Molecular biology

Demystifying the D-loop in DNA recombination

Upasana Roy & Eric. C. Greene

Homologous recombination is a mechanism for DNA repair that enables the exchange of genetic information between DNA molecules. Structural analysis reveals how the protein RecA orchestrates this process. **See p.801**

Our chromosomes are subjected to a constant barrage of DNA-damaging agents that can lead to the loss of genome integrity[1,2]. Homologous recombination is an evolutionarily conserved type of DNA repair that can fix a particularly dangerous form of damage in which both strands of a DNA molecule break[3,4]. The crucial need for homologous recombination is reflected in the cancers and cancer-prone syndromes that are associated with defects in proteins, such as BRCA1 and BRCA2, that function in recombination[5]. On page 801, Yang *et al.*[6] report structural data that shed light on how the protein RecA facilitates a key early step in the process of homologous recombination in bacteria.

During homologous recombination (Fig. 1a), DNA at the site of a double-strand break undergoes an enzymatic processing step that generates single-stranded overhangs called presynaptic single-stranded DNA. This presynaptic DNA then 'invades' an identical section of a double-stranded DNA molecule (termed homologous DNA) and pairs with the matching complementary strand, which results in displacement of the non-complementary strand of the double-stranded DNA[3,4]. This pairing event is called strand exchange, and it generates a DNA configuration referred to as a displacement loop (D-loop), which is an essential intermediate structure needed for the events that lead to DNA repair[3,4]. Yang and colleagues provide detailed data for D-loop structures that yield new insights into the mechanism of homologous recombination.

The complex DNA transactions that take place during recombination are mediated by members of the RecA family of recombinase enzymes. During this process, RecA proteins form long filaments on the presynaptic single-stranded DNA. This RecA-bound DNA then invades a double-stranded DNA molecule and separates the two strands. Whether this opened DNA matches the sequence of the RecA-coated DNA is then determined. This process must happen extremely quickly because many double-stranded DNA sequences might need to be sampled until the correct match is found.

The filaments formed by RecA enzymes stretch presynaptic single-stranded DNA by approximately 50% relative to its normal length[7]. The DNA is not stretched uniformly, however. The DNA bound within a so-called primary DNA-binding site of RecA is organized into triplets of nucleotide bases (one triplet per RecA molecule) that retain the dimensions of normal double-stranded DNA. But the DNA backbone (the chain of sugar and phosphate groups) between these triplets is greatly extended compared with its usual form[7,8].

When the complementary DNA strand pairs with the RecA-coated presynaptic single-stranded DNA, it also adopts this extended conformation and is stabilized mainly by base-pairing interactions with the presynaptic DNA. This provides a way to identify a sequence match that has the correct nucleotide-pairing interactions[7]. However, many aspects of this process remained unclear. In particular, there was a lack of structural information on how RecA-bound DNA first interacts with a double-stranded DNA molecule and then opens the strands up to form a D-loop.

Yang *et al.* used using high-resolution cryo-electron microscopy to reveal the structure of D-loop intermediates. Structural studies of RecA have proved notoriously difficult, because RecA can form filaments of different lengths bound to DNA. To overcome this problem, the authors fused nine RecA proteins together and mutated the amino-acid residues at the ends of this 'mini-filament' to prevent any further addition of RecA molecules. The use of these short, well-defined RecA mini-filaments enabled the authors to uncover some key missing information about the process of recombination. Their findings illuminate the choreography of protein–DNA interactions that occur as RecA mediates D-loop formation.

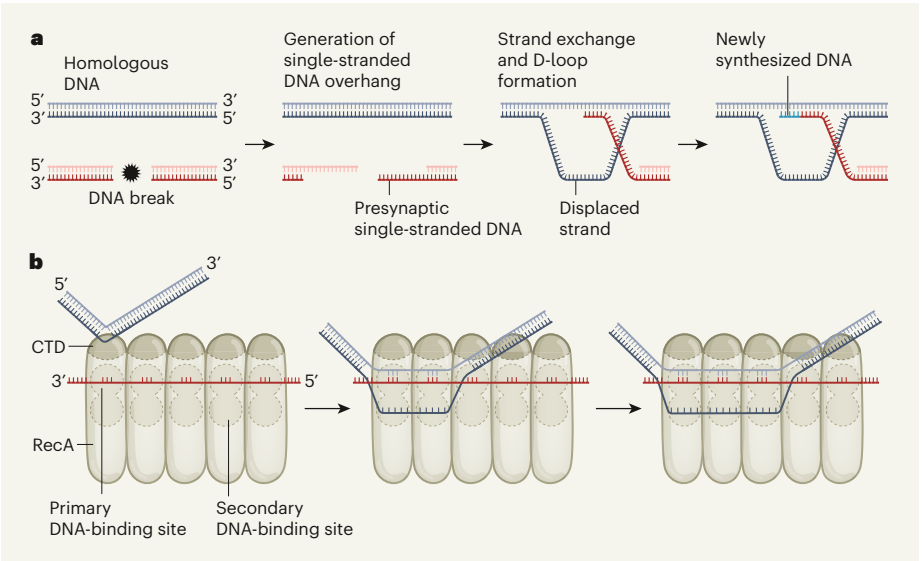

Figure 1 | The structure of a DNA D-loop. a, Broken DNA is repaired using an identical (homologous) double-stranded DNA molecule as a template. The broken DNA is processed to generate single-stranded overhangs termed presynaptic DNA. An overhang invades homologous DNA to pair with its complementary strand (light blue), thereby displacing the non-complementary strand (dark blue). This strand-exchange event generates a displacement loop (D-loop) structure. DNA synthesis then occurs, leading to DNA repair. **b**, Yang *et al.*[6] present structural data that illuminates early events during D-loop formation. RecA proteins bind to presynaptic DNA to form a filament of DNA and RecA. Double-stranded DNA initially binds to RecA's carboxy-terminal domain (CTD). It is then guided further inwards towards the centre of the filament, and the two strands of its DNA are separated. The complementary strand pairs with the presynaptic DNA in RecA's primary DNA-binding site (which holds three paired nucleotides), and the non-complementary strand binds RecA's secondary DNA-binding site (which binds to five unpaired nucleotides). When the DNA sequences match in the primary site, strand separation of the double-stranded DNA continues, and D-loop growth occurs in a direction corresponding to the 3′-to-5′ orientation of the presynaptic DNA.

The structural data suggest that at the start of the process (Fig. 1b), the carboxy-terminal domain (CTD) region of RecA binds to double-stranded DNA and guides it towards the primary DNA-binding site near the centre of the RecA-bound DNA filament. A small protein loop there helps to open up the double-stranded DNA. Once open, the displaced non-complementary strand binds to a region along the surface of the RecA–DNA filament called the secondary DNA-binding site. This site was already known to exist, but its location on the protein surface had not been fully defined. The non-displaced complementary DNA strand is positioned in a way that enables it to pair with the presynaptic single-stranded DNA.

The continued opening of the double-stranded DNA then proceeds preferentially in a 3′-to-5′ direction with respect to the orientation of the presynaptic DNA. However, if the newly interacting strands don't match, the probability of further opening of the double-stranded DNA at each successive RecA protein is substantially reduced. When the interacting strands match, the D-loop continues to be extended. In this case, the structure is further stabilized by binding of the displaced strand to the secondary DNA-binding site, as well as by base-pairing interactions between the complementary strand and the presynaptic single-stranded DNA.

Yang and colleagues also discovered that this process has a surprising feature – there is an asymmetry between the DNA interactions in the primary DNA-binding site and the secondary DNA-binding site. Three nucleotides of the stretched single-stranded DNA are bound to each RecA protein within the primary site, whereas each RecA protein binds to five nucleotides of DNA in its secondary site.

The functional purpose of this binding asymmetry is unknown. One possible benefit would be to enable some flexibility within the complementary single-stranded DNA such that if five of these bases unwind (as opposed to just three), then this unwound stretch of single-stranded DNA could, in principle, be positioned in three different registers in the primary binding site to test for different base-pairing interactions with the presynaptic DNA. This might help to ensure the quick formation of optimal base-pairing interactions in the primary DNA-binding site. In addition, once a correct sequence match is found, the binding asymmetry might help to propagate strand invasion by further opening up enough double-stranded DNA to help establish the base-pairing associated with the next RecA protein in the filament. These findings considerably advance our knowledge of the mechanism that underlies homologous recombination.

Even with the remarkable insights provided by Yang and colleagues' work, many key questions remain. How quickly can non-homologous sequences that don't match the presynaptic DNA strand be tested and rejected? Non-homologous sequences vastly outnumber homologous ones, and homologous recombination could not succeed if too much time were spent sampling the wrong sequences. How much complementarity between sequences of nucleotides is necessary for stable pairing interactions to occur, and what happens when sequences are not perfectly matched? Understanding the exact nature of these interactions could shed light on what determines whether a sequence is sufficiently homologous to support RecA-mediated recombination.

The version of RecA found in eukaryotes (organisms that have a nucleus in their cells) is known as Rad51. Understanding recombination in eukaryotes will require determination of the structures and mechanisms used by other accessory proteins that are also involved in recombination, because Rad51 is highly dependent on such protein co-factors[9,10]. The advent of cryo-electron microscopy as a tool for defining high-resolution structures of complex recombination intermediates offers the potential to address these and many other intriguing puzzles about homologous recombination.

Upasana Roy and **Eric C. Greene** are in the Department of Biochemistry and Molecular Biophysics, Columbia University Irving Medical Center, New York, New York 10032, USA.
e-mails: ur2129@cumc.columbia.edu; ecg2108@cumc.columbia.edu

1. Hoeijmakers, J. H. J. *Nature* **411**, 366–374 (2001).
2. Hoeijmakers, J. H. J. *N. Engl. J. Med.* **361**, 1475–1485 (2009).
3. Mehta, A. & Haber, J. E. *Cold Spring Harb. Perspect. Biol.* **6**, a016428 (2014).
4. Kowalczykowski, S. C. *Cold Spring Harb. Perspect. Biol.* **7**, a016410 (2015).
5. Prakash, R., Zhang, Y., Feng, W. & Jasin, M. *Cold Spring Harb. Perspect. Biol.* **7**, a016600 (2015).
6. Yang, H., Zhou, C., Dhar, A. & Pavletich, N. P. *Nature* **586**, 801–806 (2020).
7. Chen, Z., Yang, H. & Pavletich, N. P. *Nature* **453**, 489–494 (2008).
8. Xu, J. *et al. Nature Struct. Mol. Biol.* **24**, 40–46 (2017).
9. San Filippo, J., Sung, P. & Klein, H. *Annu. Rev. Biochem.* **77**, 229–257 (2008).
10. Heyer, W.-D. *Cold Spring Harb. Perspect. Biol.* **7**, a016501 (2015).

This article was published online on 14 October 2020.

Catalysis

Titanium atoms pair up in industrial catalyst

Bert M. Weckhuysen

A study of the industrial catalyst titanium silicalite-1 suggests that the conventional view of the structure of its active sites is wrong. The findings might enable further optimization of related industrial catalysts. **See p.708**

Metal ions trapped in crystalline microporous solids known as zeolites are promising solid-state catalysts for a wide variety of oxidation reactions[1–3]. In the past few decades, there has been intense research into the structure and nuclearity (the number of metal ions) of the active sites in these zeolite-based catalysts[4,5]. Despite these efforts, there is still no agreement on the nuclearity: the proposed number of metal ions in the active sites ranges from one to three[1,2,4,5]. On page 708, Gordon *et al.*[6] propose that there are two titanium ions in the active sites of a well-characterized industrial zeolite catalyst called titanium silicalite-1 (TS-1), challenging the widely accepted idea that there is only one. Their work has implications not only for TS-1, but also for other metal-containing zeolites for which the structure of the active sites is not yet fully established.

TS-1 has a rich scientific and industrial history[3,7]. It started with the seminal work[3] of industrial researchers in the early 1980s, who prepared it by partially replacing silicon atoms with titanium atoms in a zeolite that has a particular type of porous structure (the MFI structure). They found that TS-1 catalyses several oxidation reactions, most notably the epoxidation of propylene ($H_3CCH{=}CH_2$) – a reaction in which an oxygen atom in hydrogen peroxide (H_2O_2) is added to propylene's carbon–carbon double bond (Fig. 1a). The product of this reaction is propylene oxide, a compound widely used to manufacture the building blocks of polyurethane plastics. This initial work spurred further industrial interest, and led to the use of TS-1 as a catalyst for the