DNA Repair Systems

Guardians of the Genome

D N Rao and Yedu Prasad

The 2015 Nobel Prize in Chemistry was awarded jointly to Tomas Lindahl, Paul Modrich and Aziz Sancar to honour their accomplishments in the field of DNA repair. Ever since the discovery of DNA structure and their importance in the storage of genetic information, questions about their stability became pertinent. A molecule which is crucial for the development and propagation of an organism must be closely monitored so that the genetic information is not corrupted. Thanks to the pioneering research work of Lindahl, Sancar, Modrich and their colleagues, we now have an holistic awareness of how DNA damage occurs and how the damage is rectified in bacteria as well as in higher organisms including human beings. A comprehensive understanding of DNA repair has proven crucial in the fight against cancer and other debilitating diseases.

The genetic information that guides the development, metabolism and reproduction of all living organisms and many viruses resides in the DNA. This biological information is stored in the DNA molecule as combinations of sequences that are formed by purine and pyrimidine bases attached to the deoxyribose sugar (*Figure* 1). The two strands of DNA[1] run in opposite directions in an antiparallel fashion and are held together by hydrogen bonds between nucleotides across the strands. The strict rules by which the nucleotides pair with each other imparts unique qualities to a DNA molecule which makes it ideal for storage and propagation of genetic information. Thus, the importance of DNA in every single aspect of life, as we know it, cannot be overstated.

In the 1950s and the early 1960s, it was generally accepted that DNA is a very sturdy molecule. Nothing else was expected of the chemical entity which manages to maintain large swathes of information, generation after generation , through millions of years. The first inroads into unraveling the phenomenon of DNA repair was made by Tomas Lindahl. His attempts to work with RNA

D N Rao is a professor at the Department of Biochemistry, Indian Institute of Science, Bengaluru. His research work primarily focuses on DNA interacting proteins in prokaryotes. This includes restriction-modification systems, DNA repair proteins from pathogenic bacteria and and proteins involved in horizontal gene transfer and DNA recombination.

Yedu Prasad is a graduate student working on his PhD under the guidance of D N Rao. His thesis work looks into the kinetic mechanism of a novel, phase-variable adenine methyltransferase from *H. pylori*. He is also working on unraveling the possible roles of this adenine methyltransferase in the physiology and pathogenesis of *H. pylori*.

[1] See Subramania Ranganathan, The Modular Construction of DNA Double Helix, *Resonance*, Vol.7, No.6, pp.39–44, 2002.

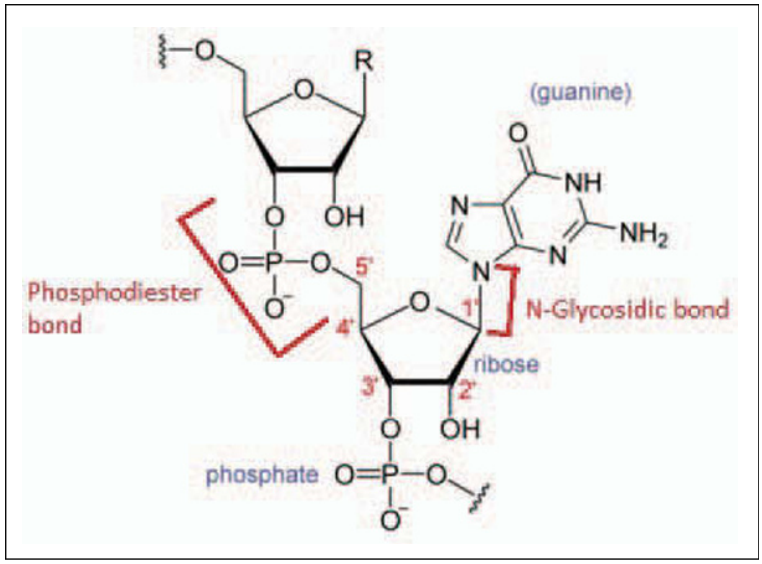

Figure 1. The chemical structure of a single nucleotide which forms the building block of a strand of DNA.
(Source: en.wikibooks.org)

The loss or damage of genetic information leads to serious consequences. The protection of DNA and thereby the maintenance of the genome is indispensable for all living organisms.

were strife with perils since it was extremely sensitive to thermal and pH fluctuations. Even though it is known that RNA is less stable than DNA, Lindahl found it odd that DNA was believed to be so much more resilient than the former. A few simple experiments were enough to confirm his suspicions that DNA, indeed, undergoes changes and can accumulate errors over time.

The loss or damage of genetic information leads to serious consequences. The protection of DNA and thereby the maintenance of the genome is indispensable for all living organisms. All processes that detect DNA damage and are involved in the correction of these errors are broadly referred to as DNA repair. DNA can undergo damage due to intrinsic (endogenous damage) or extrinsic factors (exogenous damage). Endogenous damage includes errors that crop up during essential metabolic processes such as DNA replication and damage caused by Reactive Oxygen Species produced within the cell. Exogenous damage is caused by external factors such as UV radiation, high energy radiation such as X-rays and gamma rays, mutagenic chemicals and viruses. Different types of DNA repair mechanisms are involved in the restoration of the genetic information, depending on the type of damage sustained by the DNA.

Keywords
DNA repair, DNA damage, base excision repair, nucleotide excision repair, methlyl-directed mismatch repair, Nobel Prize.

Direct Repair

Direct repair of DNA damage involves chemical reversal of the damage without breaking the phosphodiester backbone of the DNA. This process is not dependent on a template[2] since the damage does not alter the sequence within which it occurs. Albert Kelner was the first scientist to observe that bacteria that have been seemingly inactivated by fatal doses of UV radiation suddenly revived upon exposure to visible blue light [1]. Aziz Sancar was fascinated by this and took it upon himself to understand the biochemistry behind this process. He was able to identify and clone the enzyme, photolyase, which is involved in this process now known as photoreactivation (*Figure* 2). After identifying and cloning photolyase, Aziz Sancar completed his PhD and turned his attention to a light-independent process which repair UV induced damage. This dark system was being investigated at the Yale University School of Medicine, where Sancar joined as a laboratory technician. It was here that he embarked upon his Nobel Prize winning research.

Direct repair of DNA damage involves chemical reversal of the damage without breaking the phosphodiester backbone of the DNA.

[2] A single strand of DNA sequence which is used as a reference by DNA polymerase for the synthesis of the antiparallel DNA strand.

Six years later, Sancar revisited the photoreactivation project to understand the action mechanism of *E. coli* photolyase. Direct repair is the light-dependent reversal of pyrimidine dimer bonding which happens due to UV damage in the first place. Adjacent thymine or cytosine bases can covalently bond with each other when exposed to high-energy UV radiation thereby dis-

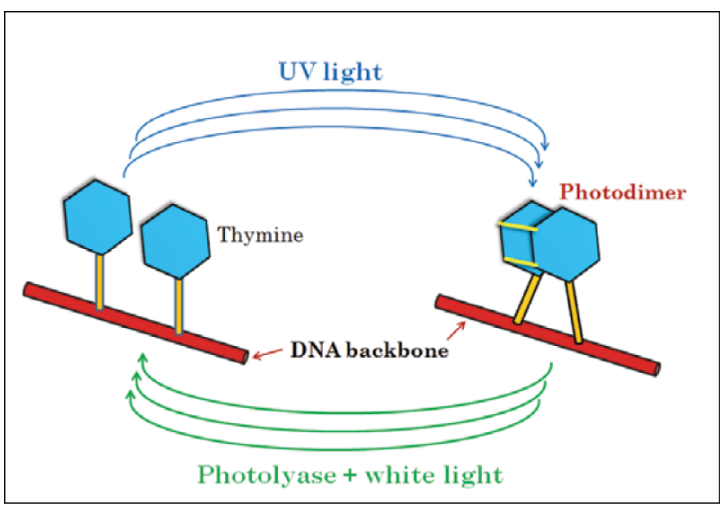

Figure 2. A simplified diagrammatic representation of photoreactivation mediated by photolyase. (Adapted from Griffiths *et al*., *An Introduction to Genetic Analysis*, 7th edition, 2000.)

Single-strand damage needs to be rectified before the next round of DNA replication.

rupting normal base-pairing across the two strands. Pyrimidine dimers are the primary causative agent of melanomas (skin cancer) in humans. Photolyase catalytically breaks down these inter-pyrimidine covalent bonds in a light-dependent manner. UV/blue light of wavelength 300–500 nm is essential for photolyase activity [2]. This chemical process is common in Nature and seems to be especially important in plants.

Template-Dependent DNA Repair

The antiparallel arrangement of a DNA molecule ensures that there is a copy of the genetic information on each strand. This is of paramount importance during DNA replication and the cell takes advantage of this property to make sure that there is no loss of information during the rectification of DNA damage. When there is damage on only one strand of the DNA, the repair can be performed using the genetic information on the antiparallel strand as the template. Single-strand damage needs to be rectified before the next round of DNA replication since the presence of mismatches will result in mutations on one of the daughter strands post-replication.

There are three types of excision repair systems extant in Nature; Base Excision Repair (BER), Nucleotide Excision Repair (NER) and Methyl-directed Mismatch Repair (MMR).

Base Excision Repair (BER)

Nitrogenous bases in DNA can be damaged by various chemical processes. These alterations interfere with the normal Watson–Crick base pairing between the two strands of DNA. Early work by Tomas Lindahl was instrumental in quantifying the incidence of spontaneous DNA-damage due to endogenous factors. He estimated that around 10,000 potentially mutagenic and cytotoxic changes occur in the human genome per day. This includes spontaneous deaminations, alkylation of bases, oxidation of bases and hydrolytic depurinations [3]. For example, methylcytosine can be converted to thymine by spontaneous deamination (*Figure* 3a). This converts the correctly base-paired C:G to T:G mismatch. Similarly, cytosine can undergo spontaneous deamination to form uracil. The presence of uracil in DNA is not tolerated. The dis-

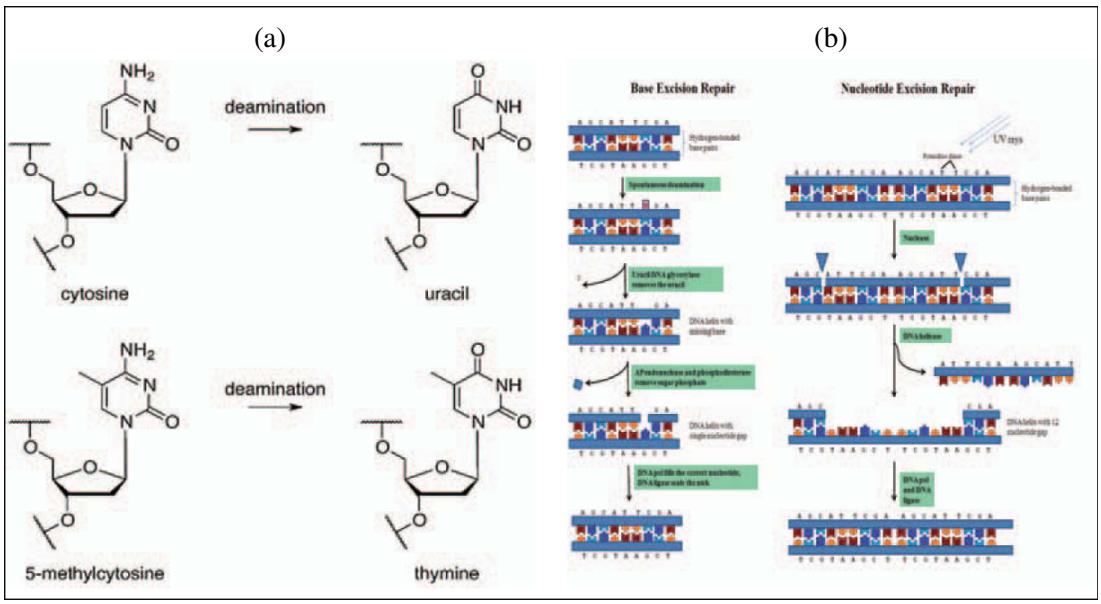

Figure 3. **(a)** The chemistry behind deamination reactions, the amino group is replaced with a keto group. (Source: www.atdbio.com) **(b)** A schematic representation of BER and NER highlighting the differences in the type of DNA damage that is acted upon, extent of the repair and the enzymes involved.
(Adapted from Alberts *et al.*, *Molecular Biology of the Cell*, 4th edition, 2002)

covery of uracil-DNA glycosylase by Lindahl was a major breakthrough in understanding BER (*Figure* 3b) [4]. This was followed by the discoveries of other DNA glycosylases such as 3-methyl adenine DNA glycosylase and others that catalyze the removal of oxidized bases [5].

Another common damage is alkylation of nitrogenous bases which causes formation of DNA distorting adducts[3] and subsequently block DNA replication and transcription. The removal of the damaged base by DNA glycosylases results in the formation of an apurinic/apyrimidinic site (AP site). An AP site can also be formed via direct action of free radicals or ionizing radiation. An AP endonuclease senses the distortion on the DNA molecule, and in one of the mechanisms to repair the AP site, a nick[4] is made on its 5′ side. A lyase or phosphodiesterase removes the base-less nucleotide.The resulting gap on the DNA is filled by DNA polymerase. In majority of cases (80–90%) the DNA polymerase fills in the single excised nucleotide and DNA ligase seals the phosphodiester backbone. In the remaining 10–20% cases, something known as 'long-patch pathway' (described very briefly in the following section) takes place. Depending on the type of DNA polymerase recruited, several nucleotides will be excised via 5′–3′ exonuclease activity and then resynthesized using the

DNA glycosylase and other enzymes are involved in the Base Excision Repair mechanism.

[3] In molecular genetics, a DNA adduct is a segment of DNA ionically bonded to a cancer-causing chemical.

other strand as template. Lindahl successfully reconstituted the entire BER (short patch and long patch) pathway in vitro using purified proteins [6].

Nucleotide Excision Repair (NER)

[4] A breakage in the phosphodiester bond backbone of only one of the strands of DNA.

Distortions in DNA due to UV-induced adducts, such as thymine dimers, are repaired by the Nucleotide Excision Repair (NER) pathway. The fundamental principles of NER are similar in bacteria and higher organisms although the latter involves more protein partners. In *E. coli*, NER is performed by the UvrABC endonuclease enzyme complex. Aziz Sancar, who was involved in characterizing the action of photolyase in UV damage DNA repair, started exploring the light-independent UV-damage DNA repair in bacteria. The group that Sancar joined at the Yale University School of Medicine already had genetic evidence for the involvement of three genes in UV sensitivity [7]. Within a few years, Sancar was able to identify and isolate the proteins involved in this process. The proteins were purified and reconstituted *in vitro* to show that they can identify UV damage and carry out repair mechanisms [8]. UvrA–UvrB complex scans the DNA for distortions. When a distortion is detected, UvrA in the complex is replaced with UvrC and the strand is cleaved on either side of the damage. UvrD helicase[5] is recruited to unwind the DNA within the two newly introduced nicks. The resulting single-stranded region is filled in by DNA polymerase I and the nick is sealed by DNA ligase. In majority of the cases (99%), the two nicks are spaced 12 nucleotides apart. Very rarely a much larger region is removed and is known as a long-patch pathway. It is not currently understood what factors determine the length of the excised patch (*Figure* 3b).

Nucleotide Excision Repair carried out by UvrABC endonuclease complex is well studied in *E.coli*

[5] An enzyme that specializes in unwinding of one DNA strand from the other by disrupting the hydrogen bonds between the nucleotide bases.

Methyl-Directed Mismatch Repair (MMR)

Mismatched nucleotide base pairs arise during every round of DNA replication. This occurs when the DNA polymerase incorporates a wrong nucleotide with respect to the template (for example, an adenine paired with a cytosine instead of a thymine).

The 3′–5′[6] exonuclease activity of DNA polymerase, also known as proofreading[7], manages to correct most of these mismatches during the process of DNA replication itself. However, there remains mismatches which escape the proofreading mechanism. Under *in vitro*, conditions it is estimated that replicative DNA polymerases operate with an error rate of 5×10^{-5} per base pair per generation (i.e., 5 errors in every 100,000 base pairs synthesized). However, it was observed that the actual error frequency in human germ line cells were closer to 1×10^{-8} errors per base pair per generation (i.e., 1 error in 100 million base pairs synthesized). These errors are different from those recognized by BER and NER since these bases are not chemically modified or are not part of DNA adducts. All these facts were indicative of the presence of a repair mechanism which specifically corrected mismatched base pairs. These mismatches, if not corrected immediately, have the potential to cause point mutations in the next round of DNA replication and get consolidated within the cell lineage. Although the occurrence of these mismatch events is probabilistically very low, the sheer size of the genome leads to these errors building up significantly.

It was Matthew Meselson who first observed that bacteria have the ability to correct mismatches that were artificially introduced in DNA. He developed a molecular tailoring technique, wherein mismatched base pairs could be introduced into the genome of DNA viruses. He observed that when these viruses were made to infect bacterial cells, the bacteria were able to correct the mismatches [9]. Naturally, this raises the question of how, within the mismatched base pairs, the bacterial cell determines which bases are the 'correct' ones. He put forward the hypothesis that the bacterial cell could either be employing a repair mechanism in close association with the replication machinery or it may be using methylated bases to distinguish the newly synthesized strands from the parent strands. Dam methylase enzyme in *E.coli* methylates the adenine within GATC sites found throughout the genome. It was discovered around this time that a mutation of *dam* gene and subsequent inactivation of adenine methylation caused a sub-

[6] In DNA biology, the 5′–3′ direction is defined as 'downstream' whereas the 3′–5′direction is defined as 'upstream'. A single DNA strand is synthesized in a particular direction owing to the biochemistry of polymerase reaction. Nucleotides are joined together via phosphodiester bonds. These bonds are forged between the 5th carbon on the deoxyribose of the incoming nucleotide and the 3rd carbon on the deoxyribose of the preceding nucleotide. This direction is defined as 5′–3′. Since DNA strands are antiparallel, the opposite strand is oriented in the 3′–5′ direction. A palindromic site is a locus where the sequence on both strands are identical in the 5′–3′ directions. Eg: 5′–GATC–3′ 3′–CTAG–5′

[7] The corrective measure employed by DNA polymerase I to detect and rectify mismatches during DNA replication. In this process the polymerase can track-back in the 3′–5′ direction to correct mismatched base pairs.

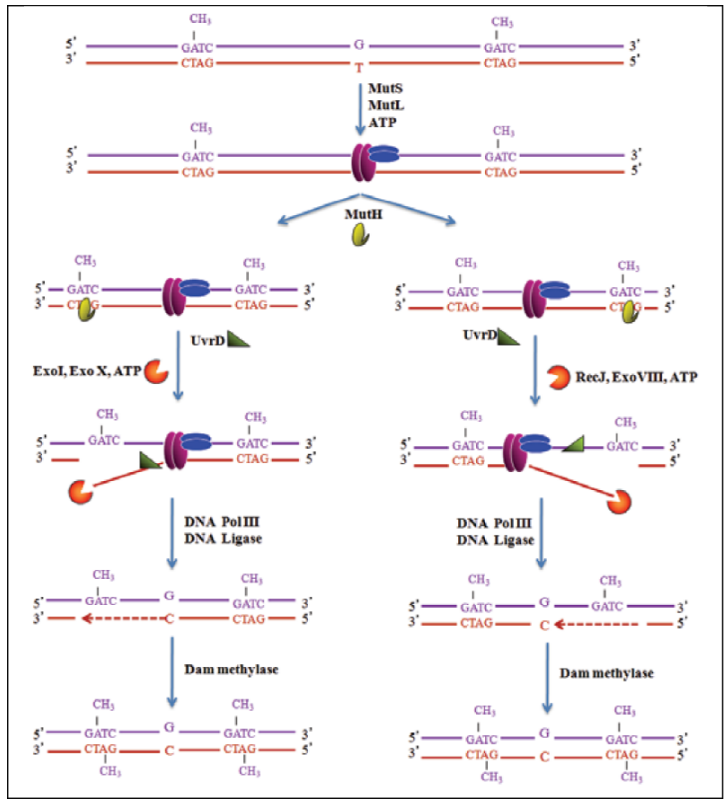

Figure 4. A simplified diagrammatic representation of Methyl-directed Mismatch Repair (MMR).

(Source: This figure has been drawn by us incorporating details about NMR that have been published by various groups over the years.)

stantial increase in the mutation rates throughout the genome [10]. Genetic studies by Barry W Glickman and Mirosav Radman demonstrated convincingly that *mutH*, *mutL*, *mutS* and *uvrD* genes were involved in MMR [11]. The confluence of all these works started painting a clearer picture of MMR in *E. coli*. However, it was still unclear how the cell managed to identify the nascent[8] strands so that the mismatches are rectified without altering the genetic blue print.

It was at this point that Paul Modrich started collaborating with Meselson to explore the methylation aspect of DNA mismatch repair (*Figure* 4).

Meselson's experiments with the viral DNA containing mismatches were repeated. However, one of the strands of these constructs were pre-methylated using Dam methylase at a GATC site far away (1000 bp) from where the mismatched base pairs were lo-

[8]The newly synthesized strand after a fresh cycle of DNA replication.

cated. They observed that the bacteria corrected the mismatches on the unmethylated strand every time [12]. Thus they were convinced that *E. coli* used the methylation status of DNA to identify the newly synthesized strands and use that information to rectify mismatched base pairs. Modrich proceeded to purify MutH, MutL, MutS and UvrD proteins and reconstituted the entire MMR pathway in cell-free conditions along with purified DNA polymerase III, Exonuclease I, DNA ligase and Single-Strand Binding protein (SSB). A DNA duplex containing mismatches along with a distant methylated site was used as substrate in this cell-free assay. The reconstituted proteins were able to correct the DNA mismatches on the unmethylated strand. This work was published in a landmark paper in 1989 [13]. In this and subsequent research, Modrich demonstrated the individual roles of MutS, MutH, MutL and UvrD proteins in MMR. MutS scans the DNA and identifies mismatched base pairs which the DNA polymerase I proofreading mechanism missed. MutH binds to the hemimethylated[9] GATC sites and then MutL is recruited to act as a mediator between MutS and MutH. A nick is introduced by MutH on the newly synthesized strand at the GATC site nearest to the detected mismatch. UvrD helicase unwinds the nascent strand starting from the nick to a point shortly past the mismatched locus. UvrD is a 3′–5′ helicase. Thus, depending on which side the nick is made (5′ or 3′ side of the mismatch) UvrD is loaded onto either strand. The newly released single stranded DNA is digested by an appropriate exonuclease. If the nick was made on the 5′ end of the mismatch, RecJ or Exonuclease VII is recruited which can degrade single-stranded DNA in the 5′–3′ direction. If the nick was made on the 3′ end of the mismatch, a 3′–5′ exonuclease (Exo I) is recruited. SSBs protect the exposed single stranded parent strand while the exonucleases perform their tasks. DNA polymerase III fills the gap and DNA ligase seals the nicks resulting in a new stretch of DNA where the mismatch no longer exists.

Methyl - Directed Mismatch Repair uses methylation status of nascent strands to repair damage.

[9] When a DNA base, like adenine, is methylated only on one strand of a palindromic sequence, it is termed as a hemimethylated site.

The pioneering works of Lindahl, Sancar and Modrich (*Figure* 5) owe their inception to genetic studies which identified crucial

Tomas Lindahl
Clare Hall Laboratory,
Francis Crick Institute,
United Kingdom

Paul Modrich
Howard Hughes Medical
Institute, Duke University
School of Medicine, USA

Aziz Sancar
University of North Carolina,
USA

Figure 5. The Nobel Prize in Chemistry 2015 was awarded jointly to Tomas Lindahl, Paul Modrich and Aziz Sancar "*for mechanic studies of DNA repair*".

genes involved in the DNA damage response. They proceeded to identify the individual proteins and understand the biochemical mechanisms involved in the process of DNA repair. It is worth noting that they embarked upon these endeavours at a time when molecular cloning and protein purification techniques were in their infancy. In recent years, numerous structural studies have reconfirmed the results that had been revealed by the genetic and biochemical research on DNA repair. The individual proteins involved in various repair pathways have been crystallized in their free as well as DNA-bound forms. Furthermore, high resolution 3D structures[10] of entire functional protein complexes have also been determined. All these genetic, biochemical as well as structural investigations have contributed to our understanding of DNA repair pathways.

The initial studies that identified proteins involved in BER, NER and MMR were done in *E. coli*. However, Lindahl, Sancar and

[10] See Hanudatta S Atreya, Structures of Biomolecules by NMR Spectroscopy, *Resonance*, Vol.20, No.11, pp.1033–1039, 2015; See Ramnathan Natesh, Crystallography Beyond Crystals: PX and SPoxyo EM, Vol.19, No.12, pp.1177–1196, 2015.

DNA Repair Pathway	Disease	Clinical Manifestations
Nucleotide Excision Repair	Xeroderma Pigmentosum	Sunlight hypersensitivity, high cancer incidence, premature aging, neurological abormalities
	Cockayne Syndrome	Sunlight hypersensitivity, mental retardation, dwarfism.
	Trichothiodystrophy	Sensitive skin, brittle hair and nails, frequent physical and mental retardation.
Base Excision Repair	Various forms of cancer. Premature aging.	Increased incidence of cancer (including lead-related carcinogenesis), neurological disorders.
Methyl-directed Mismatch Repair	Hereditary Non-Poly posis Colorectal Cancer (HNPCC)	High incidence of colon cancer.

Table 1. Human diseases associated with mutational inactivation of proteins integral to different DNA repair pathways.

Modrich have proceeded on to explore the human equivalents of these pathways. Thanks to their pioneering efforts, we now have a holistic understanding about the ways by which DNA-repair mechanisms constantly work against the tide of DNA damages our cells encounter; inactivation of even a single protein partner within these myriad pathways may cause cancer and other serious diseases (*Table* 1).

Suggested Reading

[1] A Kelner, Effect of visible light on the recovery of *Streptomyces griseus conidia* from ultraviolet irradiation injury, *Proc. Natl. Acad. Sci.*, USA, Vol.35, pp.73–79, 1949.

[2] G B Sancar, F W Smith, R Reid, G Payne and M Levy, Action mechanism of *Escherichia coli* DNA photolyase I. Formation of the enzyme-substrate complex, *J. Biol. Chem.*, Vol.262, pp.478–485, 1987.

[3] T Lindahl, B Nyberg, Rate of depurination of native deoxyribonucleic acid, *Biochemistry*, Vol.11, pp.3610–3618, 1972.

[4] T Lindahl, An N-glycosidase from *Escherichia coli* that releases free uracil from DNA containing deaminated cytosine residues, *Proc. Natl. Acad. Sci.*, USA, Vol.71, pp.3649–53, 1974.

[5] T Lindahl, New class of enzymes acting on damaged DNA, *Nature*, Vol.259, pp.64–66, 1976.

[6] G Dianov and T Lindahl, Reconstitution of the DNA base excision-repair pathway, *Curr. Biol*., Vol.4, pp.1069–1076, 1994.

[7] P Howard-Flanders, R P Boyce, L Theriot, Three loci in *Escherichia coli* K-12 that control the excision of pyrimidine dimers and certain other mutagen products from DNA, *Genetics*, Vol.53, pp.1119–1136, 1966.

[8] A Sancar, W D Rupp, A novel repair enzyme: UVrABC excision nuclease of *Escherichia coli* cuts a DNA strand on both sides of the damaged region, *Cell*, Vol.33, pp.249–260, 1983.

[9] R Wagner, M Meselson, Repair tracts in mismatched DNA heteroduplexes, *Proc. Natl. Acad*., Sci. USA, Vol.73, pp.4135–4139, 1976.

[10] M G Marinus, N R Morris, Pleiotropic effects of a DNA adenine methylation mutation (dam-3) in *Escherichia coli* K12, *Mutat. Res. – Fundam. Mol. Mech. Mutagen*., Vol.28, pp.15–26, 1975.

[11] B W Glickman, M Radman, *Escherichia coli* mutator mutants deficient in methylation-instructed DNA mismatch correction, *Proc. Natl. Acad. Sci*., USA, Vol.77, pp.1063–1067, 1980.

[12] P J Pukkila, J Peterson, G Herman, P Modrich, M Meselson, Effects of high levels of DNA adenine methylation on methyl-directed mismatch repair in *Escherichia coli*, *Genetics*, Vol.104, pp.571–582, 1983.

[13] R S Lahue, K G Au, P Modrich, DNA mismatch correction in a defined system. *Science*, Vol.245, No.4914, pp.160–164, 1989.

[14] Errol C Friedberg, Graham C Walker, Wolfram Siede and Richard D Wood, *DNA repair and mutagenesis*, 2nd Edition, ASM Press, 2006.

Address for Correspondence
D N Rao and Y Prasad*
Department of Biochemistry
Indian Institute of Science
Bengaluru 560 012, India.
Email:
dnrao@biochem.iisc.ernet.in
*yeduprasad108@gmail.com