNA. RPA is a ssDNA binding protein (SSB) which also has an affinity for damaged DNA. Together they have a greater affinity for a lesion than either alone.

Once these proteins are bound to a damaged site, TFIIH is recruited via its ability to bind XPA. TFIIH was originally identified as a general transcription factor (GTF) involved in promoter clearance by RNA polymerase II (simply, the transition between the initiation and elongation modes of transcription). TFIIH has several activities. The most relevant here is helicase activity, which unwinds the duplex around the damaged site.

XPG, which interacts specifically with RP-A and TFIIH, is recruited next. (This is the pre-incision complex). XPG is an endonuclease which makes the 3' incision.

XPF then enters the complex and makes the 5' incision. 5' incision sites range from 20 to 26 phosphodiester bonds from the damage site, and 3' incisions from 2 to 10 phosphodiester bonds from the damage. Excised fragments usually range from 24 to 32 nt in length.

Repair synthesis appears to be carried out by both DNA Pol β and DNA Pol δ . Excised fragments, and remaining excision complex proteins, are probably displaced by polymerase. DNA ligase I seals the nick.

As in *E. coli*, transcribed DNA is repaired faster than non-transcribed DNA. This preferential repair is largely confined to the template strand, and in humans, is limited to genes transcribed by RNA polymerase II. The details of transcription coupled repair in humans are not known, but it is presumed to be similar in general mechanism to the *E. coli* system. That is, a factor (in this case a complex of two proteins, CSA and CSB) recognizes a stalled RNA polymerase, disengages it from the template, and simultaneously recruits the exonuclease.

Review:

Sancar, A. (1996) DNA excision repair. Annual Review of Biochemistry 65: 43-81.