Cell Systems
Commentary

Hypothesis: Homologous Recombination Depends on Parallel Search

Johan Elf[1,*]
[1]Department Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden
*Correspondence: johan.elf@icm.uu.se
http://dx.doi.org/10.1016/j.cels.2016.10.005

It is not known how a cell manages to find a specific DNA sequence sufficiently fast to repair a broken chromosome through homologous recombination. I propose a solution based on freely diffusing molecules that are programmed with sequences corresponding to those flanking the break site. In such a process, the target search would be parallelized.

Having a double strand break (DSB) of chromosomal DNA is a serious condition for any cell. In all kingdoms of life, cells achieve error-free chromosome repair after a DSB by using homologous recombination (HR), where a reference sequence in a sister chromatid or a homologous chromosome is used as a template. The molecular mechanism of HR has been worked out in great detail (Heyer et al., 2010) since HR-deficiency is a major source of cancer (Moynahan and Jasin, 2010). However, the mystery of how a broken sequence can *find* its homologous template sufficiently fast is unsolved (Renkawitz et al., 2014), although the importance of getting the homologous sequences together was pointed out already by Holliday more than 50 years ago (Holliday, 1964).

The mystery is that HR seems to work also in cases when there is no newly replicated sister chromatid in close proximity to the break site. The efficiency is lower, but cells still manages to repair DBS breaks when they do not know where to find the homologous template sequence. This is best demonstrated in yeast where HR works also when short homologous sequences are transplanted between different chromosomes (Lee et al., 2016). This suggests that that HR probes each conceivable chromosome position although this requires unwinding of the double helix and base pairing with the broken chromosome sequence. However, even considering recent in vitro measurements revealing rapid probing for DNA homology with short oligos (Qi et al., 2015), it would take weeks to unwind and compare all sequences even in the relatively small yeast genome. Adding the diffusive and topological constraints of moving and aligning the broken chromosome in a dense nucleus, the timescales involved become astronomical, especially considering that the broken ends must also be held together (Lisby and Rothstein, 2004b). Still the yeast cell manages to find the homologous sequence in hours (Barzel and Kupiec, 2008).

Also in *E. coli* it is well established that HR works for spatially distant sequences. Here, it has been possible to use direct imaging experiments where the DSB site and target sequence are fluorescently labeled such that their movement can be followed in time (Lesterlin et al., 2014). The sequences are brought together in less than 5 min, which is as fast as an individual *lac* repressor transcription factor, LacI, can search the *E. coli* chromosome for a target (Hammar et al., 2012). However, in contrast to the HR search process, LacI does not have to unwind the DNA for base pairing or drag the whole broken chromosome around.

Something fundamental is clearly missing in our understanding of the search phase of homologous recombination. This is, however, not only an irritating gap in our biophysical understanding of basic cell biology, but a severe bottleneck for developing efficient CRISPR-Cas9 mediated genome engineering which currently is limited by the efficiency of HR.

A Possible Solution to the Search Problem

If it takes too long to solve a problem sequentially, the task of finding the solution can often be parallelized, that is the problem is broken into pieces that can be solved independently and at the same time. Thus, I suggest that the cell makes many short copies of the sequences flanking the DBS and uses these DNA or RNA "scouts" to search for the homologous sequence (Figure 1). This strategy not only parallelizes the search process but also eliminates the need to move the broken chromosome around in a time consuming manner. The situation is similar to the difference between a single person walking from house to house searching for a missing person (analogous to the Classic model) and a televised broadcast of the missing person's photo (analogous to my hypothesis). Although molecular components that could be used to implement a broadcast strategy in an intracellular setting already have been described, the transformative impact of this strategy on search speed has not been discussed.

Cellular Implementation

To exemplify how the parallelized search could be implemented with already characterized components, I will give one prokaryotic and one eukaryotic example (Figure 2). In *Arabidopsis* and in human cells, it has been reported that small RNAs are transcribed from the sequences flanking the DBS (Wei et al., 2012). These so called diRNA bind to Ago2, which from separate studies, e.g., Janowski et al. (2006), are seen targeting chromosomal promoters or nascent transcripts when programmed by complementary microRNA. Thus, the Ago2-diRNA complexes are good candidates for being the scouts that search the genome for sequences homologous to that of the broken chromosome. Ago2 deletions are further known to be deficient in HR and

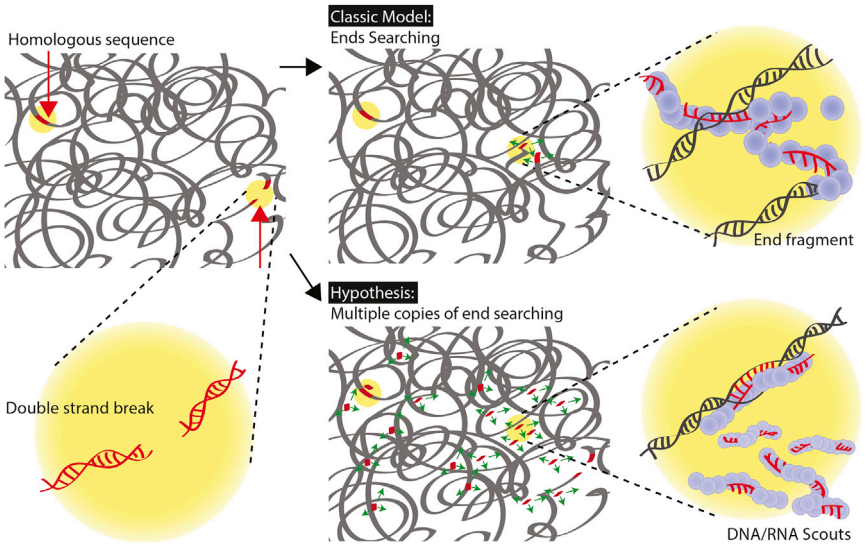

Figure 1. Before Homologous Recombination Can Take Place, the DNA Ends at the Double Strand Break Site Need to Find the Homologous Sequences
In classical models the ends search the chromosomes one position at the time, i.e., sequentially, by unwinding dsDNA and base pairing. As an alternative hypothesis, I propose that the sequence flanking the ends are copied such that these copies can be used in a parallel search.

Ago2 is directly binding to the critical HR protein Rad51 even in the absence of the diRNA[13]. It is unclear if Ago2 can search and bind chromosomal DNA or if it targets nascent RNA transcripts. In either case, the important aspect is that Ago2 could recruit Rad51 or related HR proteins to mark the homologous region.

An RNA corresponding to the diRNA has not been described in bacteria, but the many short ssDNA resection products that are made by the RecBCD or AddAB systems while processing the DSB ends (Wigley 2013) could play the same role. This would also explain why thousands of base pairs are degraded at the DSB ends and why the ssDNA degradation products have lengths of tens of nucleotides instead of single nucleotides (Dillingham and Kowalczykowski 2008), which is what is needed to identify a chromosomal position uniquely. In fact, short pieces of ssDNA have been shown to bind RecA in vitro, and this complex is also known to search for specific sequences in dsDNA (Ragunathan et al., 2011). It is not known whether the resection products of RecBCD and AddAB bind RecA in vivo. Notably, however, the resection products can bind Cas1-Cas2 (Levy et al., 2015), as part of the CRISPR spacer acquisition process. This suggests that the resection products are also accessible to RecA. If they are, they would acts as scouts that recruit RecA to the region homologous to the break site and mark it as such.

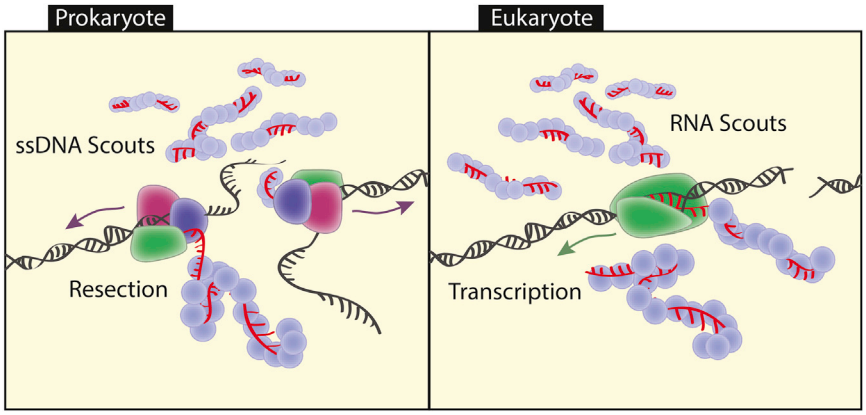

Figure 2. Making Copies of Broken Ends
Short single stranded sequences previously flanking the DSB are made in the resection process (left) or by transcribing the sequences flanking the DSB into short RNAs (right).

Bringing the Sequences Together

Once the homologous template sequence has been located and marked, there are many ways to make the region more accessible for recombination with the broken chromosome (Figure 3). This second step may be passive in the sense that broken ends simply have a higher affinity for the marked positions such that they bias their random search to the region where the template sequences are located. It does, however, seem more likely that the process of bringing the DBS and the homologous sequence together is an active process based on the ATPase activity of RecA and Rad51. The best evidence for active transport is the observation of directed movement by DSB channeled along RecA filaments toward the template sequence in *E. coli* (Lesterlin et al., 2014). One can also envision relocation of any RecA/Rad51 bound DNA to nuclear "repair centers" (Lisby and Rothstein 2004a), which would place the DSB and the template sequence in the same region.

The observation that RecA filaments often stretch from the DSB to its homologous sequence in another part of the cell (Lesterlin et al., 2014; Kidane and Graumann 2005) is very suggestive since it implies that the filament has information about which sequence to reach for. Given our current understanding of RecA structure and dynamics this information is unlikely to be communicated through the filament, which implies that the target sequence needs to be marked by other means, for example by the scouts. If the actual search by the scouts occurs before moving the filament, this scenario would agrees well with the timescales for HR after DSB in *E. coli*: it has been observed that broken ends do not move much the first hour, but then they move in a directed manner to the target

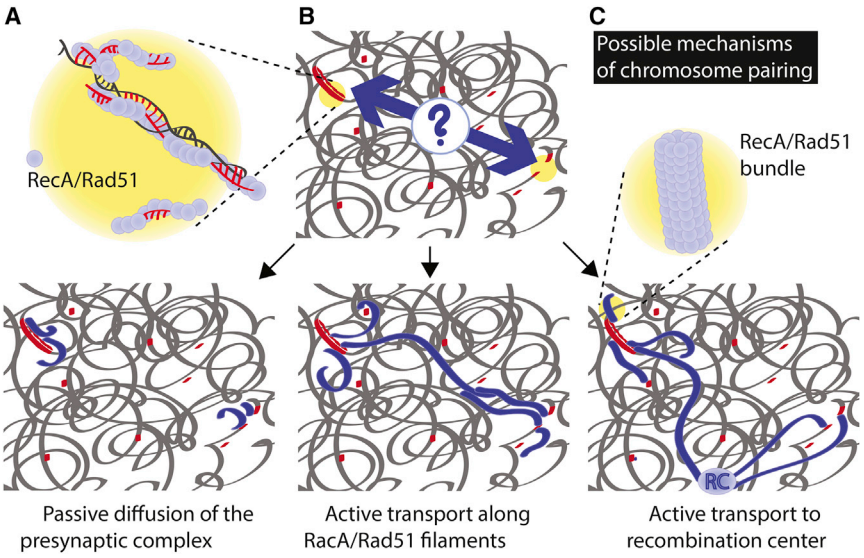

Figure 3. Bringing the Homologous Sequence to the Template
When the homologous region has been found and marked by the scouts, the DNA flanking the DSB can for example find the correct site through (A) a random walk without unwinding all the chromosomal DNA, (B) by forming a filament between the homologous sequence and the DSB that can actively pull the regions together or (C) by filaments dragging the marked regions to a common recombination center.

sequences within 5 min (Lesterlin et al., 2014).

Independent of how the DSB and the homologs site are brought together; when the template sequence is already found and marked through parallel search by many freely diffusing molecules, the overall process of homologous recombination can be made many thousands times faster.

In summary, I propose that the solution to the homologous recombination search problem is a parallelized search implemented by freely diffusing molecules programmed with sequences corresponding to those flanking the break site.

ACKNOWLEDGMENTS

I am grateful to Ran Kafri for introducing me to the existence diRNA, to Claudia Kutter for helpful discussion, to Irmeli Barkefors for drawing the figures, and to David Fange and Måns Ehrenberg for help with the text.

REFERENCES

Barzel, A., and Kupiec, M. (2008). Nat. Rev. Genet. *9*, 27–37.

Dillingham, M.S., and Kowalczykowski, S.C. (2008). Microbiol. Mol. Biol. Rev. *72*, 642–671.

Hammar, P., Leroy, P., Mahmutovic, A., Marklund, E.G., Berg, O.G., and Elf, J. (2012). Science *336*, 1595–1598.

Heyer, W.-D., Ehmsen, K.T., and Liu, J. (2010). Annu. Rev. Genet. *44*, 113–139.

Holliday, R. (1964). Genet. Res. *5*, 284–304.

Janowski, B.A., Huffman, K.E., Schwartz, J.C., Ram, R., Nordsell, R., Shames, D.S., Minna, J.D., and Corey, D.R. (2006). Nat. Struct. Mol. Biol. *13*, 787–792.

Kidane, D., and Graumann, P.L. (2005). J. Cell Biol. *170*, 357–366.

Lee, C.-S., Wang, R.W., Chang, H.H., Capurso, D., Segal, M.R., and Haber, J.E. (2016). Proc. Natl. Acad. Sci. USA *113*, E146–E154.

Lesterlin, C., Ball, G., Schermelleh, L., and Sherratt, D.J. (2014). Nature *506*, 249–253.

Levy, A., Goren, M.G., Yosef, I., Auster, O., Manor, M., Amitai, G., Edgar, R., Qimron, U., and Sorek, R. (2015). Nature *520*, 505–510.

Lisby, M., and Rothstein, R. (2004a). Curr. Opin. Cell Biol. *16*, 328–334.

Lisby, M., and Rothstein, R. (2004b). Curr. Biol. *14*, R994–R996.

Moynahan, M.E., and Jasin, M. (2010). Nat. Rev. Mol. Cell Biol. *11*, 196–207.

Qi, Z., Redding, S., Lee, J.Y., Gibb, B., Kwon, Y., Niu, H., Gaines, W.A., Sung, P., and Greene, E.C. (2015). Cell *160*, 856–869.

Ragunathan, K., Joo, C., and Ha, T. (2011). Structure *19*, 1064–1073.

Renkawitz, J., Lademann, C.A., and Jentsch, S. (2014). Nat. Rev. Mol. Cell Biol. *15*, 369–383.

Wei, W., Ba, Z., Gao, M., Wu, Y., Ma, Y., Amiard, S., White, C.I., Rendtlew Danielsen, J.M., Yang, Y.G., and Qi, Y. (2012). Cell *149*, 101–112.

Wigley, D.B. (2013). Nat. Rev. Microbiol. *11*, 9–13.