CHAPTER 5

CONCLUSIONS

During the experimentation, peaks represented by frequency response are made of discrimination between exons and introns. The exon is biologically random. At this point, we use random number generator to distinguish the different between exons and introns. The random number series directly affect to the results. We found that different random number series give different result. With different random number series, some peak is lower; some peak is shifted while some peak is not affected. The good random number set will provide the good result. As the result, the optimization process is functioning to give the best coefficients maximizing the discrimination between exon and numerical random series. Sometimes we found that result from this process does not meet criteria or out of constraints. Frequently, we found that this optimization takes very long time to give the answer, or the result is not converged. Two varying inputs of the process are a set of random number series and a set of initial values, which is manually defined. If the set of initial value is well defined, the optimization process will not be in trouble. In most cases, we found that the optimization result does not meet constraints and takes a vast period.

The wavelet analysis does not show the main great advantage about frequency-resolution analysis here because we need to use the benefit of characteristic of the data only at the $a=3$ (for wavelet transform) or at frequency $f=N/3$ (for Fourier transform). The complex Morlet wavelet emphasizes each nucleotide with different weight by using exponential window function. It treats the nucleotides in middle of window the most important. This function also provides a smooth and continuous window. This makes result smoother than the result by using Fourier transform. The wavelet analysis yield another advantage which is the mother wavelet function can be changed to express other

characteristic of DNA. Complex Morlet gives a frequency response Complex Gaussian is an interesting mother wavelet for the reason that it can blind linear trend of input information [6].

Using the wavelet analysis is possible to foretell location of exons and introns in a DNA sequence. The information sources about DNA obtained in this thesis are downloaded from GENBANK. The main result represented here is calculated by using DNA of the *Caenorhabditis elegans*, which is a kind of small worms. The outcome produced by this DNA presents better visual result than other genes. Nevertheless other genes also give the results by using this method. We expect that each genome has got its definite characteristic. It would be proper if we try varying some parameter or adjusting some procedure during the experiment.

Because the many techniques have developed to do this task, we should consider and elaborately use them together to ease the job. No technique has its disadvantage and no technique has its advantage.

This predictor is able to roughly forecast where the exons are possible. The exact location must be obtained by biological experiment. The computational techniques solely guide the user about coarse information about that gene. The user should consider what techniques should be used, depending on the resolution of the result and occupation time for the program.

Eventually, there are more effective parameters during the experiment, such as mapping algorithm, optimization technique, etc. These parameters should be profoundly studied in the future. The mapping nucleotide sequence into numerical sequence is the most important factor in representation of hidden information on DNA. If the mapping procedure is good enough, we can hopefully distinguish between coding regions and non-coding regions at once. The optimization procedure is also essential to the result for the reason that discrimination between exon and intron is distinguished by optimizing the difference of their characteristics.