Error-correction coding theory of DNA repair

Bane Vasić[1], David W. Galbraith[2], Shashi Kiran Chilappagari[3], and Michael W. Marcellin[4]

1 Motivation

Maintaining integrity of genetic material is achieved through DNA repair, a process in the cell in which damage and continually monitored and corrected. In many organisms the genes and proteins that participate in this process have been identified, but with a few exceptions the role of many of them is given only descriptively. Understanding the role of genetic and protein interactions during DNA-repair and its mathematical formulation is obscured not only by the limitations of the existing experimental methods, but also by deficiencies of their underlying theoretical frameworks or lack of thereof.

In this poster we will present an information theoretical framework for of DNA repair which views it as an error correction system [1]. The main idea is that the principles of information theory and coding theory developed for engineering of man-made information storage systems [3] can serve as a basis for incorporating phenomena observed on different levels of abstraction of the genomic error correction system.

The model encompasses the multilevel structure of the genetic error correction system and interactions not only among its different levels but also among other sub-systems in the cell. In order to understand such a complex system, specialized repair mechanisms, which have been the primary object of research in the past, must be considered in the context of the global error correction machinery. The proposed framework for rigorous treatment of DNA-repair enables describing its functionality quantitatively and algorithmically. In the rest of this abstract we briefly describe the proposed information storage framework.

2 Information Storage Theory Setting

The block diagram of the genetic error correction system is shown in Figure 1. Note that the term "coding" is used in the information theory sense which has much broader meaning than in biology. This communication theoretical setting, in which the DNA replication is viewed as a process of information storage (or transmission), is not new - the existence of a genetic error correction system has been hypothesized in the literature. However with few exceptions, these models were rather naïve, and placed redundancy only in the genetic code (in the Watson-Crick sense) and correction only on a molecular level. Our model emphasizes the fact that the strength of error correction comes from the orchestrated response of a whole genome. This response is driven by an error correction algorithm.

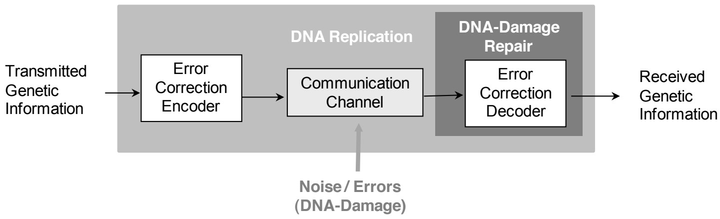

Figure 1 The existing communication theory setting of the process of DNA replication and DNA-damage repair. DNA replication is analogous to the storage of information on cion/memory channel. DNA-damage repair is analogous to error correction decoding. Encoding represents all mechanisms and structures (ranging from a molecular to a genomic level) developed by nature to make the genetic information less susceptibe to errors during replication. All molecular interactions, chemical signaling, protein activation and deactivation, transcription, and other processes involved in DNA-damage repair are viewed as steps performed in the *decoder*, with the goal of error free transmission of information from a mother to a daughter cell.

The redundancy that enables DNA-damage repair exists on many layers of molecular interactions, and manifests itself though the molecular structure of DNA, the structure of genes, the specificity of *DNA-repair proteins*, but also though gene interactions. Among many other functions in the cell, gene interactions have a

[1] Department of Electrical and Computer Engineering and Department of Mathematics. E-mail: vasic@ece.arizona.edu
[2] Department of Plant Sciences, and BIO5 Institute for Collaborative Bioresearch. E-mail: galbraith@arizona.edu
[3] Department of Electrical and Computer Engineering. E-mail: marcellin@ece.arizona.edu
[4] Department of Electrical and Computer Engineering. E-mail: shashic@ece.arizona.edu

fundamental purpose of embodying an error correction algorithm. The simplest model takes into account only interactions through transcript, i.e., each gene is regulated by a number of genes in a constellation known as a *gene regulatory network* (GRN). The information theoretical significance of this network lays in the fact that it embodies the decoder. From the engineering standpoint, the ultimate goal is to produce a formal logical and casual description of interaction among genes, or a *genetic wiring diagram.* Such wiring diagram can be viewed as a digital logic circuit of a decoder.

The main conceptual advance in our grasp of theoretical aspects of DNA-repair follows from the fact that chemical compounds and agents that threaten genetic stability damage not only the DNA molecule, but proteins, RNA, and many other molecules involved in DNA-repair. Therefore not only is the transmitted/stored information damaged, but so is the error correction decoder circuitry. The process of correcting errors is thus not error free as assumed in the classical setting shown in Figure 1.

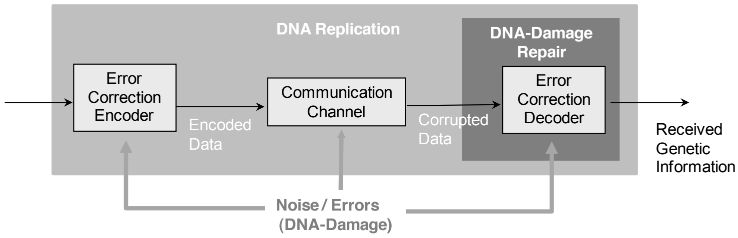

Figure 2 A block diagram of the fault-tolerant decoding for DNA-damage repair. External agents corupt the transmitted information but also damage the decoder. Even though the error correction is performed by a faulty decoder, low error probability is maintained. At *DNA-damage checkpoints*, a decision is made whether to continue replication. Reenters the cell-cycle without completing damage repair is equivalent to miscorrection. The apoptosis is equivalent to decoding failure.

Surprisingly, genetic error correction has not been considered in this way (illustrated in Figure 2). Despite its faultiness, the decoder maintains very low probability of (replication) error, as low as 10^{-10} errors, per site, per replication cycle. The ultimate goal is to find out what property of the genetic decoder makes such high fidelity of replication possible. Only a few such fault-tolerant decoder topologies are known today [4]. Their main attribute is distributed information processing typical to so called codes on graphs. The parallel nature of genetic and protein interactions suggests a highly distributed nature for the genetic error correction algorithm, and possibly a universal design principle to ensure fault-tolerance.

3 A Case Study

Our poster will focus on mathematical description and analysis of DNA-damage repair in *yeast Saccharomyces cerevisiae* induced by *ionizing radiation* (IR). Such focus circumvents deficiencies of previous studies in which on one hand, theoretical models had only loose connection with biological constraints, while on the other hand, experimental studies had too many uncontrolled parameters which made the model inference impossible. The choice of yeast as a model organism is natural in view of the amount of data available in genomic databases (see for example [2] and [5]), and its amenability to experimental manipulation. The choice of IR as a damage agent is motivated by controllability of its effects in intensity (dose), time and space, which leads to significant simplifications of the proposed model [5].

References

[1] Gallager R. 1968. I*nformation Theory and Reliable Communication*, John Wiley & Sons Inc.

[2] Pan X, Ye P, Yuan DS, Wang X, Bader JS, Boeke JD, 2006: A DNA integrity network in the yeast *Saccharomyces cerevisiae. Cell*, 124:1069-1081.

[3] Kurtas EM, Vasic B, Eds. 2005, *Advanced error control techniques for data storage systems*. New York: CRC Press

[4] Vasic B, Chilappagari SK, 2007. An information theoretical framework for analysis and design of nanoscale fault-tolerant memories based on low-density parity-check codes *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 54, No. 11. pp. 2438-2446.

[5] Westmoreland TJ, Marks JR, Olson JA Jr, Thompson EM, Resnick MA, Bennett CB. Cell cycle progression in G1 and S phases is CCR4 dependent following ionizing radiation or replication stress in *Saccharomyces cerevisiae*. 2004. Eukaryot Cell. 3(2):430-46.

[6] Workman, CT., Mak, HC., McCuine, S., Tagne, JB., Agarwal, M., Ozier, O., Begley, TJ., Samson, LD., Ideker, T. 2006. A Systems Approach to Mapping DNA Damage Response Pathways. *Science*. 312 (5776):1054-1059.