DNA Repair

International Edition: DOI: 10.1002/anie.201509770
German Edition: DOI: 10.1002/ange.201509770

DNA Repair

*Thomas Carell**

biochemistry · DNA damage · DNA repair · genetic information · medicinal chemistry

The DNA double helix encodes the genetic information of almost all organisms on earth. It was Friedrich Miescher who discovered nucleic acids, which he called nuclein, back in 1869. Erwin Chargaff uncovered the base components of DNA, adenine (A), cytosine (C), guanine (G), and thymine (T), and reported the rules that are known today as the Chargaff rules and describe the 1:1 relation between A and T as well as G and C.[1] These reports, together with X-ray data obtained by Rosalind Franklin and Maurice Wilkins, laid the foundation for the discovery of the famous DNA double-helix structure by James Watson and Francis Crick.[2] Based on this structure, they formulated the molecular principles of self-complementary base pairing, which form the basis for the faithful replication of the DNA duplex. This finally allows genetic information to be passed down from generation to generation.

Since the discovery of the DNA double-helix structure it has been a riddle how this complex molecule can safely store information. Even if the replication machinery that copies the DNA molecule by pairing A with T and G with C makes only one mistake in one million steps, it would still lead to thousands of mismatches when, for example, a full human genome with its 10^9 base pairs is copied. Even if the half life of cytosine, for example, which easily deaminates to uracil, was a million years, with 10^9 bases in a human genome of which a fourth are cytosines, thousands of deamination reactions would be expected to occur per day and cell.[3] Furthermore, the genetic material is surrounded by mitochondria, which produce reactive oxygen species (ROS), some of which attack DNA molecules to produce a wide range of oxidative-degradation products. Furthermore, DNA is also degraded by UV light ($\lambda < 300$ nm), which is a natural part of sunlight, resulting in DNA photolesions. All of these lesions lead to a constant loss of genetic information. Overall, these processes basically eliminate the ability of the DNA duplex to encode a complex human being.

At this point, DNA repair comes into play; this years' Nobel Prize in Chemistry was awarded to Tomas Lindahl (base excision repair), Paul Modrich (mismatch repair), and Aziz Sancar (repair of chemically complex lesions) for research in this area. These researchers are pioneers in the DNA repair field. They reported essential mechanistic work on how the involved repair enzymes and proteins function. Once a lesion is detected, the repair systems trigger a series of events that finally lead to the removal of the lesion from the DNA duplex and the restauration of the genetic information.

Tomas Lindahl realized early on that DNA, although surprisingly robust, is far too fragile to fulfil its function as a genetic storage unit.[3,4] He concluded that something like DNA repair must exist. Tomas Lindahl worked mostly on the base excision repair (BER) system, which carefully removes damaged bases from the genome. The whole BER system is based on a set of different lesion-specific DNA glycosylases.[5] These enzymes recognize either one specific lesion or a small set of DNA lesions with a common structural motif. The enzymes employ a sophisticated molecular recognition step that allows them to precisely distinguish damaged bases from undamaged ones. Today, many repair glycosylases are known, and for most of them, even crystal structures together with their cognate substrates have been reported.[6] In particular, Tomas Lindahl discovered the uracil glycosylase, which removes mismatched uracil from a U:G base pair that results from deamination of C to U (Figure 1).[7] Other glycosylases include the human protein hOGG1, which removes 8-oxo-dG,[8] and the bacterial glycosylase MutM, which removes FaPy-dG lesions from the genome (Figure 2).[9] Both of these compounds are formed after oxidative damage of guanine bases.[10] Other repair glycosylases, such as MDB4[11] or the thymine DNA glycosylase (TDG),[12] are able to detect T:G mismatches. They remove the T base formed after deamination of 5mC, for example. All of these glycosylases employ a somewhat similar repair mechanism, which involves flipping the damaged base out of the duplex into the active site.[13] A nucleophile in the active site then attacks the C1′ carbon atom of the glycosidic bond; the glycosidic bond is then broken, and the base leaves to give a covalent DNA–glycosylase adduct, which is subsequently hydrolyzed to give an abasic site. These abasic sites are recognized by endonucleases, which now cleave the phosphodiester bonds to create a single nucleotide gap that is subsequently filled by the insertion of a new base. Base excision repair is a versatile repair system that removes many lesions from our genome in a straightforward process. The system protects our genome from damages created by deamination, alkylation, and oxidative stress. These are lesions that only result in minor DNA duplex distortion.

Paul Modrich pioneered mechanistic work in the field of mismatch repair, which is a complex DNA repair process that

[*] Prof. Dr. T. Carell
Center for Integrated Protein Science, CiPS^M, Department für Chemie
Ludwig-Maximilians Universität München
Butenandtstrasse 5-13, 81377 München (Germany)
E-mail: thomas.carell@lmu.de
Homepage: http://www.carellgroup.de

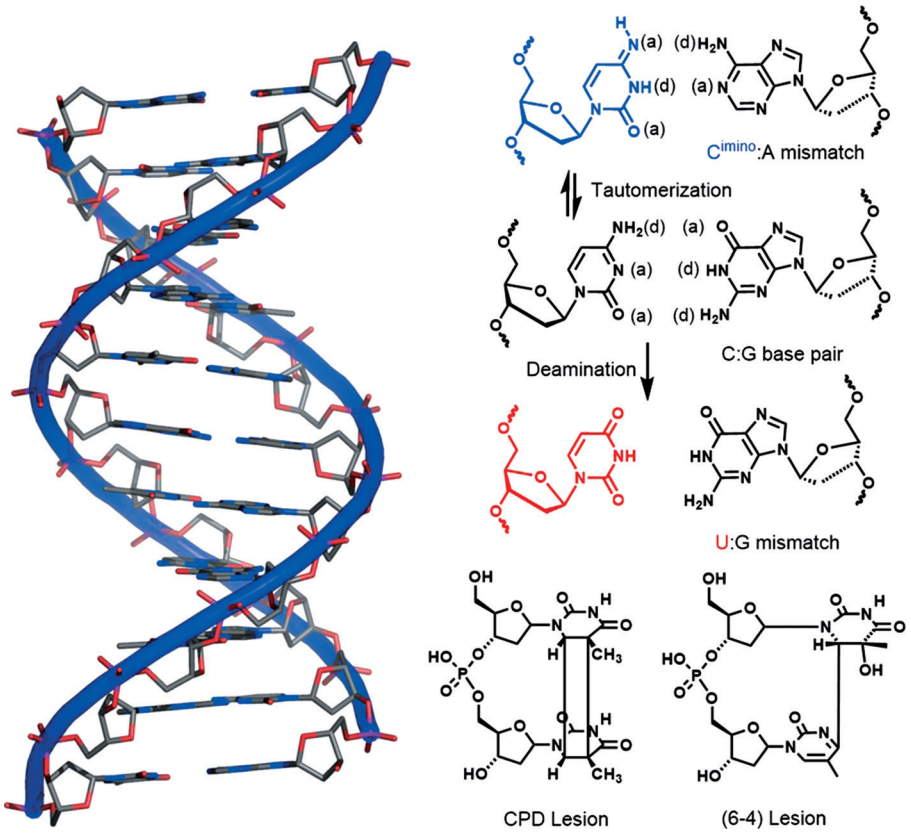

Figure 1. Left: Depiction of the DNA double helix (B-form) with the sugar phosphodiester backbone in blue and the nucleobases pointing towards each other. Right: C:G base pair (black). The C base can undergo deamination to give a U:G mismatch, which is repaired by the uracil DNA glycosylase (UDG). The U base is shown in red. The C base can also tautomerize, and this form (blue) is able to base-pair with A to give a C^{imino}:A base pair, which is repaired by the mismatch repair system, which consists of MutS, MutH, and MutL in bacteria. Also depicted are the UV-induced cyclobutane–pyrimidine dimer (CPD) and (6-4) lesions.

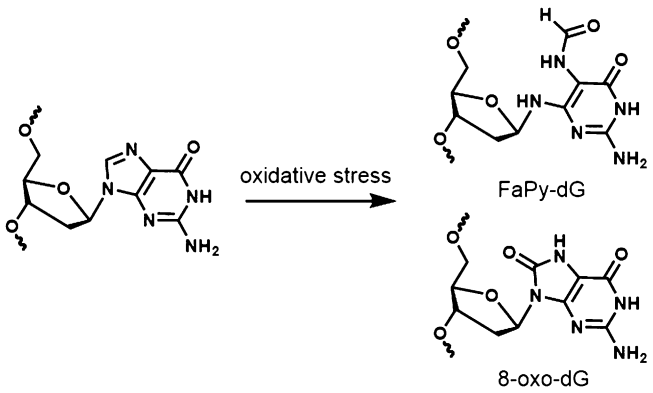

Figure 2. Oxidative stress: Reactions with reactive oxygen species lead to the formation of 8-oxo-dG and FaPy-dG lesions.[10a]

is able to eliminate base-pairing mistakes generated during replication of the DNA duplex by DNA polymerase.[14] The chemical basis for this type of DNA damage is often base tautomerization. DNA bases can exist in different tautomeric states, and although the tautomers that are the basis for Watson–Crick-type base pairing are by far the most dominant forms, other tautomeric states exist as well. If they appear inside the DNA polymerase during the replication process, they can lead to the selection of a wrong counterbase during the copying reaction. A famous example is the DNA base cytosine (Figure 1). The dominating tautomer possesses a H-bond donor (d)/H-bond acceptor (a)/H-bond acceptor (a) arrangement that enables base pairing with guanine with its complementary (a)-(d)-(d) arrangement in the Watson–Crick mode. However, when the C base tautomerizes to the imino form, an (a)-(d)-(a) arrangement results, which will be base-paired with the (d)-(a) motif of adenine to give a C:A mismatch. DNA polymerases are astonishingly promiscuous, and indeed, crystal structures of DNA polymerases with DNA bases in such rare tautomeric states were recently reported.[15]

The repair of mismatches is complex because the repair machinery needs to distinguish between the (old) DNA strand that forms the template in the copying process and the newly synthesized strand in which the mistake occurred. This is important because the system has to decide whether it needs to remove the C or the A base from the mismatch in the above example. Whereas the details of this strand-recognition process are not entirely clear, it is best known for bacteria. Here, three enzymes, MutS,[16] MutH, and MutL, are needed to perform the repair reaction. These enzymes use the fact that some adenine bases are methylated at the N6 position.[17] Shortly after DNA replication, the newly synthesized daugh-

ter strand lacks these methylated bases, which allows the repair system to identify the newly synthesized strand. During mismatch repair, bacterial MutS (as a dimer, $(MutS)_2$) recognizes the mismatched base, and MutH binds to hemi-methylated sites. MutL then functions as a mediator between MutS and MutL forcing them together, which leads to the activation of MutH, which then cleaves the phosphodiester backbone close to the mismatch in the non-methylated strand. The whole repair complex subsequently slides along the duplex to remove a whole piece of DNA around the mismatch to create a single-stranded gap. These processes are followed by de novo DNA synthesis to fill the gap and final ligation. In humans, the mismatch repair mechanism is still under intense investigation.[18]

Aziz Sancar is a pioneer in the DNA repair field who investigated how chemically complex DNA lesions, such as UV-induced lesions, are repaired.[19] For UV lesions, specific repair enzymes exist, which are able to repair the lesions directly in the double strand. The repair enzymes are the CPD and (6-4) DNA photolyases.[19b] They enable bacteria to survive after UV exposure in the presence of long-wavelength sunlight, a process that is called photoreactivation, and they are also important in plants.[20] UV irradiation leads to the dimerization of two adjacent T bases by either a $[2\pi+2\pi]$ cycloaddition reaction, which leads to cyclobutane pyrimidine dimers (CPDs) in a superfast process,[21] or by a Paternó–Büchi reaction, which gives rise to (6-4) photo adducts (Figure 1). To repair these lesions, DNA photolyases employ a complex repair process that is based on energy- and electron-transfer events.[19b] Photolyases contain two cofactors, one of which (either a methenyltetrahydrofolate or a deazaflavin) absorbs light with wavelengths of about 350–400 nm and transfers the excitation energy, by a Förster mechanism, to a reduced and deprotonated flavin ($FADH^-$) present in the active site. Crystal structures of photolyases[22] and co-crystal structures of the CPD[23] and (6-4) photolyases[24] (Figure 3) in complex with lesion-containing DNA show that the enzyme flips the whole dinucleotide lesion out of the duplex into the active site to bring it into close proximity to $FADH^-$*. The key step of the repair process is an electron transfer from $FADH^-$* to the lesion, which subsequently fragments back to the thymine monomers.

Although mismatch repair, base excision repair, and photoreactivation are central elements of the genome repair systems in nature, most species, including humans, also possess a further complex repair system called nucleotide excision repair (NER). This repair system, which was also intensively investigated by A. Sancar, removes not only UV lesions,[25] but also a variety of bulky-adduct lesions and cisplatin DNA lesions, formed during chemotherapy with cisplatin, from the genome.[26] It requires the complex interplay of many proteins.[27] Today, we distinguish between global NER processes[28] and NER processes that are transcription-coupled,[29] and we understand that defects of the NER system in whatever subsystem are particularly harmful for humans.[30] The proteins involved in both global and transcription-coupled NER have been identified in the meantime, but the molecular details that ultimately lead to damage recognition are not yet fully understood. However, recently reported crystal structures of key NER proteins in complex with specific lesion-containing DNA start to uncover how the different lesions are recognized.[31–33] The history of

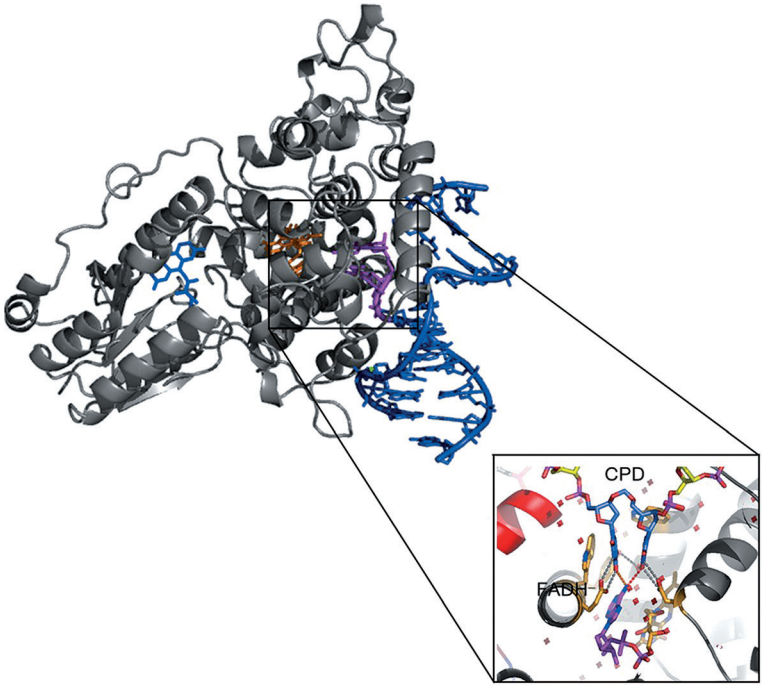

Figure 3. Crystal structure of the CPD photolyase (gray) with the deazaflavin cofactor F_{420} (blue) and the FAD cofactor (orange). The DNA duplex is shown in dark blue with the flipped-out CPD lesion (magenta) that is formed between the DNA bases T7 and T8. The active site is indicated by the black box. Bottom right: View into the active site, showing the repaired CPD lesion in complex with $FADH^-$. The $FADH^-$ ion has a U-type conformation.[22a,23] Some water molecules are also found in the active site (+).

the discovery of the NER system is interesting, and readers are referred to an early excellent review on this topic.[34]

Acknowledgements

I thank the Deutsche Forschungsgemeinschaft for generous support through the grants SPP1784 (CA275/10-1), SFB749 (A5), SFB646 (B1), and SFB1032 (A4).

How to cite: *Angew. Chem. Int. Ed.* **2015**, *54*, 15330–15333
Angew. Chem. **2015**, *127*, 15546–15549

[1] a) E. Chargaff, *Science* **1971**, *172*, 637–642; b) R. Dahm, *Hum. Genet.* **2008**, *122*, 565–581.
[2] J. D. Watson, F. H. Crick, *Nature* **1953**, *171*, 737–738.
[3] T. Lindahl, B. Nyberg, *Biochemistry* **1974**, *13*, 3405–3410.
[4] a) T. Lindahl, A. Andersson, *Biochemistry* **1972**, *11*, 3618–3623; b) T. Lindahl, B. Nyberg, *Biochemistry* **1972**, *11*, 3610–3618; c) T. Lindahl, O. Karlstrom, *Biochemistry* **1973**, *12*, 5151–5154.
[5] T. Lindahl, *Nature* **1976**, *259*, 64–66.
[6] a) R. Savva, K. McAuley-Hecht, T. Brown, L. Pearl, *Nature* **1995**, *373*, 487–493; b) S. S. Parikh, C. D. Putnam, J. A. Tainer, *Mutat. Res.* **2000**, *460*, 183–199; c) J. C. Fromme, A. Banerjee, G. L. Verdine, *Curr. Opin. Struct. Biol.* **2004**, *14*, 43–49.
[7] T. Lindahl, S. Ljungquist, W. Siegert, B. Nyberg, B. Sperens, *J. Biol. Chem.* **1977**, *252*, 3286–3294.
[8] A. Klungland, M. Hoss, D. Gunz, A. Constantinou, S. G. Clarkson, P. W. Doetsch, P. H. Bolton, R. D. Wood, T. Lindahl, *Mol. Cell* **1999**, *3*, 33–42.
[9] F. Coste, M. Ober, T. Carell, S. Boiteux, C. Zelwer, B. Castaing, *J. Biol. Chem.* **2004**, *279*, 44074–44083.
[10] a) J. Cadet, T. Douki, J. L. Ravanat, *Acc. Chem. Res.* **2008**, *41*, 1075–1083; b) J. Cadet, T. Douki, J. L. Ravanat, *Hum. Genet.* **2010**, *49*, 9–21.
[11] B. Hendrich, U. Hardeland, H. H. Ng, J. Jiricny, A. Bird, *Nature* **1999**, *401*, 301–304.
[12] P. Gallinari, J. Jiricny, *Nature* **1996**, *383*, 735–738.
[13] a) A. R. Dinner, G. M. Blackburn, M. Karplus, *Nature* **2001**, *413*, 752–755; b) K. Sadeghian, D. Flaig, I. D. Blank, S. Schneider, R. Strasser, D. Stathis, M. Winnacker, T. Carell, C. Ochsenfeld, *Angew. Chem. Int. Ed.* **2014**, *53*, 10044–10048; *Angew. Chem.* **2014**, *126*, 10208–10212.
[14] R. R. Iyer, A. Pluciennik, V. Burdett, P. L. Modrich, *Chem. Rev.* **2006**, *106*, 302–323.
[15] W. Wang, H. W. Hellinga, L. S. Beese, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 17644–17648.
[16] J. J. Warren, T. J. Pohlhaus, A. Changela, R. R. Iyer, P. L. Modrich, L. S. Beese, *Mol. Cell* **2007**, *26*, 579–592.
[17] a) A. L. Lu, S. Clark, P. Modrich, *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 4639–4643; b) P. J. Pukkila, J. Peterson, G. Herman, P. Modrich, M. Meselson, *Genetics* **1983**, *104*, 571–582.
[18] J. Pena-Diaz, J. Jiricny, *Trends Biochem. Sci.* **2012**, *37*, 206–214.
[19] a) A. Sancar, C. S. Rupert, *Gene* **1978**, *4*, 295–308; b) A. Sancar, *Chem. Rev.* **2003**, *103*, 2203–2237.
[20] a) A. Kelner, *Proc. Natl. Acad. Sci. USA* **1949**, *46*, 73–79; b) R. Dulbecco, *Nature* **1949**, *163*, 949.
[21] W. J. Schreier, T. E. Schrader, F. O. Koller, P. Gilch, C. E. Crespo-Hernandez, V. N. Swaminathan, T. Carell, W. Zinth, B. Kohler, *Science* **2007**, *315*, 625–629.
[22] a) H. W. Park, S. T. Kim, A. Sancar, J. Deisenhofer, *Science* **1995**, *268*, 1866–1872; b) T. Tamada, K. Kitadokoro, Y. Higuchi, K. Inaka, A. Yasui, P. E. de Ruiter, A. P. Eker, K. Miki, *Nat. Struct. Biol.* **1997**, *4*, 887–891.
[23] A. Mees, T. Klar, P. Gnau, U. Hennecke, A. P. Eker, T. Carell, L. O. Essen, *Science* **2004**, *306*, 1789–1793.
[24] M. J. Maul, T. R. Barends, A. F. Glas, M. J. Cryle, T. Domratcheva, S. Schneider, I. Schlichting, T. Carell, *Angew. Chem. Int. Ed.* **2008**, *47*, 10076–10080; *Angew. Chem.* **2008**, *120*, 10230–10234.
[25] J. T. Reardon, A. Sancar, *Genes Dev.* **2003**, *17*, 2539–2551.
[26] J. H. Hoeijmakers, *Nature* **2001**, *411*, 366–374.
[27] W. L. de Laat, N. G. Jaspers, J. H. Hoeijmakers, *Genes Dev.* **1999**, *13*, 768–785.
[28] L. C. Gillet, O. D. Scharer, *Chem. Rev.* **2006**, *106*, 253–276.
[29] P. C. Hanawalt, G. Spivak, *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 958–970.
[30] E. C. Friedberg, *Nat. Rev. Cancer* **2001**, *1*, 22–33.
[31] J. H. Min, N. P. Pavletich, *Nature* **2007**, *449*, 570–575.
[32] a) E. S. Fischer, A. Scrima, K. Bohm, S. Matsumoto, G. M. Lingaraju, M. Faty, T. Yasuda, S. Cavadini, M. Wakasugi, F. Hanaoka, S. Iwai, H. Gut, K. Sugasawa, N. H. Thoma, *Cell* **2011**, *147*, 1024–1039; b) A. Scrima, R. Konickova, B. K. Czyzewski, Y. Kawasaki, P. D. Jeffrey, R. Groisman, Y. Nakatani, S. Iwai, N. P. Pavletich, N. H. Thoma, *Cell* **2008**, *135*, 1213–1223.
[33] S. C. Koch, J. Kuper, K. L. Gasteiger, N. Simon, R. Strasser, D. Eisen, S. Geiger, S. Schneider, C. Kisker, T. Carell, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 8272–8277.
[34] P. C. Hanawalt, R. H. Haynes, *Sci. Am.* **1967**, *216*, 36–43.

Received: October 19, 2015
Published online: November 19, 2015

