

Yanda LI

Life and information

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2010

Abstract A computer virus is merely a small piece of program in nature, and similar to that of a computer virus, an organism may be considered as an information system in nature. This paper analyzes above idea from different ways. 1) DNA sequence satisfies the basic requirements of an information system; 2) The controls of a man and a robot both obey the principle of cybernetics; 3) How a man can have ideas but a robot has no such capacity; 4) The advantages of understanding a living organism from the point of view of information systems.

Keywords organism, information system, life

1 Elicitation from a computer virus

Every day, we come into contact with many objects, and we recognize that each object has a shape and weight. We burn coal to operate machinery, and we are affected by heat and force. Integrating these sensations, we are able to generate the concepts of matter and energy. By understanding matter and energy, we, in turn, recognize more about the world.

The concept of information differs from either matter or energy. After we begin to understand the concept of information, we also begin to find that we can understand the world in profound depth if we analyze it from the view of information science. This is because information science provides us with the means to understand the inside of an object. Viewed from this perspective, we can, therefore, analyze organisms from the view of information science.

One useful analogy is the computer virus. Most people who have used a computer are quite familiar with

the computer virus that infects their computers by way of the internet. Once a computer virus infects one computer, it can spread and even duplicate to infect other computers. Eventually, infected computers will develop problems or cease to work. In this way, the performance of a computer virus is similar to that of a real virus.

The world in which the computer virus resides is cyberspace, a world formed by numerous computers, telephones, and other electronic devices, all networked to each other through telephone lines, airwaves, or power lines. Although the shape of various devices is different, the behavior of each device is determined simply by its software — a piece of program inside. In actuality, a computer virus is also merely a small piece of program in nature.

Similar to cyberspace, the real biological world is also formed by a great network of animals, plants and microbes connected to each other through air, water, etc. If we can view the behavior of a device as being determined by a small piece of information code, the obvious question then becomes whether the behavior of each organism in the biological network is also determined by a piece of internal code. If this is so, then an organism that can live independently needs not only that one piece of code but also a complete information system. The next question then becomes whether an organism may in fact have such an information system residing within it.

2 Information system inside

The most wonderful phenomenon in the world is the appearance of living creatures. There are over a million animal species, 300000 species of plants, 100000 types of microbes and over 4 million insect species in the world. Each organism has its own fine structure and special behavior. Through cell division and development, an embryo becomes a living thing. As far as we know, except for several RNA viruses, the implementation of reproduction, development, metabolism, evolution, and adaptation to surroundings are all controlled by the DNA

Received February 8, 2010; accepted March 6, 2010

Yanda LI (✉)

Department of Automation, Tsinghua University, Beijing 100084, China

E-mail: daulyd@tsinghua.edu.cn

inside the organism. Therefore, DNA can be seen as a biological information system that stores all of the genetic information of the organism.

However, although DNA only contains 4 characters (A, T, C, and G), it is comprised of millions, even billions, of bits in length. An information system must contain a memory that stores up all of the data and the relational operations. In addition, in order to visit every unit of the memory, it requires a fit address space that can cover all addresses in the memory. If we view DNA as an information system, we also need to confirm that it satisfies these basic requirements.

For example, consider the situation in the human genome. There are different genes in DNA; some genes are called “coding genes”, and their functions are to code for proteins. These genes can be seen as the memory that stores all of the protein data. The functions of other genes are to regulate the expression of specific genes, including when and to what extent expression will occur. These genes are called “regulatory genes” or “operation genes” and can be seen as “operators” that enhance, silence the expression, or degrade unnecessary mRNA of other genes. The entire genome can also be divided into two areas — coding areas and non-coding areas. The data for proteins and other elements are stored in the coding area, whereas the regulatory factors of the operators are mainly stored in non-coding and coding areas. In humans, there are 25000 genes in the genome. In order to visit every gene, we need an address space that can cover 25000 addresses. The way this is done in the human genome is as follows:

The address of each gene is located in an untranslated region (UTR) of a gene, and a binding site within the UTR can turn on the gene when proper regulatory factors bind to it. In fact, the regulatory factor needs only to match a 6–7 nt seed piece within the binding site to anchor. The number of codes for 6–7 bits with characters A, T, C, G is 4096–16384.

In addition, several histone modifications (methylation, acetylation, phosphorylation, etc.) are needed to loosen the winding duplex DNA to allow the regulatory factor to anchor. If three kinds of histone modifications are considered, then the number of addresses will be 32768–131072 to cover the 25000 addresses. Therefore, the histone modifications plus the 6–7 nt seed form the address code of a gene. Thus, a genome actually becomes a set of biological programs — or “a set of information codes”. Its basic elements are, therefore, quite similar to computer software, although it is much more complicated than common software.

If we can say a computer virus is a piece of information code in nature, then an organism is also a piece of information code in nature, albeit more complicated. From the point of view of an information system, development of an organism from an embryo can be seen to start a

program that turns on different groups of genes in different locations in an orderly fashion to build up various tissues such as skeleton, muscles, arteries, nerves, etc. At the same time, the program also instructs apoptosis after a number of cycles of cell division.

We can thus assume that an information system for life consists of three parts:

- 1) A memory of genetic information;
- 2) A development program;
- 3) A program for exchange information with the environment. The memory and development programs both join in the exchange of outside information.

Thus, organisms not only hold genetic information of themselves but also exchange information with their external environment. If any variation occurs during reproduction, then the variation can be held or eliminated through inheritance in order to adapt the organism to the environment — the so-called “evolution”.

3 Role of cybernetics in control of a man and a robot

The control of a robot is well known to obey cybernetics or control theory. A human being is formed by cells, tissues, organs, etc., which are different from parts of a machine. For a body, in order to move, to eat, and to hunt, coordination is needed between all cells, tissues, organs, and biological factors inside the cells. All parts communicate with each other and are modulated by many regulatory factors. It is unclear whether this type of regulation could obey cybernetics or control theory.

Cybernetics is a discipline that analyzes the interaction among different parts in a system and the system behavior generated by this interaction. If we note that the subhead of cybernetics is “The control and communication in the animal and the machine”, Dr. Wiener, the author of cybernetics, seemed to believe that a man and a robot both obey the same rules. Wiener has analyzed the behavior of a man caused by damage of cerebellum and found that it can be repeated in a machine. However, this does not establish whether the regulating process inside a cell or among cells can be explained by cybernetics or control theory.

After studying the control of lactose content inside the bacterium, biologists found that this control can be explained by the “lactose operon model”, as shown in Fig. 1.

The regulation process is shown in Fig. 2.

Obviously, Fig. 2 is a control chart with negative feedback control, from the view of control theory. Therefore, we can conclude that both the coordination of a body and the regulation among biological molecules obey the

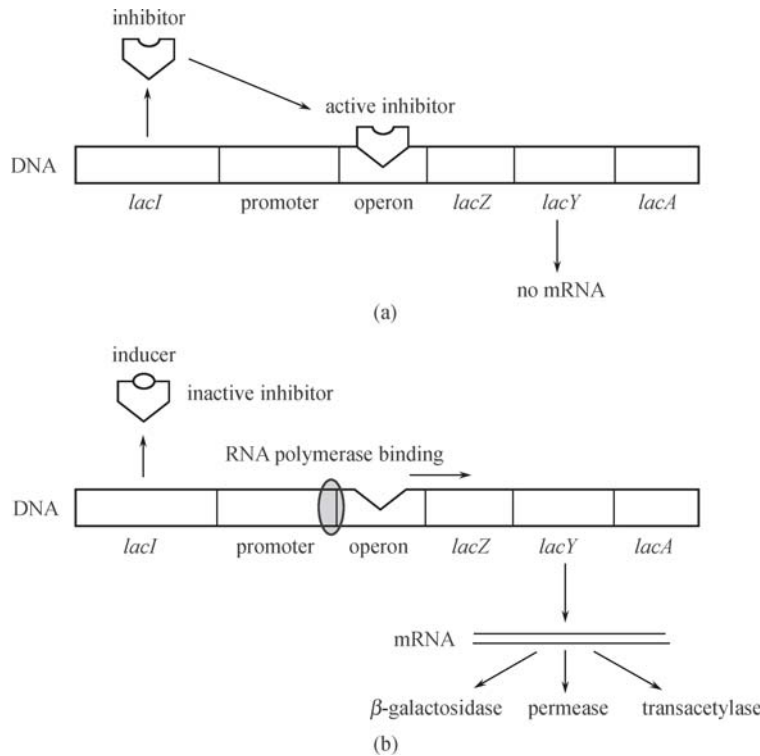


Fig. 1 Lactose operon model. (a) Lactose absent, operon off; (b) lactose present, operon on

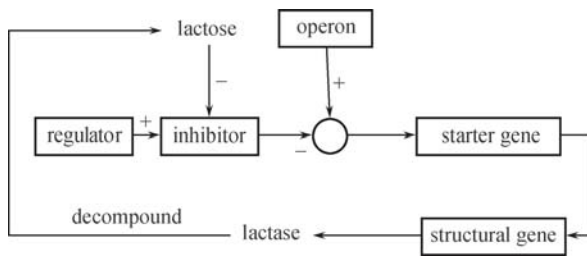


Fig. 2 Regulation process

principle of cybernetics or control theory.

The core concept in cybernetics is “feedback”. “Feedback” means that output information affects input; that is to say that the information flow inside the system will determine the behavior of the control system. In other words, the information system inside the body will determine the behavior of the body.

4 A man can have an idea but a robot cannot

Although a man and a robot both obey the principle of cybernetics, and their behaviors are determined by the information system inside, the question arises as to how a man can have ideas but a robot has no such capacity. I believe that there are two major differences between a man and a robot, as follows:

- 1) Property of a complex system;
- 2) An organism accumulates experience through evolution.

The concept of a complex system has attracted much attention in recent years. A new feature appears in a complex system when numerous elements inside the system interact on each other. This new feature is different from the features of its elements and yet cannot be reasoned from the elements. Sometimes, people use the term “emergence” to describe this phenomenon. For example, a man would generate a concept “horse” while he sees many different horses. Perhaps this is the process of “emergence”.

Much evidence has shown that a living organism is a complex system that has an “emergence” feature. For instance, analysis of heartbeat, cardiograms, electroencephalograms, the cycle of breath, dormancy, or procreation have shown that they all act as complex systems, even though at the molecular level, such as at the genome level, it still can be seen as a complex sequence. We have analyzed single base density displacement of the *Escherichia coli* (*E. coli*) genome and estimated the *H* parameter of the sequence using four methods. The values of the *H* parameter from the four methods are larger than 0.5 (shown in Fig. 3); this indicates that the DNA sequence is a complex sequence. We have also analyzed the genome sequence of several species; all of the results show that they are complex sequences.

Besides the “emergence” of complex systems, a living organism can accumulate experience through evolution. A living organism has to learn from experience in order to adapt itself to its environment. The change in the organism has been recorded in the genome and is

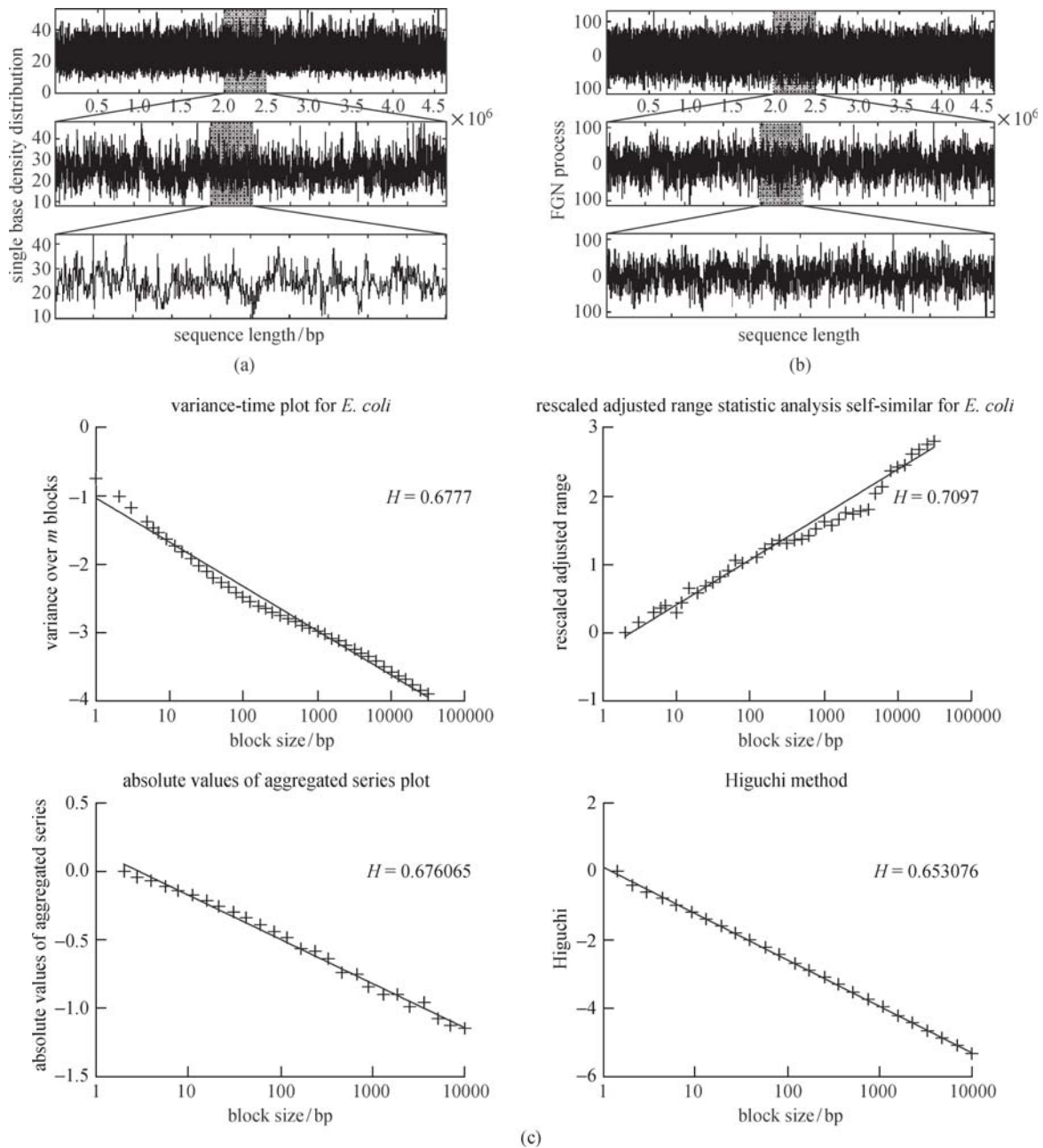


Fig. 3 H parameter estimated by four methods (revised from Ref. [1]). (a) 100 bp density distribution of A from *E. coli*; (b) fractional Gaussian noise (FGN) process with the same length and a H parameter of 0.7; (c) linear fitting for H parameter estimation of A from *E. coli*

heritable. Thus, the evolution of the organism is not only the change of structure but also the accumulation of experience and learning. This accumulation has a great impact on the development of intelligence.

Because of lack of “emergence” of its complex system and the “evolution” of experience, a robot has no consciousness and higher intelligence at this time. We should pay more attention to emergence of complex systems and the “evolution” of experiences of robots. For example, we should consider whether it is possible to design a computer that has the property of a complex system. We need to determine how robots might accumulate experience through evolution.

5 Advantages of understanding a living organism from the point of view of information systems

Firstly, we should be aware that information inside the organism will play a very important role, and we should endeavor to build an information flow model of the organism [2,3]. Actually, biologists have always understood the importance of pathways. Once we have built up the information model, then we can analyze the behavior caused by the interaction of different parts inside

the organism. Using an information model, we can link different parts at different levels into a model and analyze it by system theory or control theory. We can more easily find the important factors that affect disease and consequently find clues to cure diseases. Secondly, from this new viewpoint, it is possible to arrive at conclusions that ordinarily would have been limited by one's own experience based on observations. For instance, the development process is non-reversible based on experience, i.e., few reversible results can be observed. However, from the point of view of information systems, any development process is merely a program that starts up a different group of genes. Thus, it is possible to instruct a recent stage to return to a previous stage through a reprogramming. In other words, any development process is reversible; the key point is how to reprogram it.

Analysis of the development program is the best way to solve this problem. In recent years, scientists have found several regulatory factors that can reverse the development stage through experience; this is a breakthrough event from the point of view of biology, but it is not surprising from the viewpoint of information systems. Similarly, the appearance of cancer cells possibly comes from a defect in the regulation process of cell division and apoptosis. To gain insight into this problem, we have to study the regulation program that runs cell division and apoptosis. Some clues should be found to treat cancer from this type of analysis. In addition, people may begin to consider the possibility of reversing cancer cells to normal cells. If this is possible, we should be able to find the program to do it.

There is also likely to be a program that decides our lifespan. Perhaps we can lengthen our life expectancy, even extend our lifespan infinitely. These are all very interesting problems and possibilities. People may derive new concepts that become quite different from the conclusions previously based on experience when we begin to look at these problems from this new viewpoint — information systems.

Thirdly, if we want to analyze the dynamics of different processes inside the organism, then building up an information model is a necessary step toward this. An information system with life features is the main concern in this paper, but information systems within an organism are particularly complicated because the organism needs material and energy from outside for metabolism. This involves extensive biochemistry and physical reactions between nucleic acids and (or) proteins. We have to understand the signal forms and signaling pathways represented in these reactions. Many chemical and physical scientists have been working hard to build these types of information systems. However, to analyze these information systems, we also need a new viewpoint, which is the emphasis of this paper. The problems that we have mentioned are very interesting areas, but are beyond the

current scope of this article.

Acknowledgements This work was partially supported by the National Natural Science Foundation of China (Grant No. 60775002).

References

1. Lu X, Sun Z R, Chen H, Li Y D. Characterizing self-similarity in bacteria DNA sequences. *Physical Review E*, 1998, 58(3): 3578–3584
2. Li Y D. Information and life. *Chemistry*, 2001, 64(10): 601–607 (in Chinese)
3. Li Y D. Understanding genome from information system point of view. *Acta Electronica Sinica*, 2001, 29(12A): 1731–1734 (in Chinese)



Yanda Li was born on Oct. 1936 in Dongguan, Guangdong Province in China. He was graduated from Department of Automatic Control, Tsinghua University in 1959. From 1979 to 1981, he visited the Massachusetts Institute of Technology, USA, and engaged in signal processing. He is Professor in Tsinghua University.

He is also Member of Chinese Academy of Sciences (CAS), Member of The Executive Committee of the Fifth CASAD Presidium, the Chairperson of Division of Information Technical Science of CAS. He is Editor-in-Chief of *Frontiers of Electrical and Electronic Engineering in China*.

He has a long interest in signal processing theory and its applications, including reconstruction of signal using partial data, wavelet analysis, and fractal signal processing. His theories have been used in many areas, particularly in oil exploration. He has also delved into areas of intelligent signal processing and systems, developed theoretical models of artificial neural network, fuzzy system and evolutionary algorithm, and applied them to the intelligent control of information networks. He also engaged in areas such as information mining and compression, organization of multimedia data in high-speed network, and intelligent control of industrial process and equipment. Since 1997, He has devoted himself to bioinformatics, applying information processing methods for complex system to molecular biology. He pioneered areas such as analysis of the information structure of genome sequence and modeling and simulation of gene control networks.

He has published a series of books including *Signal Reconstruction and Its Application* and *Neural Network Signal Processing* and more than 100 scientific papers.

He consecutively won the National Natural Sciences Award, Scientific and Technological Progress Award of National Education Commission, Beijing Scientific and

Technological Progress Award, National Excellent Scientific and Technological Books, and National Special-Class Teaching Achievement Award.