

RELATIVE WATSON-CRICK PRIMITIVITY OF WORDS

LILA KARI^(A) MANASI S. KULKARNI^(B) KALPANA MAHALINGAM^(B)

^(A) *School of Computer Science, University of Waterloo, Waterloo, ON, N2L 3G1 Canada
Department of Computer Science, University of Western Ontario
London, Ontario, N6A 5B7, Canada
lila.kari@uwaterloo.ca*

^(B) *Department of Mathematics, Indian Institute of Technology Madras
Chennai, Tamilnadu, 600036, India
ma16ipf01@smail.iitm.ac.in (M. S. KULKARNI)
kmahalingam@iitm.ac.in (K. MAHALINGAM)*

ABSTRACT

We introduce the concept of relative Watson-Crick primitivity of words and its generalization, the relative θ -primitivity of words, where θ is a morphic or an antimorphic involution. Similar to relatively prime integers which do not share any common factors, we call two words u and v relatively θ -primitive if they do not share a common θ -primitive root. We study some combinatorial properties of relatively θ -primitive words, as well as establish relations between each of the two words u and v and the result of some binary word operation between u and v , from this perspective.

Keywords: primitive words, relatively primitive words, θ -primitive words, antimorphic involution

1. Introduction

Periodicity and its opposite, primitivity of words, are fundamental properties of words in combinatorics on words and formal language theory. In addition, the detection of repetitions in strings plays an important role in, e. g., pattern matching and text compression [2, 3, 14]. On the other hand, tandem repeats in DNA, that is, sequences of two or more contiguous, approximate copies of a pattern, occur in the genomes of both eukaryotic and prokaryotic organisms, and have biological and medical significance. In this paper we bring together these concepts from two different fields with the notion of relatively prime numbers from number theory, to define the relative Watson-Crick primitivity of words.

Recall that, in the framework of formal language theory, a single DNA strand – a sequence of nucleotides that can be of four different kinds, Adenine, Cytosine, Guanine or Thymine –, can be modelled as a word w over the alphabet $\Delta = \{A, C, G, T\}$. DNA strands can be either single-stranded or double-stranded, with the latter being formed