Chapter 6
Why Are Multiple Regulators Required for Transcription of Each Gene?

Hiroshi Kobayashi
Chiba University, Japan

ABSTRACT

The human genome has been proposed to contain numerous regions called “consensus sequences” that are recognized by DNA-binding proteins required for gene expression. Results obtained with a computer simulation showed that: (1) all random sequences consisting of less than 15 base pairs were present in the human genome, and (2) consensus sequences reported to date were found in an artificial genome created randomly by a computer. These results suggest that conserved sequences consisting of more than 15 base pairs are required for accurate gene expression. No consensus sequence consisting of more than 15 base pairs has been identified to date. Thus, the co-association of multiple proteins bound side by side is required for appropriate gene expression because the total number of conserved sequences can exceed 15, whereas the binding of a single protein has no physiological role because its consensus sequence is present randomly in the human genome.

DOI: 10.4018/978-1-5225-0353-8.ch006

Copyright ©2016, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
INTRODUCTION

Reports of various regulatory networks for gene expression are now accumulating in mammals, demonstrating that branched complex networks are present. It has been suggested that mammals have a large number of DNA-binding proteins for gene expression. In contrast, one regulatory protein is generally involved in individual gene expression, and such expression is often mediated by only RNA polymerase without other DNA-binding proteins in bacteria. The basic question is why multiple DNA-binding proteins are required for the regulation of single gene expression in mammals. One explanation may be that mammals have more complex functions than bacteria. However, the functions operating in a single mammalian cell are less complex than those in a single bacterial cell, although the functions in the whole body are highly complex in total. It is indispensable for our improved understanding of gene-regulating networks to clarify why multiple regulators are required for each gene expression in mammalian cells.

The human genome has been almost completely analyzed (International Human Genome Sequencing Consortium, 2004), and numerous mechanisms for transcriptional regulation have been proposed (Abnizova & Gilks, 2006; Barrett, Fletcher, & Wilton, 2012). In the first step of transcription, various regulatory proteins are bound to their specific sites with a nucleotide sequence called a “consensus sequence”. Experimental data have demonstrated that a given regulatory protein recognizes its consensus sequence to regulate transcription. The bioinformatics of mammalian genomes suggest numerous sites for the binding of regulatory proteins which cooperatively regulate gene expression (Abnizova, Subhankulova, & Gilks, 2007; Barrett, Fletcher, & Wilton, 2012).

The human genome has approximately 3 billion base pairs, and many consensus sequences reported to date have less than 16 base pairs (Barrett, Fletcher, & Wilton, 2012). Theoretically, a sequence consisting of less than 16 base pairs is present randomly in the human genome if its sequence is close to random. In contrast, if the human genome is not random, consensus sequences will be conserved only at sites required for gene expression. It is therefore indispensable to clarify whether or not the human genome is close to random.

To clarify this, the following were theoretically assessed using a computer in the present study: (1) whether or not sequences consisting of 11 to 16 base pairs created randomly are present in the human genome; (2) whether or not consensus sequences reported to date can be found in an artificial genome created by random numbers using a computer. The results suggested that the human genome has areas whose sequences are close to random and the areas account for more than 2.7x10^8 (4^14) base pairs.
Analysis of Valuable Techniques and Algorithms Used in Automated Skin Lesion Recognition Systems
www.igi-global.com/chapter/analysis-of-valuable-techniques-and-algorithms-used-in-automated-skin-lesion-recognition-systems/159764?camid=4v1a

Georgian Experience in Telecytology
www.igi-global.com/chapter/georgian-experience-in-telecytology/151946?camid=4v1a