Second level code of life: Existence and role of protein program

Bo Zhang[1,*]

[1]The Genetic Program Lab, Xi'an 710119, China

* Email address: zhb609@qq.com ORCID: 0000-0002-4895-460X

Abstract

The traits of an organism are determined by DNA or RNA genetic information. Based on modern genetic theory, such a determination path is from genetic information to protein structures and protein functions and then to the traits or abilities of an organism. However, the microscopic activities of life molecules in many phenomena and functions of specific proteins are not well explained only by the structures of proteins and other life molecules.

I gave a classification of rules things operate in accordance with, which includes rules of natural laws, rules of structures and rules of information codes, and made inferences based on modern genetic theory and the concept of the three types of rules to reach that life molecules may act following information codes. Then, I proposed the hypothesis of a new conceptual model of life, namely, the genetic program of genetic substance molecules, which is the second level code of life, compared with the genetic information of DNA/RNA which can be called the first level code of life. Genetic programs are carried and executed by genetic substance molecules, including DNA, RNA, and protein, and define certain rules of actions of life. The protein program is now the mainstay of the genetic program. The protein's role as the information substance of life is examined in this paper.

The theory provides a new perspective and justifications to explain the activities of biomolecules and biological functions. I also gave a new definition of life based on the concept of genetic program.

Keywords

Genetic program; Protein program; Second level code of life; First level code of life; Genetic information

1   Introduction

The molecular structure of DNA was discovered in the last century (Watson and Crick 1953). The genetic code was cracked (Nelson and Cox 2017). The central dogma of molecular biology explains the flow and delivery of genetic information (Crick 1970). DNA replication (Bell and Dutta 2002) and the processes of genetic information flow from DNA to RNA to protein, transcription (Dignam et al. 1983) and translation (Ramakrishnan 2002) have been described. Gene regulation was also discovered (Hobert 2008).

The following are descriptions based on genetic theory. Deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are genetic materials carrying genetic information. A gene is a nucleotide sequence with functional meaning in the genetic material molecule and can be expressed as a polypeptide with a

corresponding amino acid sequence. One or more polypeptides make up a protein. Proteins and other life constitutive substances form the organism and express the traits of the organism.

The traits of an organism are determined by DNA or RNA genetic information, and such traits are also determined by multilevel structural units of atoms, molecules, organelles, cells, tissues, organs, organ systems and their interactions. It is the protein that bridges both the genetic information and the multilevel structural units. The proteins' functions are considered to be determined by proteins' structures, which come from genetic information. At the same time, modern microscopy techniques are used to study the microscopic activities of life molecules. However, the microscopic activities of life molecules in many phenomena are not explained only by the structures of proteins and other life molecules, and the functions of specific proteins are not completely explained by their structures as well. Is there another perspective to explain activities of life molecules and protein functions? In addition to the DNA/RNA genetic information that determines the structure of proteins, whether there is kind of information code carried by life molecules and on which their activities are directly based? This is what this paper will argue and aim to answer. That kind of information code would be argued to be genetic program, which defines certain rules of life activities and is a second level code of life if we call the genetic code of DNA/RNA the first level code of life.

2 Methods

I took the method of inferential argument to obtain the main point: the existence of genetic program.

The life traits of an organism summarize its life structures and life behaviors. My argument started with the information determination relationship of genetic information and life traits and then to the action determination relationship of genetic substance molecules and life traits. Information determination is implemented by action determination. In the information determination relationship, the information is genetic information carried by genetic substance molecules of DNA, RNA and protein. In the action determination relationship, there are actions performed by genetic substance molecules of DNA, RNA and protein. To implement the determination and achieve life traits, the actions must follow certain rules.

Then, I turned to Rules by which things operate follow in a general sense. I proposed three types or levels of rules, including rules of natural laws, rules of structures and rules of information codes.

The actions of life molecules were elucidated and analysed under these three levels of rules. Some microscopic activities of life molecules were explored under rules of structures and rules of information codes. There are behavior phenomena of life molecules that cannot be explained by rules of structures only but can be explained with rules of information codes, so we can make a hypothesis that the rules of information codes of life molecules exist.

If exist, the rules of information codes of life molecules must come from genetic information. I called genetic program for information codes that define the rules. Therefore, hypothesis can be made for the existence of the genetic program.

I classified the genetic program as DNA program, RNA program and protein program by the types of the genetic substance molecules that carry and execute the genetic program. Then, the argument for the

protein program and protein's role as the information substance of life was developed later in the text.

3 Results

3.1 Existence of genetic program

3.1.1 Genetic substance molecules realizing life structure and conducting life behaviour

This expression is not wrong: biological genetic information defines the structures and behavioural abilities of a life. This relationship is depicted as

$$\text{Genetic information} \dashrightarrow \text{Life structures} + \text{Life behaviours} \tag{1}$$

The dashed arrow represents the direction of information determination.

Genetic information comes from or is carried by genetic materials DNA or RNA.

For life structures, the composition materials of life include water, amino acids and proteins, sugar and polysaccharides, nucleosides and nucleic acids, lipids and fatty acids, inorganic salts and ions, etc. These substances constitute three-dimensional life structures.

For life behaviours, we can list the following behavioural abilities of life: obtaining resources (energy, materials), metabolism, growth and development, response to the environment, reproduction, storage of information (memory), construction activities for external objects, etc. Life behaviours also include building life structures.

The central dogma explains the flow of genetic information. Generally, the information transfer sequence is from DNA to RNA and then from RNA to protein through the processes of transcription and translation. After these processes, the genetic information of the organism is in materials including DNA, RNA and proteins and can be expressed as

$$\text{Genetic information} = \text{DNA genetic information} + \text{RNA genetic information} +$$
$$\text{Protein genetic information} \tag{2}$$

Combined with Exp. 2, Exp. 1 is expressed as:

$$\text{DNA genetic information} + \text{RNA genetic information} + \text{Protein genetic information} \dashrightarrow \text{Life structures}$$
$$+ \text{Life behaviours} \tag{3}$$

The information determination of Exp. 3 must be realized by the action of physical matter. After transcription and translation, which implement the genetic information flow of the central dogma, DNA molecules, RNA molecules and protein molecules carry genetic information. Information determination is realized by the action of these information-carrying molecules. We convert Exp. 3 from information determination expression to action determination expression and obtain:

$$\text{DNA molecules} + \text{RNA molecules} + \text{Protein molecules} \longrightarrow \text{Life structures} +$$
$$\text{Life behaviours} \tag{4}$$

The solid arrow represents the direction of action determination.

DNA molecules, RNA molecules and protein molecules are collectively referred to as genetic substance molecules. Then, we obtain:

$$\text{Genetic substance molecules} \longrightarrow \text{Life structures} + \text{Life behaviours} \qquad (5)$$

Genetic substance molecules are the subject of action here.

Exp. 5 implies that the actions of genetic substance molecules determine life structures and life behaviours. The determination also depends on other matters, such as the composition materials of life mentioned in the explanation of Exp. 1, but these matters are not the subject of action; they are prerequisites for the deterministic relationship, similar to environmental conditions such as sunlight and temperature, and are not included in Exp. 5.

As the subject of action, genetic substance molecules to realize life structure and conduct life behaviour must follow certain rules. Let us first analyse the concept of rules and then apply it to the action of genetic substance molecules.

3.1.2 Types of rules

Rules are laws that things operate in accordance with. According to where they are, rules can be classified into the following three types. They also represent the three levels of rules.

1. Rules of natural laws

First, natural laws, or natural physical laws, are rules that govern matter/energy, time and space. They are rules of the bottom of the world. Whether or not they are discovered by people, behaviours and interactions of all the particles, atoms, molecules and their composition objects follow them.

2. Rules of structures

Rules also lie in the structures of objects. Natural physical laws are nature's information and are invariable in our universe. How natural physical laws act on objects is determined by structural elements, such as the composition, combination, and shapes of specific objects. Thus, structures are rules.

Structures determine the rules of objects running and interacting. For example, a room has its maintenance structure with a door. The maintenance structure defines the division of internal and external matters, and people and objects that interact with this room can only enter and exit at the position of the door. Similarly, bilayer lipid membranes define the cell space and division of its lumen and inclusions, preventing the passage of interactive objects such as polar, hydrophilic substances and macromolecules. These substances and molecules can only pass in a selective manner through transmembrane proteins on membranes. For another example, the internal combustion engine defines the way in which the fuel burns and energy works in its interior by its structure, producing power and driving the vehicle wheels.

These structure-determined rules are rules of structures. The rules are realized by the direct effect of structures of interacting objects under natural physical laws.

3. Rules of information codes

Rules can also lie in information codes. A direct example is the code of a computer program that instructs the operation of the computer. There are two terms, "information" and "code", need to be explained. What is information? As Luciano Floridi said, "Information is still an elusive concept." (Floridi 2011) "Information is notoriously a polymorphic phenomenon and a polysemantic concept so, as an explicandum, it can be associated with several explanations, depending on the level of abstraction adopted and the cluster of requirements and desiderata orientating a theory." (Floridi 2009) He summarized an informational map covering concepts from data to semantics to knowledge and theories from the mathematical theory of communication (MTC) to the philosophy of information. I agree with his statement: "information is not about representing the world: it is rather a means to model it in such a way as to make sense of it and withstand its impact" (Floridi 2011). Therefore, the information I am referring to here is biased towards semantics. It is a model that expresses semantics or meaning, which can be extracted from the structure of the objects, rather than the structure itself or data of the structure. The model is represented by codes. What is code? The definition of code can be as follows: "A code is defined as a correspondence between two independent worlds" (Barbieri 2003). In this sense, the information or meaning relating to a structure can correspond to another structure by code, and the code can be defined by a set of rules that establish a correspondence between two structures. Or based on definition of the word "code" in the dictionary: "a system of words, letters, numbers or symbols that represent a message or record information secretly or in a shorter form" (Hornby 2014), the message or meaning itself represented by symbols or structural forms of some system can also be called "code". The message or meaning here can be what we talked about the rules a thing operates in accordance with. The rules can be mapped to multiple sets of information codes by multiple symbols or structural forms.

Therefore, rules can lie in information codes. Depending on different sets of codes, the rules can be carried by multiple physical carriers and can have multiple manifestations. When acting on a related object, the rules need to be extracted from its physical carrier. Decoding and execution are needed for the rules to be extracted and take effect. The process of transforming a meaning from one structural form into another is called decoding. Information code is eventually converted to some structural elements and works through the final structures presented. This can be achieved by decoding or a series of decoding through which the meaning of the code is realized. Doing the corresponding decoding or series of decoding by a subject can be called the execution of the information code, and the verb "execute" can be used for the doing.

Text and language are rules of information codes. For example, "Give me five" can be considered a rule for greeting: the rule is expressed as an open hand expanding five fingers. Let us look at "five": "five" is number 5, there are many forms of expression of number 5 in text or other information code, such as

"five" in English, different representations in other languages, "5" in Arabic numbers, abacus beads "⚱",
"101" of binary number and ". . . . ." of Morse code. The physical carrier can be ink on paper, light and shadow on the screen, a wooden abacus, a semiconductor chip, "di di di di di" from a telegraph, and human voices. These are information codes, and the rules they represent are rules of information codes. Decoding and execution are needed to make the rule work. Codes of text and language are decoded and executed by humans, and the final manifestation of the rule is the structure of a palm with 5 fingers

expanded (Fig. 1).

Fig. 1 Give me five

Computer programs are information codes. A rule can be embodied in programs in different programming languages. The carrier of the program can be paper tape, magnetic tape, CD, transistor chip, etc. Decoding and execution of the program are needed for implementation of the rule, which is done by compiling the program and logic-enabled electronics. They are ultimately converted to electric currents producing light and shadow, sound, or object motion to implement the rule they express.

The genetic information of a gene is the information code. It can be expressed by the ACTG base sequence of a DNA double strand, and it can also be expressed by the ACUG base sequence of an mRNA single strand. Polypeptide chains with specific amino acid sequences are its structural presentation. The decoding from code to structure relies on tRNAs as adaptors of translating triple base code to a certain amino acid. The binding of amino acids to their tRNAs is carried out by a family of enzymes named aminoacyl-tRNA synthetases. When tRNAs and mRNA with the information enter the ribosome and match to generate a polypeptide chain of a specific amino acid sequence, decoding is realized. The final products are polypeptide chain structures of proteins.

3.1.3 The genetic rules and existence of the genetic program

As seen in Exp. 5, genetic substance molecules determine life structures and life behaviours in accordance with certain rules. They are rules of life action. Rules of life action are from genetic information, so they can be named genetic rules, which will be analysed according to the types of rules presented above.

1. Rules of natural laws

Life materials consist of matter particles. The actions of genetic substance molecules inevitably follow natural physical laws.
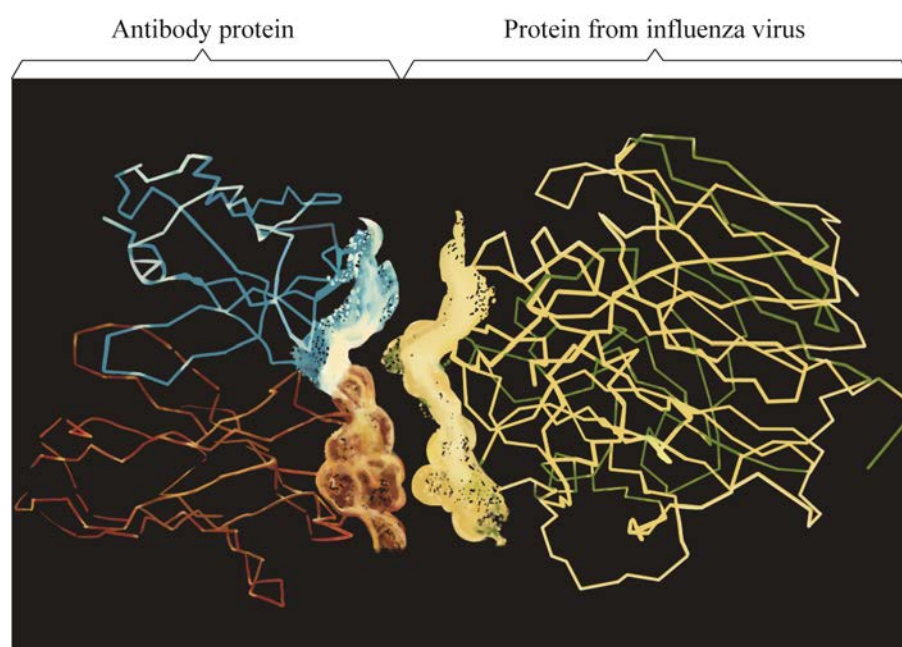
2. Rules of structures

As mentioned earlier, rules of structures are rules defined by structures of interacting objects under natural physical laws.

After the transmission of genetic information as the central dogma explains, and after protein processing and folding, structures of genetic substance molecules (DNA, RNA and proteins) have been determined. Interacting objects are other genetic substance molecules or other life constituent molecules, ions, and

6

their structures are also formed through the same processes or processes of nature. They generally act and interact according to rules of structures. The working of bilayer lipid membranes of cells described earlier is an example. Let us look at proteins' work.

A protein's specific structure determines how it works. In very common cases, the function of a protein depends on its ability to recognize and bind to some other molecule, which is determined by whether the structural form of the protein matches the structure of a foreign molecule. For example, Fig. 2 is a computer model of an antibody protein (blue and orange, left) bound to an influenza virus protein (yellow and green, right). This is a wireframe model modified by adding an "electron density map" in the region where the two proteins meet and then using computer software to back the images away from each other slightly. It shows some exact match of shape between an antibody protein and a substance on influenza virus (Urry et al. 2017). More details of the structure between influenza virus and an antibody can be seen in the article of Colman et al. (1987). Structural matching determines that the antibody can bind the protein from the influenza virus and marks the virus for elimination.

Fig. 2 Antibody protein matches with protein from influenza virus by complementarity of structural shape (Photo credit: Peter M. Colman. Image source: Campbell Biology, 11th Edition.)

This illustrates the rules of structures for proteins. We think that the functions of proteins are determined by their structures, and thus, we studied proteins and described and explained the functions of proteins based on their structures and interacting substances. However, is this sufficient? Can rules of structures explain all life phenomena?

3. Rules of information codes

This part studies the existence of rules of information codes for genetic substance molecules' action. The behaviour phenomenon of life is determined by the action of genetic substance molecules. If some behaviour phenomenon cannot be explained by rules of structures only but can be explained with rules of information codes, we can then know the existence of rules of information codes. Using robots as an

7

analogy, there are electromechanical robots and information robots. The functions of electromechanical robots are realized by their electromechanical structure. However, for information robots, in addition to mechanical and electrical construction, there are programs defining rules in information codes. You can make an appointment for a sweeping robot to sweep the floor. After cleaning the room, it can return to the charging position and charge itself. These behaviours cannot be explained by the electromechanical structure and rules of structures. It can be considered an information robot instead of an electromechanical robot. It contains program code, and its action rules include rules of information codes.

We are studying the rules in the action deterministic relationship of Exp. 5. For rules of information codes under this situation, they are different from the original genetic code on DNA/RNA. DNA/RNA code is the source and has been interpreted and implemented at an early stage before action. In this sense, the DNA/RNA genetic code can be called the "first level code of life", and the code of rules here is the "second level code of life".

If there are rules of information codes of life, the codes must exist in genetic substance molecules, including DNA molecules, RNA molecules or protein molecules, and be decoded and executed by genetic substance molecules. The codes come from genetic information and are decoded and executed like computer program, so we call them "genetic program". The genetic program can be called the "DNA program", "RNA program" or "protein program" corresponding to the specific genetic substance molecules of DNA molecules, RNA molecules or protein molecules. Genetic substance molecules act following the rules defined by the genetic program through the execution of the program.

Since the program in the name "genetic program" comes from the analogy to computer program, let's look at what is "program" in computer science. According to the Oxford Dictionary of Computer Science, "program" is "A set of statements that (after translation from programming-language form into executable form) can be executed by a computer in order to produce a desired behaviour from the computer" (Butterfield et al. 2016). It follows the previous argument that decoding and execution are needed for the rules of information code to be extracted and take effect. In the same way, the genetic program can be a set of statements that can be executed by a genetic substance molecule that can produce a specific behaviour from the molecule.

Another definition of the word "program" in dictionary is "a set of instructions in CODE that control the operations or functions of a computer" (Hornby 2014). In this sense, program can be considered equivalent to code that represents encoded meanings.

Now I take an example of cell mitosis. Mitosis is part of the cell cycle, which is the period of cell division. Mitosis is conventionally broken down into five stages: prophase, prometaphase, metaphase, anaphase, and telophase (Urry et al. 2017). The process of mitosis can be seen and photographed more clearly with modern microscopy (Rieder and Khodjakov 2003). The following is the situation from metaphase to anaphase in an animal cell (Fig. 3), with an analysis of protein behaviour in the process. At metaphase, all chromosomes have arrived at the metaphase plate. which is a plane that is equidistant between the spindle's two poles. The chromosome centromeres lie at the metaphase plate. Then, anaphase begins, the cohesin proteins are cleaved by enzymes called separase, the microtubules attached at the centromere shorten, and the two sister chromosomes begin moving toward opposite poles of the cell. The

kinetochore motor protein moves through the microtubule to drag the chromosome poleward, and the microtubule is shortened via depolymerization of the microtubule after the kinetochore motor proteins move through the microtubule (Gorbsky et al. 1987).

Fig. 3 is described as follows. a: Fluorescence micrographs showing dividing lung cells from a newt at metaphase to anaphase. The newt has 22 chromosomes. b: The drawings of metaphase to anaphase. For simplicity, the drawings show only 5 chromosomes. The drawing of metaphase shows the positions of centrosomes, microtubules constituting the mitotic spindle and the chromosomes at metaphase. The drawing of anaphase shows the action scene at anaphase: after the cohesin proteins are cleaved, the two sister chromatids of each pair part and move along spindle microtubules towards opposite ends of the cell. c: Detailed drawing of kinetochore motor proteins moving through the microtubules to drag the chromosome poleward. Shortening of the microtubule is disintegrating of microtubule synchronously after kinetochore motor proteins move through the microtubule.
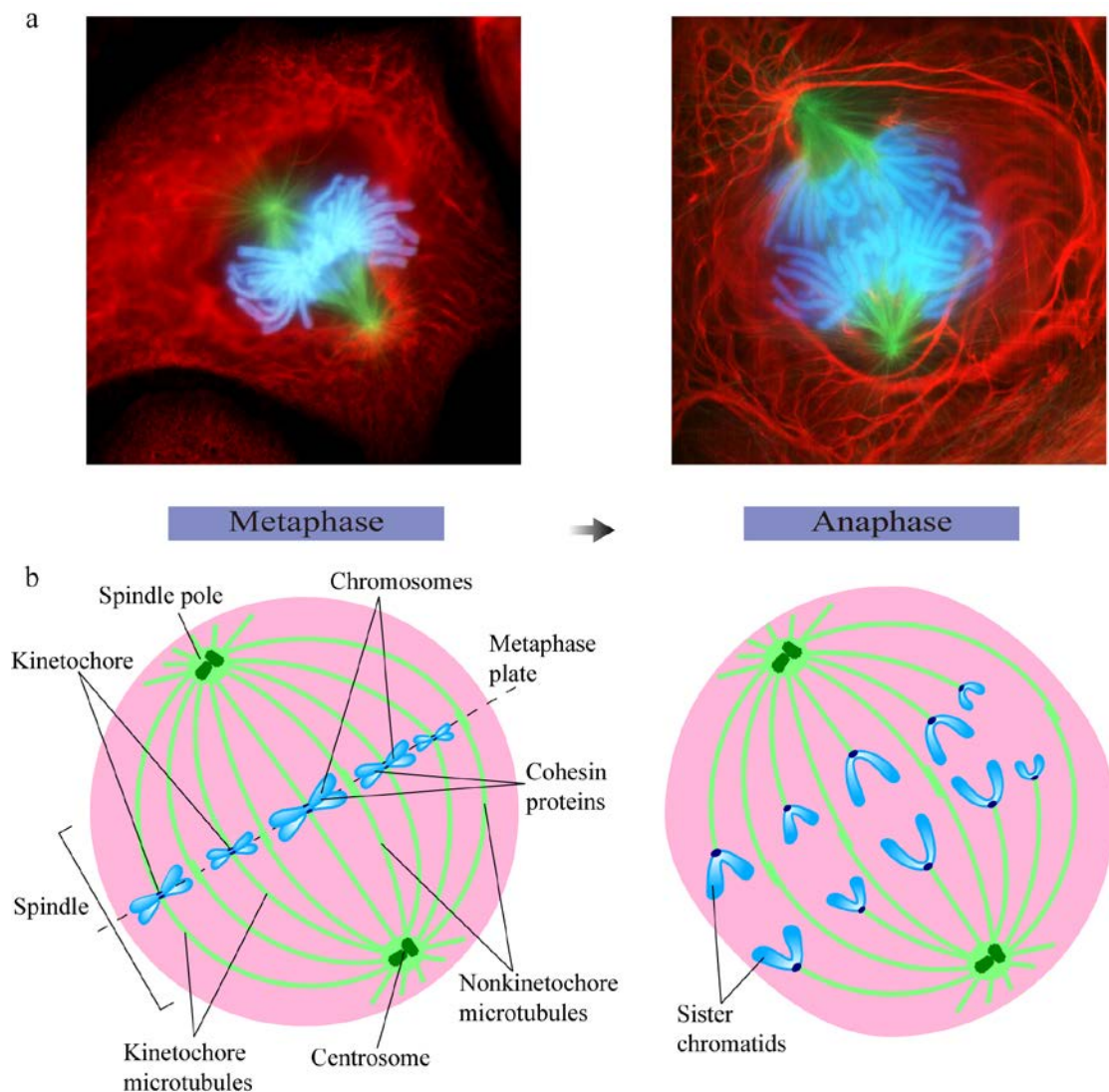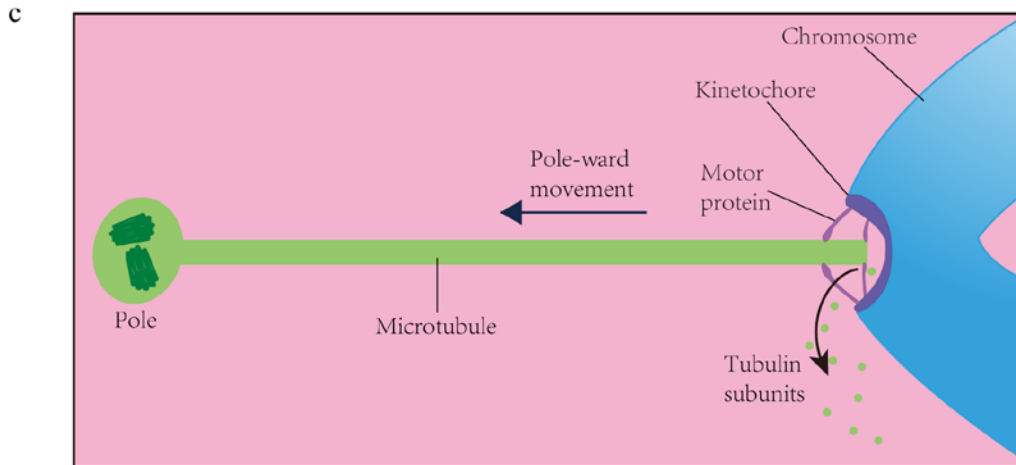
Fig. 3 Metaphase and anaphase of mitosis in an animal cell (Photo credit: Courtesy of Dr. Conly L. Rieder, Wadsworth Center, Albany, New York, USA.)

In this process, the following two aspects cannot be explained by rules of structures only.

1. Relative positions

In metaphase, each part of participation showed relative positions. Two centrosomes are at opposite poles, the microtubules are positioned to form the mitotic spindle, and the chromosomes dragged by the kinetochore are distributed at positions on the central plate.

Regarding the positions of proteins, proteins are either integrated into the lipid membrane, included in lysosomes, or secreted through the cell membrane. The targeting mechanism of proteins was first postulated by Günter Blobel and colleagues in 1970. The important element of the targeting mechanism is a short sequence of amino acids called a signal sequence or signal peptide (Walter et al. 1984). This targeting or positioning mechanism can be explained by structures. Related structural units are SP (signal peptide), SRP (signal-recognition particle) and SRP receptor protein.

There is no mechanism explaining the relative positioning described in metaphase, as if proteins have relative position awareness. Relative position information is important in life. For example, during embryonic development, the same cells develop into different types of cells depending on the relative location, and cells located in some locations are apoptotic during development (such as human cells that make up the webbed part between the fingers in the early embryo). This precise relative location information processing cannot be fully explained by rules of structures that are construction-based.

2. Time-based coordination

A scene can be seen in mitosis: Separases that cleave cohesin proteins and kinetochore motor proteins of each chromosome pair seem to wait for each other to arrive at the metaphase plate and start acting nearly simultaneously.

Molecular structures can define attachment relationships of matter molecules, such as the connection of molecules of the aster centrosome and spindle microtubules extending from the centrosome, the attachment of molecules of the kinetochore motor protein and the centromere of the chromosome, the attachment of molecules of the kinetochore motor protein and kinetochore microtubule, and the

attachment of chromosome and cohesin proteins that condense sister chromatids of the chromosome. However, molecular structures cannot determine the time-based coordinated actions of participating molecules, such as the simultaneous appearance of spindle microtubules, coordinated actions of separases cleaving cohesin proteins of all discrete chromosomes and motor proteins moving pole-ward along the microtubules. Time-based control plays an important role in life; for example, the turning on time of a gene has great impacts on the characteristics of the organism. Precise time control and time-based coordinated actions of many and multiple proteins cannot be explained by physical or chemical processes dominated by their structures.

If rules of information codes exist and work or, in other words, a genetic program exists, the corresponding two aspects of questions can be answered. The genetic program contains logic of position and time processing, making some proteins able to deal with relative positions and time information and then obtain the calculation results as conditions for action. They can receive or send signals to other proteins to coordinate proteins of different components. Actions can be carried out by signals coordinately.

This is similar to the example of the robot mentioned earlier. The sweeping robot with the timed appointment function that can return to the charging position on its own should be an information robot, and a coded program that defines function rules is carried by the robot.

Based on this reasoning, we can make a hypothesis that rules of information codes exist for the actions of life molecules. Based on Exp. 5, the carriers of the codes are genetic substance molecules. We call the information codes genetic programs. We can say that there are genetic programs that are kinds of second level code of life and are carried and executed by genetic substance molecules.

As information code, genetic programs can contain time- and position-related or other variables and logics to process these variables. Life molecules can take action by changing structures based on the results of handling. This provides a new perspective and justification to explain the activities of life molecules and life functions.

3.2 Protein program and its role

3.2.1 Protein program is the mainstay of genetic program

DNA molecules, RNA molecules and protein molecules are carriers of the genetic program and subjects for decoding and execution of the genetic program. We can further study the genetic program of these genetic substances by biomolecular activity. Biomolecular activity indicates the degree and extent to which action is performed for one kind of genetic program. It shows that the physiological behaviours of organisms are mainly performed by proteins. There are a few active nucleic acids as described below. Ribozyme, which is a small molecule RNA with catalytic function, was discovered by T. R. Cech and S. Altman in the early 1980s (Cech 1990; Doherty and Doudna 2001). Most known ribozymes now catalyze intramolecular reactions, including self-cutting, self-splicing, and self-circulation of RNA. Ribozymes that catalyze intermolecular reactions usually bind to proteins and form ribonucleoprotein complexes. Through in vitro molecular evolution techniques, DNA has been shown to have catalytic properties similar to RNA (Breaker 1997). Deoxyribozyme or DNAzyme catalyzes the cleavage reaction of specific

parts of RNA. However, no naturally occurring catalytic DNA has been found.

Combining the above facts, judging from the activity of molecules of genetic substances, DNA molecules, RNA molecules and protein molecules can all execute the genetic program, but molecular activities of DNA molecules and RNA molecules are mainly shown in processing RNA after synthesis. Actions of life are mainly performed by proteins. It can be said that proteins are the main body that carry and execute the genetic program, and the protein program is the main executable program of the genetic programs. In the following, the execution of the protein program by proteins is used to explain the execution of the genetic program.

If we simplify Exp. 4 and ignoring the execution of DNA molecules + RNA molecules, we obtained:

$$\text{Protein molecules} \longrightarrow \text{Life structures + Life behaviours} \tag{6}$$

The protein program can be described in another way. Genetic information contains not only structural information but also program information that defines the action rules of genetic substance molecules. The program information is the genetic program. In addition to the structures being expressed, the genetic program must be dispatched from genes to proteins. Proteins are discrete in the organism; they are discretely distributed in the cytoplasm or nucleoplasm, embedded in the lipid membrane, polymerized into the fibrous skeleton of cells, or distributed as signal molecules in body fluids. The genetic program is also dispersed into these individual proteins to take effect. The genetic program distributed into proteins is dispatched from genetic material with the process of protein production.

Genetic information defines the structures of proteins. However, how do the structures of proteins define the structures of other parts and functional information of life? If there are only structures, the protein is still functionally static. Execution of the protein programs shows the activity of life. Proteins are always active or dynamic molecules in functional performance. Proteins interact with other molecules to achieve their function. They can receive other molecules and react. The result of the reaction involves altering the chemical configuration or composition of the interacting molecule, such as what enzymes are doing, transporting substances and producing movements by reversible binding of other molecules, such as hemoglobin's binding with heme and the attachment of myosin and actin. The functional dynamics of protein molecules indicate that the protein can execute the information code - protein program, which is a second level code of life.

3.2.2 Protein program coexists with protein structure

Action determination is a result of information determination. From Exp. 6, we can return to the point of information determination (dashed arrow) as

$$\text{Protein information} \dashrightarrow \text{Life structures + Life behaviours} \tag{7}$$

Because there is a protein program, protein information consists of protein structures and protein programs. Then, we can obtain

$$\text{Protein structures + Protein programs} \dashrightarrow \text{Life structures + Life behaviours} \tag{8}$$

Exp. 8 indicates the decisive role of rules of structures and rules of information codes for life.
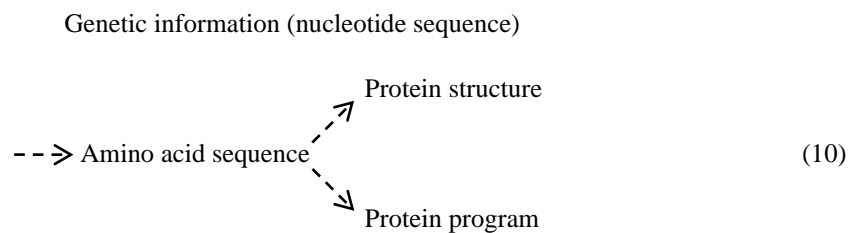
According to Exp. 1, and the upstream and downstream relationship of information, which can be derived from Exp. 8 is

$$\text{Genetic information} \dashrightarrow \text{Protein structures} + \text{Protein programs} \qquad (9)$$

As described previously, protein structures are determined by genetic information from the DNA/RNA fragment of a gene. The nucleotide sequence of the gene was decoded to the amino acid sequence of the polypeptide. One or more polypeptides that are twisted, folded, and coiled make up a protein with a special spatial structure. However, it is unclear where the protein program comes from.

I think that the protein program is also determined by an amino acid sequence. Proteins have three-dimensional structures that can be described as having four levels of structure: primary, secondary, tertiary, and quaternary structures (Urry et al. 2017). The protein program should be one-dimensional from the perspective of information. In this sense, the amino acid sequence of the polypeptide determines the protein program, or the protein program has nothing to do with the secondary, tertiary or quaternary structure of the protein but exists in the primary structure of the protein, which is a linear chain of amino acids of the polypeptide.

The information determination of this process can be expressed as

$$\text{Genetic information (nucleotide sequence)}$$

$$\dashrightarrow \text{Amino acid sequence} \nearrow \begin{array}{l} \text{Protein structure} \\ \\ \searrow \text{Protein program} \end{array} \qquad (10)$$

Exp. 10 can be considered an extension of the central dogma, pointing out the source of information for proteins from both structural and programmatic perspectives.

This conclusion can then be reached: the protein program coexists with the protein structure. In addition, both the protein structure and the protein program come from the amino acid sequence, which is derived from genetic information.

Realized by rules of structures and rules of information codes, the function of a protein is determined by its structural function and its programmatic function.

$$\text{Function of protein structure} + \text{Function of protein program}$$
$$\dashrightarrow \text{Function of protein} \qquad (11)$$

From the previous examples (Fig. 2), the binding of antibody and protein on the influenza virus by shape matching explains the function of protein structure. Enzymes cleaving cohesin proteins and kinetochore motor proteins moving pole-ward on microtubules simultaneously during mitosis explain the function of the protein program.

In the entire amino acid sequence of a protein, let us assume that some part of the amino acid sequence plays a role in the protein structure and that another part plays a role in the protein program, or in some

cases the same sequence part may play both roles. Overall, the amino acid sequence totally defines the protein structure and the protein program. If considered a functional unit containing information codes, the amino acid sequence defines both the hardware and software of a protein.

3.2.3 Protein is an information substance of life and the execution subject using genetic information

As seen from Exp. 10, the protein program is a genetic program of life that is determined by genetic information. Not only is nucleic acid (DNA and RNA) the substance of genetic information, but protein, which carries the protein program, is also the substance of genetic information. The two kinds of substance of genetic information, namely, nucleic acid and protein, can be summarized as follows:

DNA (or RNA for some lives) is the information source of life. Nucleotide sequences carried by DNA or RNA constitute the first level code of life. Therefore, DNA/RNA is the first level substance of genetic information – indeed, it is the genetic information book of an organism.

The protein program carried by the protein is a second level code of life. It can be said that protein is a kind of second level substance of genetic information. Protein carries information fragments of corresponding nucleic acid gene units.

The first and second level codes or the first and second level substances of genetic information represent the determination order of genetic information and the whole and partial relationship of genetic information.

Protein also plays a role of execution subject using genetic information. Because the protein program is the mainstay of the genetic program and the protein program is carried by protein, protein is the main substance molecule unit executing the genetic program. (DNA and RNA can also execute their own program, see previous discussion.)

Fig. 4 depicts the relationship between nucleic acid and protein, the two major genetic substances of life, and their relationship with the body of organism from both life structures and life behaviours (the body represents other parts of the organism). Among them, protein as the execution subject is the executor of the building of life structures ("build" in Fig. 4) and the executor of life behaviours ("behave" in Fig. 4). When building proteins, proteins read the genetic information of the corresponding gene from the nucleic acids ("read" in Fig. 4). Proteins can regulate nucleic acid genes to affect proteins reading gene information for protein building ("regulate" in Fig. 4). The processing of RNA by DNA/RNA (ribozyme and deoxyribozyme) is also reflected ("process" in Fig. 4).
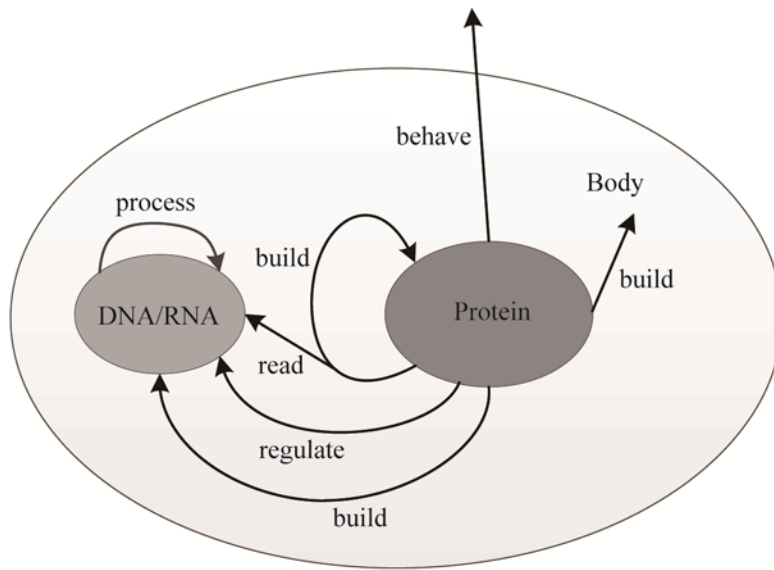
Fig. 4 Information substances and protein as the execution subject of life

3.2.4 Proteins constitute the life information processing network

Position and time of life activity are used to infer the existence of rules of information codes previously, but if rules of information codes of life exist, they are not limited to dealing with position and time issues. The rules of many aspects of life behaviour can be in rules of information codes. An organism can sense, process, and send information. It can sense information such as light, sound, pressure, smell, taste, cold and heat (microscopic particle motion) by sensory organs, processes perceived information and then acts or reacts accordingly. Computing and logical reasoning are also the information processing capability of life. Capabilities of information processing are also determined by the genetic information of life, and based on the previous analysis, they are realized by the action of genetic substance molecules. Specific rules of the action can be rules of structures and rules of information codes.

Neuroscience and brain science are important research directions in the study of life information processing. The nervous system is mainly composed of neurons and glial cells, and it is currently believed that neurological functional activities are mainly undertaken by neurons. Chemicals such as electrical ions and neurotransmitters are transmitted between neurons. Through the communication of electrical signals and chemical signals, neurons are connected and form the neural network system. The brain is the most advanced part of the nervous system, with a large number of neurons making up the most complex part of the neural network. For example, there are approximately 20 billion neurons in the human cerebral cortex (Pakkenberg and Gundersen 1997). The brain neuron network is used for life information processing. In other words, the structure of the nervous system determines the information processing function of the brain. This is the embodiment of rules of structures.

Here, I apply rules of information codes to the information processing capability of life. As discussed earlier, the protein program is the mainstay of the genetic program and is carried and executed by proteins. Information processing of life can be realized by proteins executing the protein program.

As mentioned earlier, proteins are discrete in the organism; they are either dispersed in the liquid of the body or embedded in the lipid membrane. Proteins in the lipid membrane are also in contact with liquid.

15

The liquid in the body provides the medium for proteins to connect. Proteins are connected by other molecules or ions in the body fluid as signals. Proteins can bind some molecules or ions, release some molecules or ions, and exchange information through these molecules or ions. The overall information processing is ultimately distributed to many proteins to process. In addition, proteins communicate through life's signaling pathways, take input and output physical or chemical signals to the outside world, and form the overall information processing network.

Cells are currently considered the most basic units of life. A cell is also a protein network, shows the overall appearance of its basic characteristics, and accepts external signals and outputs signals. In fact, the signal receiving and outputting of cells are also completed by specific proteins. There are controlling proteins that can coordinate the activity of multiple proteins in cells. According to the division of tissues and organs, each organ or tissue can be said to be a functional network formed by specific connected cells. These functional networks include the neural network as an advanced information processing system. The whole body of an organism is a large information processing network. However, proteins are still the most basic information processing execution units – protein program execution units.

The single-cell eukaryote Stentor roeseli behaves in a complex hierarchy of avoidance (Dexter et al. 2019). When encountering stimulation, S. roeseli behaves by bending away or ciliary alteration most of the time; if the stimulation continues, it then behaves as contractions or detachment from where it anchored itself. Therefore, it shows a priority order of avoidance behaviours. The hierarchy of behaviours seems to indicate that some relatively complex decision-making calculations are performed inside a single cell. From the perspective of the protein program and the protein information processing network, this nonneural system calculation can be understood naturally.

4 Discussion

4.1 Code, decoding and execution of the protein program

As stated in the previous discussion about the rules of information codes, the information code can be extracted from its physical carrier. It can be represented in a variety of forms and carried by a variety of physical carriers. The code needs a mechanism for decoding or execution.

The genetic code, which is the first level code of life, has been cracked. The presence of the code in the carrier DNA and RNA and the processes of information decoding are also known.

This paper demonstrates the existence of the second level code of life – the genetic program and the protein program, which is the main genetic program. The carrier substance of the protein program is protein. It cannot be said that the protein program is carried by multiple substances, but it needs to be decoded or executed, and this conforms to the characteristics of rules of information codes.

However, how is the protein program encoded, decoded? There should be a deep mechanism for this. But it's an unknown part for us now.

According to Exp. 10, the protein structure and protein program come from the amino acid sequence. I will briefly analyse the code of the protein program by amino acid sequence. Let us write down the amino acid order of the protein and consider this as the basic source code of the protein program. Just as all

creatures on earth use one genetic code, it can be assumed that the protein program is coded by one coding system, and protein can decode this encoded information.

Let us look at the basic source code of this program. Protein has 20 amino acids, 6 fewer than 26 English letters in the human language. Twenty amino acids have been agreed to be represented by 20 English letters (Urry et al. 2017), so the basic source code of protein can be written in 20 English letters. However, this is just a form, and the true meaning of coding still needs to be cracked. What is handled by the program should include electrochemical signals representing the perception of the environment inside and outside the organism, initiating behaviour using energy, and logical processing.

4.2 A new definition of life

As a member of life, we may be able to identify a life intuitively, but it is not easy to define life in one sentence. Life behaviours explained under Exp. 1 are characteristics that can be used to define life. To define life, using one or some of the characteristics is not enough, but using all characteristics seems too strict and will miss some.

The first principle of semantic biology summarized by Marcello Barbieri is that epigenesis is a defining characteristic of life. In more detail, what is crucial to life is the ability to produce a convergent increase in complexity. The principle is nothing more than a new definition of life. He also listed many definitions of life in history in the appendix of the book (Barbieri 2003). The fourth principle of semantic biology he summarized is that there cannot be a convergent increase in complexity without codes; in other words, organic epigenesis requires organic codes. Combining these two principles, life is inseparable from codes. The organic codes of four models in Barbieri's book are not the same as the codes of the genetic program of this paper, but the principles they followed are the same. The codes of the genetic program can be executed at the bottom for organic epigenesis and other kinds of life formation and behaviours.

According to the concept of genetic substance molecules carrying and executing genetic programs in this paper, I try to give a new definition of life as follows: *A life is a body that contains molecules carrying genetic programs, and at least one of the molecules has not lost the ability to execute its genetic program.*

This definition covers all life forms, including prokaryotes, unicellular and multicellular eukaryotes, viruses, cells without genetic material such as red blood cells and platelets, proteins without genetic material such as prions, gametes of plants and animals, fertilized eggs, plant seeds and frozen cells at -196 °C.

5 Conclusions

Rules that things operate in accordance with can be classified into three types or levels: rules of natural laws, rules of structures, and rules of information codes.

Based on the rules of structures, the bottom of life activity can be explained by the structures of the many organic and inorganic molecules, ions that make up living organisms. The important parts are protein structures, which come from genetic information.

Based on rules of information codes, the bottom of life activity can also be explained by the genetic

programs that are carried and executed by genetic substance molecules, including DNA, RNA and protein molecules. The genetic programs come from genetic information and are a "second level code of life" compared with the "first level code of life", which is the genetic code of DNA/RNA.

The protein program is the mainstay of the genetic program, though the existence of the DNA program and the RNA program should not be excluded.

The protein program coexists with the protein structure. The protein structures and protein program can both realize some parts of a protein's function that are structural function and programmatic function.

Nucleotide sequences carried by DNA or RNA constitute the first level code of life, and DNA/RNA is the first level substance of genetic information. The protein program carried by protein is a second level code of life, and protein is a second level substance of genetic information. Protein also plays a role of execution subject using genetic information.

Many aspects of life behaviour can be studied by protein program and protein's role as execution subject using genetic information, including processing perceived information, and making reactions accordingly. With protein programs and protein communication through life's signalling pathways, discrete proteins constitute the life information processing network.

As information code, the protein program needs a mechanism for decoding and execution. The underlying mechanism is the part we don't yet know.

Acknowledgements

References

Barbieri M (2003) The organic codes, An introduction to semantic biology. Cambridge, New York, pp 93,244-263

Bell SP, Dutta A (2002) DNA replication in eukaryotic cells. Annu Rev Biochem 71:333–374

Breaker RR (1997) DNA enzymes. Nat Biotechnol 15:427–431

Butterfield A, Ngondi GE, Daintith J, Wright E (2016) A Dictionary of Computer Science 7th edn ebook. Oxford, New York, pp 862

Cech TR (1990) Self-splicing of group I introns. Annu Rev Biochem 59:543-568

Colman PM, Laver WG, Varghese JN, Baker AT, Tulloch PA, Air GM, Webster RG (1987) Three-dimensional structure of a complex of antibody with influenza virus neuraminidase. Nature 326: 358–363

Crick FHC (1970) Central dogma of molecular biology. Nature 227:561–563

Dexter JP, Prabakaran S, Gunawardena J (2019) A complex hierarchy of avoidance behaviors in a

single-cell eukaryote. Curr Biol. https://doi.org/10.1016/j.cub.2019.10.059

Dignam JD, Lebovitz RM, Roeder RG (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res 11:1475–1489

Doherty EA, Doudna JA (2001) Ribozyme structures and mechanisms. Annu Rev Biophys Biomol Struct 30:457-475

Floridi L (2009) Philosophical Conceptions of Information. In: Sommaruga G (ed) Formal Theories of Information: From Shannon to Semantic Information Theory and General Concepts of Information. Springer, New York, pp 13-15

Floridi L (2011) The Philosophy of Information. Oxford, New York, pp xiv, 30

Gorbsky GJ, Sammak PJ, Borisy GG (1987) Chromosomes move poleward in anaphase along stationary microtubules that coordinately disassemble from their kinetochore ends. J Cell Biol 104:9–18

Hobert O (2008) Gene regulation by transcription factors and microRNAs. Science 319:1785-1786

Hornby AS (2014) Oxford Advanced Learner's English-Chinese Dictionary. Commercial, Beijing, pp 379, 1635

Nelson DL, Cox MM (2017) 27.1 The Genetic Code. In: Lehninger Principles of Biochemistry 7th edn ebook. W. H. Freeman and Company, New York, pp 2822-2847

Pakkenberg B, Gundersen HJG (1997) Neocortical neuron number in humans: Effect of sex and age. J Comp Neurol 384:312-320

Ramakrishnan V (2002) Ribosome structure and the mechanism of translation. Cell 108:557-572

Rieder CL, Khodjakov A (2003) Mitosis Through the Microscope: Advances in Seeing Inside Live Dividing Cells. Science 300:91-96

Urry LA, Cain ML, Wasserman SA, Minorsky PV, Reece JB (2017) Campbell Biology 11th edn. Pearson, New York

Walter P, Gilmore R, Blobel G (1984) Protein translocation across the endoplasmic reticulum. Cell 38:5–8

Watson JD, Crick FHC (1953) Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. Nature 171:737–738

Watson JD, Crick FHC (1953) Genetical implications of the structure of deoxyribonucleic acid. Nature 171:964–967