

CHAPTER 1

INTRODUCTION

1.1 Problem Identification and Objectives of Work

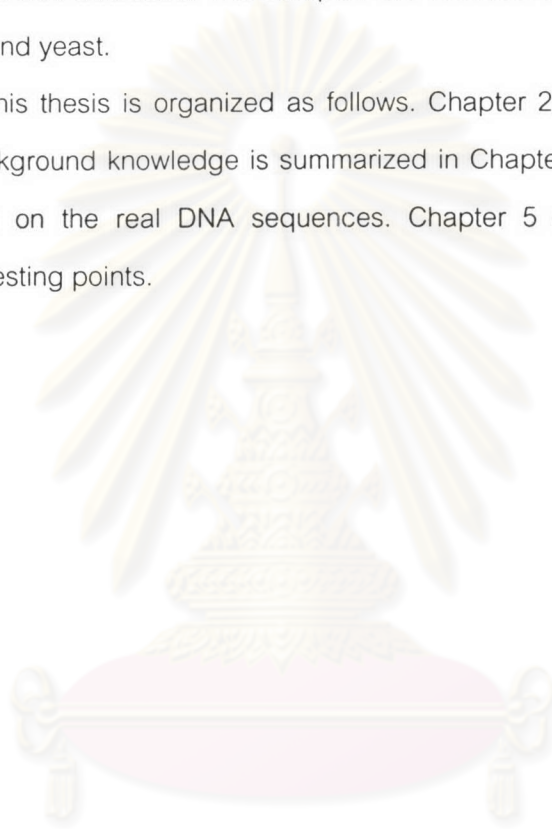
For the reason that identification of coding and non-coding regions [1] on DNA sequences takes a great amount of time and cost by infeasible biological techniques in laboratory, the computational techniques had been applied to this area in order to let the essential genetic information effortlessly and rapidly come out. Earlier, the frequency analysis was introduced by means of Fourier transform. Unfortunately, the Fourier transform has its limitation on non-stationary signal whereas a new multi-resolution frequency analysis, wavelet transform [2], is able to analyze this type of signal. In contrast to Fourier transform, which uses a single analysis window, wavelet transform uses short windows at high frequencies and long windows at low frequencies. This flexibility of the wavelet transform is the great benefit of the analysis. In this study, we propose an application of wavelet transform in an area of biology. The wavelet transform will be employed onto DNA sequences to show different regions of the DNA. The difference between coding and non-coding regions on DNA sequences is able to be distinguished noticeably during the transformation and optimization processes. The goal of this thesis is to predict coding and non-coding regions on DNA sequences.

1.2 Scope of Work

In this study we apply a computational technique to an analysis of DNA sequences. This technique is wavelet transform, one of frequency analysis. The study begins with

understanding the structure of DNA sequences, followed by understanding how the wavelet transform works, and finally we apply the wavelet transform into the DNA sequences in order to predict locations of exons and introns on DNA sequences and express both advantages and disadvantages of this method. The samples of DNA used in this work are obtained from GENBANK database. The samples are focused on Prokaryotes, especially DNA of small worm and yeast.

The rest of this thesis is organized as follows. Chapter 2 reviews the literatures of previous works. Background knowledge is summarized in Chapter 3. Chapter 4 shows our experimental results on the real DNA sequences. Chapter 5 concludes the work and suggests some interesting points.



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย