Achieving Fidelity in Homologous Recombination Despite Extreme Complexity: Informed Decisions by Molecular Profiling

Robert P. Rambo,[1] Gareth J. Williams,[1] and John A. Tainer[1,2,*]
[1]Life Sciences Division, Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[2]Department of Molecular Biology, The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA
*Correspondence: jat@scripps.edu
DOI 10.1016/j.molcel.2010.10.032

In this issue of *Molecular Cell*, Savir and Tlusty (2010) apply signal detection theory to show that homologous recombination machinery is optimally tuned to find homologous DNA sequences within an exceptionally high background of heterologous sequences.

Homologous recombination (HR) is a biologically critical, conserved, genomic maintenance system found in all three domains of life that is essential for genome integrity. An essential step in HR is the direct search of the homologous sequence within an extremely high background of heterologous DNA sequences. Although the proteins comprising the HR machine vary among species, HR search effectors include the filament forming RecA/Rad51 protein superfamily (RecA in bacteria, RadA in archaea, and Rad51 and DMC1 in eukarya). HR is a very complex process that occurs through formation of the RecA nucleoprotein filament (NPF) on a select single-stranded DNA (ssDNA). NPF formation determines the DNA sequence for recombination. Within the context of the human genome (46 chromosomes at ~6 billion base pairs), finding a homologous DNA sequence becomes a challenging search problem akin to "finding a needle in a haystack," when the needle is just another piece of hay! Crystal structures of RecA bound to ssDNA revealed that the NPF stabilizes the ssDNA in a distorted conformation (Figures 1A and 1B), splaying out bound nucleotides as short trinucleotide repeats with B-DNA geometry (Chen et al., 2008). This architecture facilitates recognition of homologous regions through Watson and Crick base pairing; however, actual searching requires probing of double-stranded DNA (dsDNA). NPF short trinucleotide segments suggest that testing or profiling of homologous DNA segments is incremental (Sagi et al., 2006). Presumably, when sufficient bound segments base pair (Figure 1C) (Kowalczykowski, 2008), HR machinery must decide to switch from *search* mode to *found* thereby committing DNA to the next stage, strand exchange. Likewise, if the segment is not homologous, HR machinery must decide to continue searching. Failure to make the proper decision can lead to nonhomologous recombination and loss of genetic information.

Savir and Tlusty (2010) reconceptualized the homologous recombination framework using signal detection theory. The RecA mediated homologous recombination search problem is essentially the recognition of a signal (complementary DNA sequence) within a naturally high background (heterologous DNA sequences). However, the exceptionally low signal-to-noise ratio creates uncertainty in recognizing the signal, which implies decisions come at a cost. Consider a repeated coin toss game, heads we make a positive decision (stand inside a hot house) or tails we make a negative decision (stand outside in the cold) (Figure 1D). Failure to make the proper decision i.e., *standing inside on tails*, results in a lost article of clothing. Success would be easy, if we could see the coin, as heads would always imply the positive decision and tails would always imply the negative decision. But in the HR version of this game, we must feel the upside of the coin, such that the implications are weakly defined. With no tactile experience the first time the game is played, our reasoning is insufficient in guiding a choice (Jaynes and Bretthorst, 2003). Yet, with increasing trial decisions, experience helps define the cost of making a decision.

Under signal detection theory, Savir and Tlusty (2010) proposed that recognition of homologous DNA by HR machinery is essentially a cost-optimization problem. The positive decision to commit the HR machinery to a recombination event is a cost balance of NPF correctly recognizing a homologous sequence that preserves genetic information versus the heterologous sequence that generates new genetic information. Through the experience of evolutionary selection, HR has been optimally tuned to emphasize homologous DNA while minimizing heterologous DNA. Savir and Tlusty (2010) suggest a mechanism for cost optimization by noting that the ssDNA in the NPF is extended, thus requiring a large-scale geometric rearrangement of the DNA found within the targeted dsDNA sequence. Consequently, the total free energy gain by pairing the NPF trinucleotide unit with a test sequence will be the sum of the binding free energy (pairing) and the extension free energy (deformation). This suggests HR machinery not only examines the target sequence at the Watson and Crick base-pairing level but also at the conformational level. Can the target sequence deform similarly to bound NPF sequence? Savir and Tlusty (2010) provide a thermodynamic argument that these events, which can be performed without ATP hydrolysis by utilizing natural dsDNA breathing, provide improved detection through a mechanism called *conformational proofreading*.

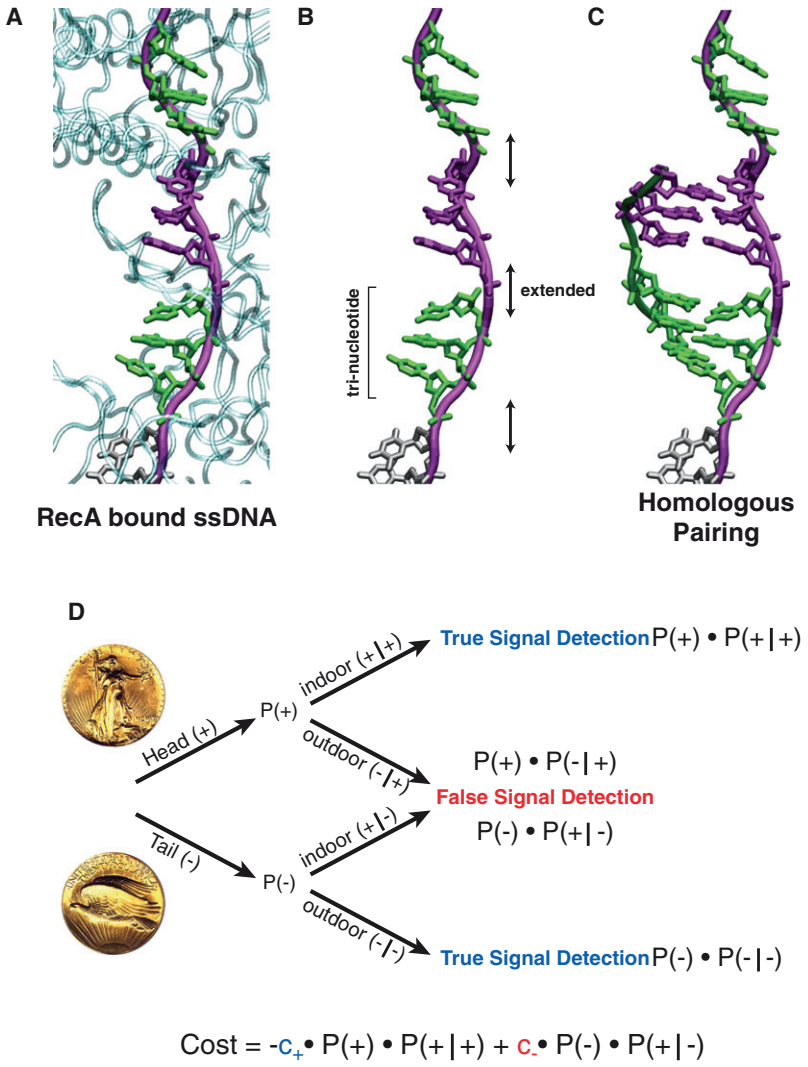

Figure 1. HR-Directed Search for the Homologous Sequence, Conformational Proofreading, Cost Optimization, and Signal Detection Theory

(A) Structure of RecA (light blue) bound ssDNA complex. Alternating trinucleotide repeats (green or purple) illustrate B-DNA like geometry.

(B) Arrows demarcate regions of extended deformation along the DNA backbone between trinucleotide units.

(C) Base pairing between homologous DNA strands with deformed backbone geometry.

(D) Binary game decision tree. Possible misinterpretation of coin toss can lead to false signal detection.

Like physics before the discovery of unifying principles, biology is discovering extreme complexity of interactions governing pathway selection, signaling, regulation, progression, and biological outcomes. In the face of challenges from extreme biological complexity, we must strive to seek unifying, mechanistic principles that yield a simplifying understanding of complex data from experimental and bioinformatical approaches. In this context, application of information, signal detection, or economic theory to biological problems creates novel opportunities for discoveries that unify our understanding of complex processes, such as HR. Savir and Tlusty (2010) have thus provided us with a unifying framework for understanding key aspects of the molecular recognition problem for HR.

How generally does this analysis apply to cell biology? In fact, many DNA pairing and repair events involve dsDNA geometric rearrangement, such that decisions are made by testing total free energy gain by partner pairing compared to deformation free energy. Conformational proofreading likely occurs from related DNA processes such as DNA synthesis (Johnson and Beese, 2004) and base repair (Qi et al., 2009), to comparable systems such as promoter recognition by transcription factors (Pabo and Sauer, 1992) and DNA mimicry (Putnam and Tainer, 2005), to the reversible assembly of macromolecular machines, such as bacterial pili (Craig et al., 2006). Indeed, throughout the cell, molecular surfaces are ubiquitous. Thus, biological macromolecular machines have evolved to recognize the correct signal within a noisy background where the cost has been optimized against nonspecific recognitions that, during the course of evolution, result in cellular toxicity or mutagenesis. Looking ahead, the need to discover unifying principles that aid the understanding and prediction of biological interactions in the face of extreme complexity and low signal-to-noise remains a critical challenge for cell biology in the next decade and beyond.

REFERENCES

Chen, Z., Yang, H., and Pavletich, N.P. (2008). Nature *453*, 489–4.

Craig, L., Volkmann, N., Arvai, A.S., Pique, M.E., Yeager, M., Egelman, E.H., and Tainer, J.A. (2006). Mol. Cell *23*, 651–662.

Jaynes, E.T., and Bretthorst, G.L. (2003). Probability Theory: The Logic of Science (Cambridge: Cambridge University Press).

Johnson, S.J., and Beese, L.S. (2004). Cell *116*, 803–816.

Kowalczykowski, S.C. (2008). Nature *453*, 463–466.

Pabo, C.O., and Sauer, R.T. (1992). Annu. Rev. Biochem. *61*, 1053–1095.

Putnam, C.D., and Tainer, J.A. (2005). DNA Repair (Amst.) *4*, 1410–1420.

Qi, Y., Spong, M.C., Nam, K., Banerjee, A., Jiralerspong, S., Karplus, M., and Verdine, G.L. (2009). Nature *462*, 762–766.

Sagi, D., Tlusty, T., and Stavans, J. (2006). Nucleic Acids Res. *34*, 5021–5031.

Savir, Y., and Tlusty, T. (2010). Mol. Cell *40*, this issue, 388–396.