

integrity

Vivienne Baillie Gerritsen

Mistakes and misinterpretation are common incidents. We are usually aware of those we can see, or hear, such as a bird flying into an undetected window or a person who has misunderstood a question. Yet, mistakes are also a recurring phenomenon at a level no living being can readily observe: the molecular level. Take, for instance, mutations that occur in an organism's DNA. Though the evolution of species may sometimes thrive on such flaws, they are also frequently the source of serious drawbacks. This is why, over time, Nature has devised different systems whose role is to detect mistakes in DNA and offer a means to fix them. Until recently, repair systems had only been considered on double-stranded DNA where the untouched strand acts as a template to re-establish an error found on the second. But when damage is found on single-stranded DNA, what happens? Is it even possible to make amends? Against all odds: yes. Such a repair system involves a protein that has been coined HMCES, which is found throughout the three domains of life, and even in certain viruses it seems.



'Children on the Beach'

by Joaquin Sorolla Y Bastida (1863-1923)

Life depends heavily on DNA. It is also very dependent on how faithfully DNA is replicated when cells are dividing. The DNA of two daughter cells should be as identical as possible to the parent cell – especially in the regions that are biologically crucial. Some mutations turn out to be of no consequence whatsoever while others give rise to changes that turn out to be useful. Natural selection flourishes on this kind of mutation. But there are also mutations that have serious consequences on an organism's well-being. An example: DNA is made

up of two types of bases known as pyrimidines and purines. However, sometimes abasic sites (neither purines nor pyrimidines) are found where classical bases should be, i.e. A, C, G or T. In fact, it has been estimated that about 18'000 abasic sites are generated in one human cell every day! These errors are spontaneous or the fruit of sources such as ionizing radiation or UV radiation for instance. With this in mind, it is not difficult to understand that repair systems are more than necessary and are an integral part of every cell – from bacteria all the way to humans.

It may seem obvious to us today, but barely more than half a century ago, many researchers were unwilling to admit that a molecule as 'simple' as DNA could be at the heart of life. Here is a molecule composed of only four different bases, and they were to believe that the complexity and intricacies of life are of its making. The molecular biologists James Watson (USA) and Francis Crick (UK) published their paper describing the double helical structure of DNA in 1953. It was a landmark in the history of molecular biology. However, how DNA was handed down from generation to generation, how it replicated, was what would really open almost limitless horizons.

Watson and Crick had outlined a model of replication now known as 'semi-conservative'. In this model, each DNA single strand is used as a template to synthesize the second. Consequently, daughter cell DNA is composed of one 'new' DNA

strand and a second that is from the parent. Watson & Crick's model was theoretical however and needed to be tested – especially as their theory was not the only one in those days. One theory actually propounded that certain proteins could be the moulds for novel DNA. The problem was solved five years later thanks to a very elegant experiment carried out by two American molecular biologists Matthew Meselson and Franklin Stahl, who showed that DNA did indeed replicate in a semi-conservative fashion.

60 years on, we have a fairly good idea how this proceeds on the molecular scale. Several steps lead to DNA replication. In a nutshell, first the double-stranded DNA has to be uncoiled, then it has to be opened up so that each strand can be used as a template, then the daughter DNA has to be synthesized. Each step is regulated and performed by many different systems, and emergency systems are also required to detect errors and repair them. This said, how a lesion located in single-stranded DNA (ssDNA) could possibly be repaired, has remained a mystery: there is no second strand to help out.

Recently, researchers observed that the presence of abasic sites on ssDNA causes the replicative machinery to stall, much as you would slow down to ponder over a road sign that makes no sense. Stalling has an SOS effect on the cell so that it ushers in an emergency repair squad, namely an enzyme known as HMCES (from 5-hydroxymethylcytosine binding ESC-specific). In fact, HMCES is expressed more abundantly when (and where) DNA is being replicated, no doubt to be able to respond to emergencies faster, and it can sense DNA lesions on ssDNA which is about to be replicated, i.e. at what is called the replication fork.

Once spotted, HMCES binds to a protein known as PCNA (Proliferating Cell Nuclear Antigen) which normally travels with the replicative machinery. At

the same time, HMCES forms a strong covalent crosslink with ssDNA at the abasic site. As HMCES does this, it literally shields the lesion from the system that would otherwise replicate it, thus avoiding downstream genetic instability. HMCES binds to ssDNA by way of a SRAP domain, or 'SOS Response Associated Peptidase' domain. SRAP proteins are widely spread across bacteria and eukaryotes. They all sport a conserved histidine and glutamic acid in their catalytic domain, besides carrying a cysteine residue which is almost invariably encoded as the second amino acid in their sequence. Then what? To date, no one knows. HMCES is not a repair protein per se. In fact, the HMCES-DNA crosslink seems to be an intermediate stage in the repair process, and the enzyme ultimately commits suicide via proteolysis. So HMCES' fundamental role is to prevent polymerases from processing abasic sites and causing downstream genetic instability, but it does not have the means to correct the mistake.

Understanding how HMCES-DNA crosslinks are resolved will help to unveil how ssDNA lesions are actually repaired. Further studies should also reveal how SRAP domains are able to detect abasic sites – possibly by sensing uncharacteristic DNA architecture. What is more, there are many different types of abasic sites which would indicate HMCES flexibility in substrate recognition. The more we know about a system such as that involving HMCES and the more we know about DNA replication and how cells deal with its flaws, the more we will be able to understand diseases such as cancer, for instance, whose driving force draws heavily on the ways cells have of copying DNA. Certainly, in only six decades, researchers have demonstrated – and in molecular detail – how crucial DNA is, and the part it plays, in life and its flow. We are light years away from the initial skepticism such a 'simple' molecule met. A perfect illustration, perhaps, of how simple is also so often beautiful.

Cross-references to UniProt

Abasic site processing protein HMCES, *Homo sapiens* (Human) : Q96FZ2

References

1. Mohni K.N., Wessel S.R., Zhao R., Wojciechowski A.C., *et al.*
HMCES maintains genome integrity by shielding abasic sites in single-stranded DNA
Cell 176: 144-153(2019)
PMID: 30554877
2. Davis T.H.
Meselson and Stahl: The art of DNA replication
PNAS 101:17895-17896(2004)
PMID: 15608065