Krajowa Konferencja Zastosowań Matematyki w Biologii i Medycynie

Sandomierz, 5th–9th September 2016

# CHAOS GAME REPRESENTATION AS A METHOD OF CLUSTERING AND DISTINGUISHING DNA BASE PAIR FREQUENCIES

**Magdalena Nowak**

Jan Kochanowki University in Kielce
ul. Świętokrzyska 15, 25-406 Kielce,
mnowak@ujk.edu.pl

## ABSTRACT

The algorithm of Chaos Game Representation (CGR) grown out of the branch of Mathematics called Fractal Theory. Fractals are objects of infinitely complex structure in a certain mathematical sense. Moreover, they are usually made of repeated pieces which have the same shape as the overall figure. It turns out that many natural objects like clouds, rivers, trees, blood vessels and bacteria colony, have fractal structure.

Many of fractal structures can be readily illustrated by using stochastic algorithm, known as the Chaos Game. Its deterministic version called Chaos Game Representation, may be used in data analysis to reveal patterns in long data strings such as DNA base pair sequences. Moreover, it allows to analyze amino acids in proteins or words in languages and represent long symbolic sequences on a two-dimensional plots conserving their statistical properties.