

DNA Computing and Errors

Elena Losseva

April 30, 2004

Keywords: DNA computing, reliability of computation, DNA encodings

The field of DNA computing can be described as a pursuit of finding strategies to use DNA (deoxyribonucleic acid) for performing general-purpose computations. This task has two broad components: data storage and data manipulation. My research is connected to both.

In order to store data in DNA, it is necessary to find optimum methods of encoding information in DNA. The DNA strands used for computation need to have certain characteristics in order to avoid data loss and erroneous results. One source of problems comes from DNA code words attaching (also called annealing) to each other in unexpected ways and, as a result, making them unsuitable for computation. It is therefore advantageous to have code words such that the corresponding DNA strands do not bind to each other in undesirable ways. Languages (word sets) that have this property are studied from the point of view of formal language theory. A hierarchy of related non-annealing languages emerges as a result. This research was published in [2].

Another concern in the design of DNA strands appropriate for data storage is a possibility that DNA strands may form secondary structure. This is a situation where, within a single DNA strand, different segments of the molecule anneal to each other. The problem with a molecule having secondary structure is that this may prevent the molecule from participating in inter-molecular interactions. One particular type of secondary structure is called a hairpin loop. The name arises from the DNA molecule resembling a hairpin. The study in [3] investigates the problem of how to construct DNA molecules that avoid hairpin loops.

Due to the inaccuracy of biochemical reactions, the experimental implementation of a DNA computation may lead to incorrectly calculated results. The survey of computer science techniques for managing errors in DNA-based computation is given in [1].

The inherently complex nature of biological processes makes it increasingly more apparent that most plausible implementations of bio-computing are likely to produce some unexpected and erroneous results. An old idea in computing, dating back at least as early as J. von Neumann's writings in 1956, is to be able to build reliable systems from unreliable components. Von Neumann proposes to view error "not as an extraneous and misdirected or misdirecting accident, but as an essential part of the process under consideration". These words gain

nearly axiomatic significance in the new paradigm of DNA computing. If we are to succeed in building functional and practical computational devices with DNA, they will be inevitably preceded by more simplistic and restricted mechanisms of limited reliability. Discovering how to combine the components of moderate reliability into larger systems with a desired reliability is key to the extension of computing capacity. This idea was explored in [4] by looking at the construction of reliable finite automata from component finite automata that do not correctly accept their intended languages.

References

- [1] Kari, L., Losseva, E., & Sosík, P. (2004). DNA computing and errors: a computer science perspective. Book chapter in *Molecular Computational Models: Unconventional Approaches*, Idea Group, Inc., in print.
- [2] Kari, L., Konstantinidis, S., Losseva, E., & Wozniak, G. (2003). Sticky-free and overhang-free DNA languages. *Acta Informatica*, 40, p.119 - 157.
- [3] Kari, L., Konstantinidis, S., Losseva, E., Sosík, P., & Thierrin, G. (2004). DNA involutions and hairpin structures, submitted.
- [4] Losseva, E. (2003). Building reliable systems from unreliable components and the DNA computing paradigm. *Bulletin of European Association of Theoretical Computer Science*, 81, p.305-317.