Global Philosophy (2023) 33:43
https://doi.org/10.1007/s10516-023-09692-0

ORIGINAL PAPER

Categorical Abstractions of Molecular Structures of Biological Objects: A Case Study of Nucleic Acids

Jinyeong Gim[1]

Received: 16 January 2023 / Accepted: 27 July 2023

Abstract
The type-level abstraction is a formal way to represent molecular structures in biological practice. Graphical representations of molecular structures of biological objects are also used to identify functional processes of things. This paper will reveal that category theory is a formal mathematical language not only to visualize molecular structures of biological objects as type-level abstraction formally but also to understand how to infer biological functions from the molecular structures of biological objects. Category theory is a toolkit to understand biological knowledge at the type-level formally, not individual token-level, as well as typical heuristic strategies in molecular biology.

Keywords Abstraction · Biological structure · Category theory · Graphical representation · Visualization · Macromolecule

1 Introduction

The most central tasks in molecular biology, genetics, and biochemistry, are to represent molecular structures of biological objects in a specific biological mechanism and predict or explain their functions. The search for the structure of biological objects and their functions is a fundamental goal at the molecular level of research. This typical research style assumes that structures of biological things imply their biological functions. For instance, the lock and critical docking system consisting of an enzyme and its substrate is closely involved in their geometric–mechanical shape, and electrochemical states. Due to the secondary cloverleaf structure of transfer RNA, composed of two sites such as the amino acid-attachment site and the anticodon site, tRNA can play a role in carrying genetic codes from messenger RNA to proteins. To successfully predict and explain the biological functions of an object

✉ Jinyeong Gim
jinyeong.gim@gmail.com

[1] College of Liberal Studies, Seoul National University, Gwanak-ro 1, Gwanak-gu, Seoul 08826, South Korea

Published online: 17 August 2023

within biological systems, researchers should attempt to comprehend its structural information, which is called the structure-function relationship.

What kind is the structural information of biological objects? It includes the target object's size, shape, location, and orientations, called *spatial organizations*. Those features of spatial organizations are determined by *hierarchical compositions*. Most biological objects are two or three-dimensional structures. Of course, it is well-known that those structures profoundly depend on their primary or secondary structures. Furthermore, lower-level building blocks determine a biological object's primary or secondary structure. Macromolecules (e.g., poly-nucleic acids, proteins, polysaccharides, etc.) consist of their units, such as nucleotides, amino acids, and saccharides, respectively.

Understanding the structural features of biological things is the beginning of the pursuit of biological knowledge at the molecular level. Biological knowledge is generally about two targets: (i) objects themselves and (ii) generalized laws or patterns among them. The latter, for instance, include Mendel's laws, Darwin's theory of natural selection, and diverse mechanisms among molecular entities. Those empirical or theoretical generalizations are based on type-level abstractions of biological objects. Mechanistic representations of common reactions among biological molecules depend on their structural features. So, knowledge of the structural characteristics of biological objects is worth investigating in biological practice.

How can we understand the structural features of biological objects? Is it possible to abstract biological objects and generalized laws or patterns, too? Yes. Pursuing formal frameworks is also a promising way to represent the structural features of objects in biology abstractly. And then, why do we consider formal frameworks to represent biological knowledge? Most biologists deal with a particularly individuated instance of some specific objects such as *a* bacteria, i.e., *E. coli*, or *a* DNA sample purified from *Tetrahymena* in their laboratories. Many diverse individual objects in biological systems are used in investigating diverse research topics. But we also know that heterogeneous biological objects share *common* structural features. All, for example, macromolecules are synthesized through a common chemical reaction, the so-called condensation reaction. As noted, they have their units, but their primary sequences of them are abstractly similar to each other. It is economically preferable to have a highly *abstract* framework to embrace almost structural features of biological objects. Knowledge is commonly described regardless of individual instances of their target objects. In other words, biological knowledge is described by naming abstract classes, so-called *types*, of individually instantiated objects, *tokens*. Thus, type-level abstraction is commonplace to represent structural features irrespective of individual experimental objects.

This abstract pursuit of scientific knowledge is a popular topic in the philosophy of science. Many philosophers of science have tried to formulate the essence of scientific knowledge by concentrating on theories and laws of nature. Logic is a fundamental way to formulate the nature of scientific theories. Logical empiricists believe that an axiomatization is a typical approach to abstract scientific theories (Carnap 1953). This approach showed that a deductive framework could formalize knowledge in *physical* sciences. However, many philosophers of biology argue that this approach can be partially applicable to some theories based on mathematical

equations and that most biological objects in living systems are independent of mathematical laws and axioms (see Culp and Kitcher 1989). Recently, proponents of biological mechanisms to explain phenomena have argued that the logical approach is methodologically fruitless and explanatorily useless. So, biological knowledge cannot be logically formulated at all (see Craver 2001).

Is it also impossible to formulate molecular structures of biological objects despite the failure of the logical axiomatization of biological theories or laws? As knowledge of biological objects is distinguished from theories or laws, this paper concentrates on formalizing not biological theories or generalized laws but molecular structures of biological objects. Although many philosophers of biology showed limitations of formal approaches to biological theories, my main goal is to show how to formally interpret and understand vital structural characteristics of biological objects. In molecular biology and biochemistry, many researchers assume that successful representations of biological structures are approximately identical to the structures of the targets in nature. That is why biological knowledge in those fields is about biological objects and systems in the world. Of course, any formal approach to theories or laws across general biological areas is still meaningful. But it is more valuable to analyze biological structures in nature rather than biological knowledge in mind because the former is more fundamental than the latter in sciences. Furthermore, pursuing a formal framework to represent biological structures is still a challenging task in the philosophy of biology. For these reasons, I will suggest a switch of targets to be formalized from theories to objects. Unlike philosophers of science to formally reconstruct scientific theories or laws, I propose that biological structures are more good targets to be formalized.

This paper will focus on three constitutively structured features of individual objects: compositional hierarchy, spatial organization, and functionality. In Sect. 2, those three structural features of biological objects will be revealed in the case of DNA. In pursuit of adequate formal frameworks to attain both topics above, a methodological turn will be emphasized by suggesting that the mathematical representation of biological objects or processes is more beneficial than traditionally linguistic or axiomatic ways in Sect. 3. As logical axiomatization was problematic to formalizing theories or generalized laws, I do not adopt a conventional axiomatization dependent on logic to formalize structural features of biological objects. That is because proper tools need to help in visualizing the structural aspects of biological objects, not just refer to their names. On behalf of the axiomatic approach, I introduce a mathematical alternative to formalize biological structures, category theory. And, then critical structural features of biological objects will be formalized by category theory in Sect. 4.

I will emphasize the significance of categorical abstractions for representing molecular structures of biological objects by enumerating either some philosophical advantages or biological strategies in Sect. 5: (i) In a descriptive aspect, category theory is the more promising language to represent structures of biological objects than set theory because the former allows us to visualize internal relations among sub-objects whereas the latter does not; (ii) in a methodological aspect, category abstractions of molecular structures of biological objects as well as mathematical concepts like product and mapping relations give a mathematical understanding of

an unexplored inference, the inference from structure to functions in biology; (iii) In an explanatory aspect, categorical abstractions along with a critical concept of category theory, *invariance* or *symmetry*, provide with a clue why highly abstract mechanistic schema like the central dogma in molecular biology are explanatory in the absence of concrete spatiotemporal information of biological mechanisms; (iv) In a classificational aspect, the invariant properties of functional relations or morphisms among objects allow to understand how higher-order types of biological molecules are generalized in practice.

It is a significant fact that no biological objects solely exist in nature alone. The structural characteristics of biological things and their functional processes are determined by regular patterns with other things or within a specific biological system. Since a categorical framework provides epistemic norms to analyze diverse structures mathematically, like invariance, these norms help in figuring out the central characteristics of structures within biological systems so that it is still philosophically significant to formally interpret biological structures in nature.

2 Type-Level Common Features of Molecular Structures in Biology

Macromolecules are fundamental objects of biological mechanisms. Macromolecules include carbohydrates, proteins, nucleic acids, and fats. Each macromolecule consists of its building blocks. Sugars are in carbohydrates, amino acids in a protein, nucleosides, nucleotides in nucleic acids, and fatty acids in fats or phospholipids. All building blocks are compounds of monomers, single ("mono-") parts ("-mers") of macromolecules. Generally, macromolecules are called 'polymers' with many ("poly-") parts ("-mers"). Monomers are a molecule, so polymers are also molecules. It implies that the construction of the more complex structures of molecules may be algebraically recursive to an arbitrary operation if it exists. Among the three kinds of macromolecules, we concentrate on the case of nucleic acids such as DNA or RNA because the other two types share standard structural features of them.

Common structural features of macromolecules are realized by synthesizing monomers, so-called polymerization. Nucleic acids include nucleosides and nucleotides that play significant roles in storing genetic information and synthesizing proteins. The former is a basic type of RNA, whereas the latter is that of DNA. Most macromolecules are synthesized through a stepwise polymerization of monomers into a long chain such that the constant set of enzymes repeatedly adds the building blocks. In chemistry, the synthesis of building blocks is called a *condensation* reaction,[1] A condensation reaction is a chemical activity from monomers to polymers, referred to as *bonding*. Linkages between two monomers are types of covalent bonding mediated by biological enzymes. The linkages between nucleotides are called *phosphodiester bonds*.[2] Outputs from the condensation reactions are *backbones* of

[1] Its reverse reaction is said to be *hydrolysis* which means the reverse chemical activity among molecules, called *breaking*.

[2] The linkages between amino acids are called peptide bonds.

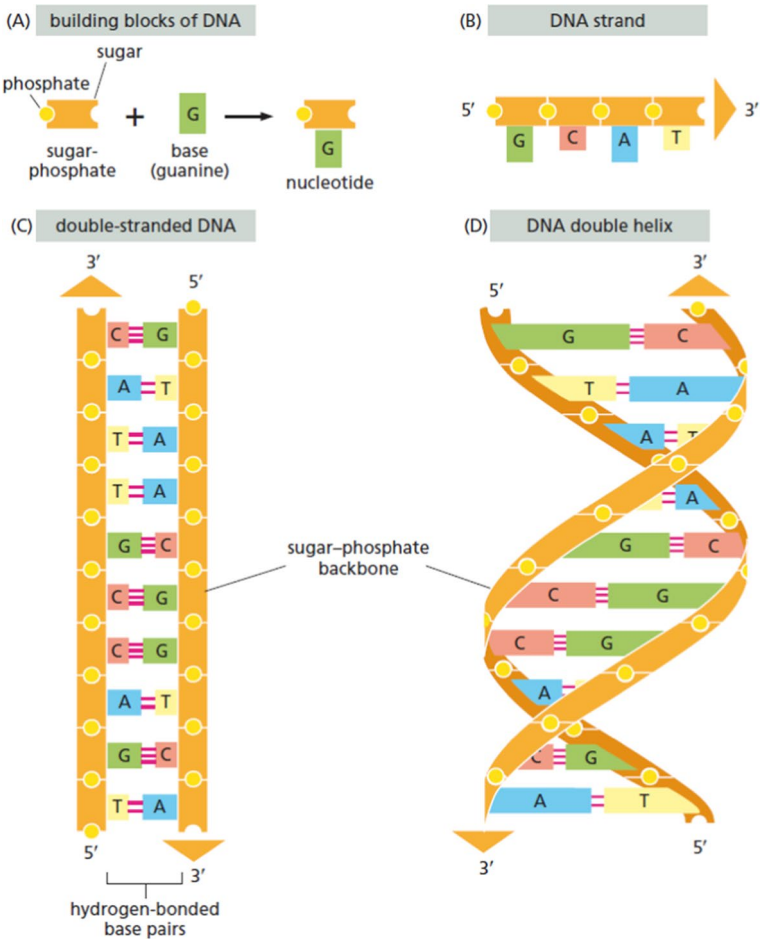

Fig. 1 DNA (Albert et al. 2014, p.173)

macromolecules. Phosphodiester bonds between the phosphate group attached to the sugar of one nucleotide and a hydroxyl group on the sugar of the next nucleotide produce the sugar-phosphate backbone of nucleic acids (see (A) and (B) in Fig. 1). A nucleotide is a compound of a nucleoside and a phosphate molecule. A nucleoside is a type of glycoside, a class of biochemical compounds consisting of a carbonic sugar attached to an amino group of a nitrogenous base through a glycosidic bond, also a type of covalent bond.[3]

[3] Sugars are of two types, ribose, and deoxyribose, within a nucleic acid. The prefix 'deoxy' indicates that the 2'-carbon atom of sugar lacks an oxygen atom linked to the 2'-carbon atom of ribose, containing both a hydrogen atom and an oxygen atom. The difference between the two sugars distinguishes the type of nucleic acid. It is well known that there are two constituents of nucleic acids, DNA-containing deoxyribose, and RNA-containing ribose. Nitrogenous bases are derivatives of two classes, purine, and pyrimidine. DNA and RNA contain two purine bases, adenine (A) and guanine (G), and two major pyrimidines.

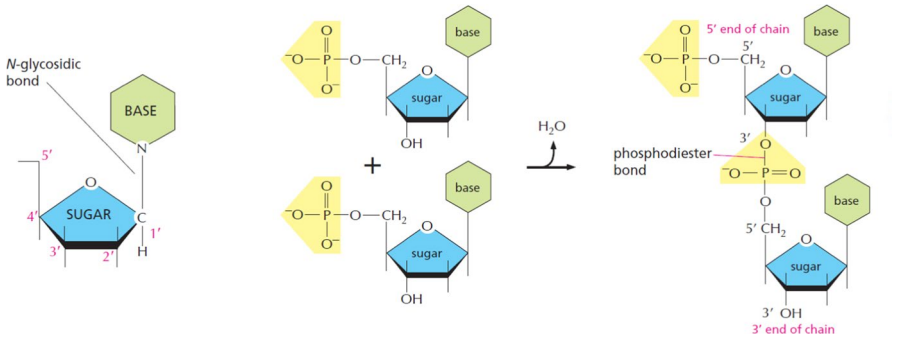

Fig. 2 Sugar-base linkage (left) and a phosphodiester bond of nucleic acids (right) (Albert et al. 2014, p.77)

To repeat, all macromolecules consist of their building blocks, so-called monomers. Most biological objects are composed of their parts. Part–the whole relation implies a hierarchical structure. Each part is regarded as a unit of the whole in the lower level. The whole is relatively considered at the higher level. *Compositional hierarchy* is a structural feature of biological objects.

For a macromolecule to be formed, each monomer must be *spatially organized*. Spatial organizations of biological molecules include orientation and shapes. Covalent bonds form primary sequences of macromolecules. Phosphodiester bonds form the sugar-phosphate backbone of nucleic acids. The bonds are shaped in particular carbon atoms in the sugar ring (left in Fig. 2). One end of a polynucleotide chain is the 5'–the end with a free phosphate group, and the other is the 3' with a free hydroxyl group. All single-stranded sequences of nucleic acids are chains from 5'–an end to 3'–end (right in Fig. 2). Each triangle arrow from 5'–an end to 3'–end in Fig. 1 indicates *oriented* polarities.[4] When two individual sequences of nucleic acids are arranged and linked together by hydrogen bonds, the two strands run antiparallel to each other in the two dimensional DNA structure ((C) in Fig. 1). The stacked-based pairs attract one another due to van der Waals interactions so that DNA becomes a double-stranded helix, which is a fundamental *shape* of DNA ((D) in Fig. 1).

Functionality is also discovered in the case of nucleic acids. An arrangement of bases among nucleotides indicates genetic information. The two nucleic acids, however, have different roles in the cell. Whereas RNA is a single-stranded polynucleotide chain, DNA is a double-stranded chain. The DNA double helix comprises two polynucleotide chains associated with each other in opposite directions by hydrogen bonds between the bases of the two chains. Because DNA is a relatively more stable

Footnote 3 (continued)
One of the pyrimidines in DNA and RNA is cytosine (C). However, the second common pyrimidine is thymine (T) in DNA or uracil (U) in RNA.

[4] Each polypeptide chain of proteins also has a structural feature, orientation from an amino (or N–terminal) end to a carboxyl (or C–terminal) end.

hydrogen-bonded structure than RNA, it acts as a piece of long-term hereditary information. On the contrary, a single-stranded RNA is usually a more temporary messenger of genetic instructions from DNA to proteins.

To summarize, the spiral shape and the orientation of DNA from 5'–the end to 3'–end are typical instances of spatial organization among biological structures. Since all macromolecules are composed of their building blocks, compositional hierarchy is revealed, too. Furthermore, biological structures give a methodological clue to infer how an object functions. In the next section, I will introduce a type-level toolkit to abstract structural features and formalize an inference from structure to function.

3 Category Theory for Abstraction of Molecular Structures of Biological Objects

3.1 Limitations of Logical Axiomatization for Formalizing Structural Features

Philosophers of science have searched for a general way to formulate scientific theories. If a universal framework applies to different fields, then the framework helps in comprehending the gist of scientific knowledge. When we confine our interest to the structural features of biological objects, how can we philosophically abstract structural knowledge in those fields?

From the early to mid-twentieth century, logical empiricists employed logic (particularly first-order predicate logic) when they articulated the structure of scientific theories, the relations between theoretical and empirical structures, and so forth. For instance, classical physics is a theoretical system consisting of Newton's three laws as definitions and deducible consequences from them by adding auxiliary conditions. This idea can still be applied to modern physics such as the theory of relativity and quantum physics. These theories in physics can be transformed into a class of axioms and several deductive consequences. Until the late twentieth century, many philosophers of physics still tried to translate successful physical theories into logically axiomatic structures (Suppes 2002; van Fraassen 1989; Costa et al. 2003).

The axiomatic approach to scientific theories has been specified in two ways. The first way to systemize scientific theories was the received view of scientific theories, suggested mainly by Carnap, who logically analyzed theories by distinguishing theoretic and observational terms (Carnap 1953). However, there were many critics of the received view in that the distinction between theoretic and observational terms needs to be more specific (see Suppe 1977, ch. 4). The second way was the semantic view of scientific theories, which was suggested by Suppes and van Fraassen, to overcome weaknesses in the received view and to articulate practical modeling processes with model theory (McKinsey et al. 1953; van Fraassen 1980).

However, these two specified ways within the axiomatic approach have been criticized by philosophers of biology since the 1980s. For example, Culp and Kitcher (1989) urged to abandon the semantic view and the received view of scientific theories because axiomatization or class-theoretic modeling is atypical of biology. First, most biologists do not just describe facts as a highly generalized form

such as a universally conditional sentence such $(x)(L_x \rightarrow D_x)$, where L_x is 'x is a living thing' and D_x is 'x will die.' Biological knowledge in molecular biology is typically expressed by graphical illustrations or diagrams rather than logical sentences. Second, even if biologists make a model to represent biological phenomena and mechanisms, they do not formalize them as a model consisting of linguistic sentences about biological objects and their activities. Instead, they show diagrammatic or pictorial expressions to be added to linguistically detailed descriptions. Recently, many philosophers of science who defend the New Mechanism emphasize the fact that typical explanations in molecular biology are mechanistic explanations about the cause to be responsible for observed phenomena, not a type of deductive argument (see Machamer et al. 2000; Bechtel and Abrahamsen 2005, 2012; Craver and Darden 2013; Glennan and Illari 2018). Third, it seems impossible for the axiomatic approach to be parallel to the methodological strategies employed in the practices of biological laboratories. According to Craver (2001), when tracing historical procedures to discover biological mechanisms, axioms about experimental apparatus or procedures have no implications on what mechanisms produce the higher-level phenomena. Instead, heterogeneous heuristics contribute to discovering novel mechanisms (see Darden 2006). In other words, the axiomatic approach fails to systemize biological knowledge and modes of explanations in biology and to capture actual practices within molecular biology or biochemistry.

Proponents of the New Mechanism tend to agree with this skeptical viewpoint that there is no formalism to systemize biological knowledge at the molecular level (see Craver 2001). However, the skepticism is due to the sterility of the linguistically axiomatized reconstruction of biological knowledge. Thus, the failures of axiomatically analyzed biological theories through the received and the semantic views never imply the absolute impossibility of any formalisms of biological sciences. Recall that when philosophers of science tried to reconstruct scientific theories in logical or mathematical ways, they mainly dealt with theories or laws within physical science which are expressible by equations. However, there are few mathematical equations in molecular biology and biochemistry rather than physics. And biological phenomena are connected to causal relations between variables or objects, whereas causal frameworks should not necessarily elucidate physical phenomena. For these reasons, the formal approach is exempt from skepticism even though the received and the semantic approaches to scientific theories are not acclaimed.

Generally, most biologists are interested in their structural features when focusing on biological things. As we saw in the previous section, structural features of biological objects must be fundamentally significant to answer the following questions; how organisms maintain, how proteins are synthesized, how genetic information is immortally stored, and so forth. Suppose there is a general framework to formulate spatial organizations, compositional hierarchy, and functionality by elucidating those questions. The logical axiomatization, however, fails to represent the structural features of biological objects.

At first, logical axiomatization is a linguistic way to express various features of objects in the world. This axiomatic method has a logical structure among axioms and theorems, the entailment relation. Each sentence within this structure denotes types of biological objects and their diverse functions in terms of names

and predicates. However, no matter how we formalize a logical structure among sentences about biological objects in molecular biology, it is still challenging to infer various features of *molecular* structures of them based on the linguistically formalized structures. Logically axiomatized structures of sentences are distinguished from spatially organized structures of biological objects. Biological objects are synthesized based on chemical reactions such as condensation reactions. Covalent bonds are essential linkages among molecules, generating spatial organizations such as directional orientation. Non-covalent bonds such as hydrogen bonds, van der Waals force, etc., make the primary sequences of polymers create secondary, tertiary, and higher dimensional structures of molecules. Within the framework of deductive logic, no logical operations exist to illustrate how different molecules chemically interact with each other, how they are spatially organized, and so forth.

Particularly, let's consider three logical operations, the material conditional ($\rightarrow$), conjunction ($\wedge$), and disjunction ($\vee$). When we are interested in formalizing chemical reactions among molecules as essential ingredients of biological structures, it seems impossible to stipulate compositional hierarchy based on part-whole relations only with logical operations conjunction ($\wedge$) and disjunction ($\vee$). It is a widely recognized common sense that the whole is more than the sum of parts. However, disjunctive connections of individual amino acids are never identical to a series of polypeptide bonds. $\phi \vee \psi$ is true unless both ϕ and ψ are false. A true disjunctive connection always allows for faulty components within the connection. But a chain of polypeptide bonds does not allow for false or absent units within the chain because false units imply unorganized structures. Further, no matter how a conjunction connects individual nucleotides, a single-stranded sequence cannot be represented because a key structural feature of the whole sequence, such as the directional orientation, is omitted.

When we agonize over how to formulate functionality logically, it will be revealed that logical axiomatization is only sometimes helpful. Logical implication ($\rightarrow$) is a remarkable candidate that seems to apply to functionality at most. However, the structure-function relationship is also not identical to the relationship between the antecedent and the consequent. The truth value of '$\phi \rightarrow \psi$' is the same as that of '$\neg\phi \vee \psi$,' but biological objects are irrelevant to this kind of truth-value calculation. Therefore, structural features are independent of logical operations and truth-value calculations. For this reason, logical axiomatization is fruitless in formulating biological structures.

We saw lengthy descriptions of nucleic acids. However, no matter how deductive logical operations can describe biological molecules, it is difficult to imagine how macromolecules are formed, how macromolecules are composed, and how macromolecules interact with other molecules. Even linguistic representations of objects are restrictively useful to illustrate them only when we have diagrammatic representations. Most biological science textbooks include many diagrammatic representations and linguistic descriptions. In other words, visualizations with diagrams are more powerful ways to represent biological structures than verbal descriptions. Logical axiomatization is a subsidiary way to formally illustrate the molecular structures of biological objects.

3.2 Preliminaries to Type-Level Abstraction of Molecular Structures

Instead of a formal linguistic method, logical axiomatization, I will introduce a *mathematical* language, category theory, to formalize type-level abstractions of structural features of biological objects. Why is category theory a suitable language? Category theory is a language for analyzing structures of mathematical objects' patterns. Of course, set theory has been widely used in mathematics until now. I admit that set theory is also a proper mathematical language when formalizing molecular structures of biological objects. However, in the next section, I will emphasize that category theory is more advantageous than set theory when formally illustrating structural features of biological objects at the molecular level. I do never say that I adopt category theory by discarding set theory. Set-theoretical expressions are helpful when discussing various categories, but I believe that category-theoretical expressions are more potent for graphically visualizing diverse structural relations among objects.

Several researchers have already considered category theory a formal tool for analyzing biological sciences. Rosen (1991) is a pioneer who suggests the first categorical framework of biological theories, the so-called (M, R) systems, where M is metabolism, and R is the repair mechanism. Louie (2009) develops Rosen's work by adopting Aristotle's four forms: the material, formal, efficient, and final causes when making relational diagrams of (M, R) systems. Independent of Rosen and Louie, Ehresmann is another biologist who applies category theory to biology (Ehresmann and Vanbremeersch 2007, 2018, 2019). Ehresmann focuses on a categorical concept, colimit, to analyze hierarchical levels by paying attention to neuroscience. Gómez-Ramirez (2014) also explores the concept of representation by applying category theory to cognitive science.

I agree with the need for category theory to formalize biological systems or levels. However, My interest in category theory differs from both Rosen and Ehresmann. I focus on structural characteristics that individual biological objects commonly share, whereas Rosen pays attention to two critical procedures in biological systems, metabolism, and repair. Mainly, I do not consent why the four Aristotelian causes are required when analyzing biological structures. I favor Ehresmann's works since she does not depend on additional frameworks, such as four kinds of causes, except category theory. As Rosen defines (M, R) systems with category theory, Ehresmann also struggles to define her term, *memory evolutive systems*, that are hierarchical and temporally successive system configurations. Of course, their new categorical systems are conceptually impressive, but I employ category theory to *represent* structural features of biological objects as type-level abstractions without defining strange terms.

Category theory is a toolkit to capture mathematical structures among different mathematical objects. I will introduce the definition of a category and some basic concepts in category theory (Awodey 2010). First, the definition of a category is as follows:

> A category $\mathcal{C}$ (i) consists of a collection of objects. (ii) For each pair of $\mathcal{C}$ -objects A and B, there exists a set $\mathcal{C}(A, B)$, which is the Hom-set of morphisms

from A to B. If $f \in \mathcal{C}(A, B)$, one may also write $f{:}A \to B$. For simplicity, or when category $\mathcal{C}$ does not need not be emphasized, Hom-set $_{\mathcal{C}}(A, B)$ is denoted as $Hom_{\mathcal{C}}(A, B)$. (iii) For any three $\mathcal{C}$-objects A, B, and C, a mapping $g \circ f : \mathcal{C}(A, B) \times \mathcal{C}(B, C) \to \mathcal{C}(A, C)$, taking f: $A \to B$ and g: $B \to C$, exists. These entities satisfy the following two axioms: (a) As the *associativity*, if f: $A \to B$ and g: $B \to C$, h: $C \to D$, such that both $h \circ (g \circ f)$ and $(h \circ g) \circ f$ are defined, then $h \circ (g \circ f)$ and $(h \circ g) \circ f$. (b) As the *identity*, for each object A, there exists 1_A: $A \to A$, such that, for any f: $A \to B$, g: $C \to A$, one has $f \circ 1_A = f$, $1_A \circ g = g$. 1_A, which is demonstrably unique, is called the identity morphism on A.

In this definition of categories, the associativity among morphisms is a chief property of categories. Category theory is a language about morphisms or arrows from one object to another. By specifying morphisms between objects, it demonstrates the structural relationships between things. This property is generally called the compositional rule of categories, and will be applied to the case that the longer macromolecule is synthesized orderly through condensation reactions.

Second, a preorder relation on a set X is one sort of relationship on X, denoted with '$\leq$' in category theory. The set X becomes any object of the preorder relation. The most significant relationship between preordered sets is a monotone morphism. This morphism preserves preorder relations, a so-called structure-preserving map for preorders, because it is an injective morphism. For example, a monotone morphism $X \to Y$ between two preorders connects each element of preorder X to an element for the preorder Y. If the monotone morphism is a function, f, then $f(x)$ is identical to elements of Y, y, where $x \in X$ and $y \in Y$. This invariant property of relationship is useful when representing spatial organizations of biological objects.

Third, colimits are a useful concept to represent a compositional hierarchy. They are the dual concept of limit. They are universal constructions representing various behaviors in a category (Awodey 2010). A limit is as follows:

Given any diagram such $\phi_{AB} : A \to B$ in a category $\mathcal{A}$, where A and B are arbitrary objects of the category, a limit (or a cone) as an object L over a diagram F is a choice of morphisms $\eta_X : L \to X$ from L to each object X in the diagram F, such that all newly formed triangles commute satisfying $\eta_B = \phi_{AB} \circ \eta_A$. Moreover, for any object Z and for any collection of morphisms $\alpha_X : Z \to F$ satisfying $\alpha_B = \phi_{AB} \circ \alpha_A$, there exists a unique morphism $f : Z \to L$ such that $\alpha_X = \eta_X \circ f$ for all objects X in the diagram F.

A colimit is a dual concept of a limit. In other words, a colimit is an object C with reversed arrows from each object X in a diagram D to the C. A colimit is useful to represent inter-connective information among all objects within a diagram.

Give any diagram such $\theta_{AB} : A \to B$ in a category $\mathcal{A}$, a colimit C (or a cocone) as an object C under a diagram D is a choice of morphism $c_X : X \to C$ from each object X in the diagram D to C, such that all newly formed triangles are commuted satisfying $c_A = c_B \circ \theta_{AB}$. Moreover, for any object Z and any collection of morphisms $\beta_X : X \to Z$ satisfying $\beta_A = \beta_B \circ \theta_{AB}$, a unique morphism g: $C \to Z$ exists, such that $\beta_X = g \circ c_X$ for all objects X in the diagram D.

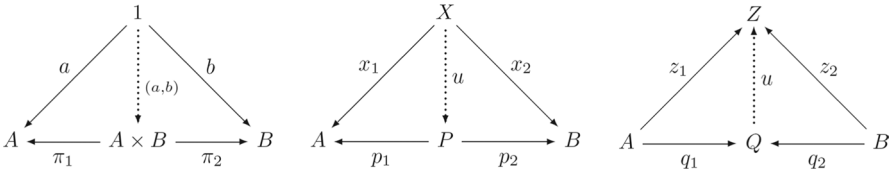

Fig. 3 A product from 1 (left), a product from X (middle), and a coproduct Q (right) (Awodey 2010)

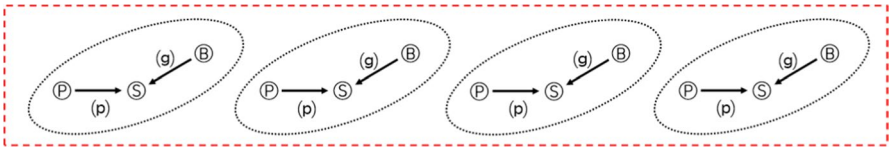

Fig. 4 Individual nucleotides

Finally, a product, which is a sort of limits, will be used to represent the functionality of macromolecules. A product is also the dual concept of a coproduct. Both product and coproduct are the most typical examples of new constructions of category from the old. A product is as follows:

> In any category $\mathcal{C}$, a product diagram for the objects A and B consists of an object P and arrows $p_1 : P \to A$ and $p_2 : P \to B$ satisfying the following properties such that: given any diagram of the form $x_1 : X \to A$ and $x_2 : X \to B$, there exists a unique $u : X \to P$, making the diagram in the above to be commuted such that $x_1 = p_1 \circ u$ and $x_2 = p_2 \circ u$.

Concretely speaking, given sets A and B, the Cartesian product of A and B is the set of ordered pairs, $A \times B = \{(a, b) \mid a \in A, b \in B\}$. Note that there are two coordinate projections $p_1(a, b) = a$, $p_2(a, b) = b$ from $A \times B$ into A or B, that is given any element $c \in A \times B$ we have $c \in (p_1 c, p_2 c)$. This situation is captured precisely in the following two commutative diagrams. One is proved from an element, 1, and the other is proved from generalized elements, X, in Fig. 3 (Awodey 2010). Coproduct is a dual construction to the product of A and B.

> In any category $\mathcal{C}$, a coproduct diagram for the objects A and B consists of an object Q and arrows $q_1 : A \to Q$ and $q_2 : B \to Q$ satisfying the following properties such that: given any diagram of the form $z_1 : A \to Z$ and $z_2 : B \to Z$, there exists a unique $u : Q \to Z$, making the diagram in the above to be commuted such that $z_1 = u \circ q_1$ and $z_2 = u \circ q_2$.

Given sets A and B, a coproduct of A and B is exactly their product in the opposite category. That is the disjoint union $A + B = \{(a, b) \mid a \in A, b \in B\}$. We generally write $q_1 : A \to (A + B)$ and $q_2 : B \to (A + B)$, which are called injections or coprojections for the coproduct. Those two morphisms q_1, q_2 consist of dual-product of A and B, there exists a unique $u : Q \to Z$ with $z_1 = u \circ q_1$ and $z_2 = u \circ q_2$ if for any Z and $z_1 : A \to Z$ and $z_2 : B \to Z$. An example of the coproduct in

sets is $A + B$ of two sets in their disjoint union, which can be constructed such as $A + B = \{(a, 1) \mid a \in A\} \cup \{(b, 2) \mid b \in B\}$ with evident coproduct injections $i_1(a) = (a, 1), i_2(b) = (b, 2)$.

Those categorical concepts will be used when we formally visualize biological structures later. The next section will formally illustrate three structural features in molecular biology and biochemistry.

4 Abstracting Molecular Structures of Biological Objects

4.1 Compositional Formation of Fundamental Units

How can spatially organized structures be created? To figure out the directional structures of the primary sequence of macromolecules, we need to think of chemical linkages among their units of them. Let me explicate this question by focusing on nucleic acids in detail. As noted in the previous section, biological structures are determined by chemical reactions and bondings. I suggest dividing biological structures into two sub-cases; (i) *intra*-structures of units, and (ii) *inter*-structures of units of macromolecules. Let's discuss the former first.

Nucleosides[5] consist of two subunits, sugars (represented by a circled 'S') and bases (represented by a circled 'B') within an aqueous solution (see Fig. 4). A nucleoside is formed if and only if a 1' carbon hydroxyl group within each sugar, including ribose and deoxyribose, is linked to either a 9' nitrogen amino group within a base of purines or a 1' nitrogen amino group within a base of pyrimidines. The linkage between a sugar and a base, called a glycosidic bond ((g)), is a covalent bond in which two discrete molecules share one or more electrons when a specific enzyme mediates these two molecules into a compound. A covalent bond can be realized through a condensation reaction from discrete molecules to a compound of such molecules. Unlike set theory, category theory provides morphisms to represent chemical linkages.[6]

Nucleotides are also essential in the second step of DNA formation because they are units used to convey individual genetic information and sustain the DNA's structure. We can easily define a nucleotide by adding a chemical bonding to a nucleoside, a phosphodiester bond ((p)) from a monophosphate group (represented by a

[5] Nucleosides are important because they individually indicate genetic information within DNA (and RNA). In particular, four bases within nucleosides determine the type of genetic information. However, individual nucleosides should be arranged within a certain frame or backbone to stabilize DNA and play a role in normal metabolic processes, such as DNA replication or transcription in protein synthesis. Nucleotides contribute to this requirement by attaching a phosphate to individual nucleosides. Therefore, nucleotides are new compounds at a higher level than nucleosides because of the additional bond by which a phosphate (represented by a circled 'P' in Fig. 4) links to a nucleoside.

[6] For individual nucleotides, depicted by the dashed lines Fig. 4, to be categories by definition, an individual phosphate molecule, a sugar molecule, and a base molecule must have their identity morphism. Those identity morphisms can be assigned by considering the chemical interactions of the three molecules with water in the condition of the solution. However, those precise descriptions are omitted here for simple illustrations.

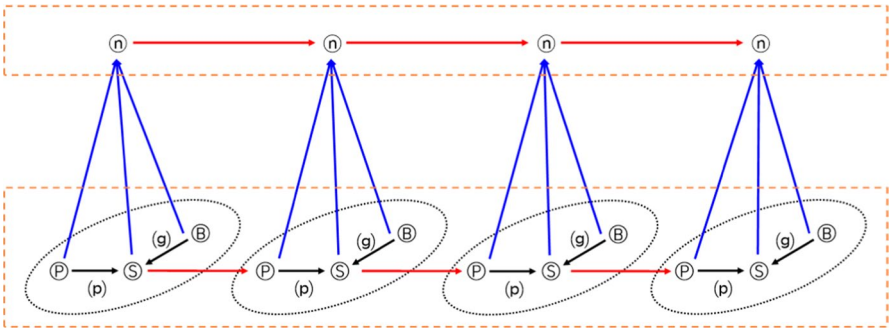

Fig. 5 The formation of a single strand of DNA

circled 'P') to the C-5' of sugar within nucleosides (see Fig. 4). The codomain of a phosphodiester bond is either a ribose in the case of RNA or a deoxyribose in the case of DNA. Individual nucleotides of RNA are adenosine, guanosine, cytidine, and uridine, and those of DNA are deoxyadenosine, deoxyguanosine, deoxycytidine, and deoxythymidine.

To summarize, phosphodiester bonding from a phosphate group to a sugar molecule and glycosidic bonding from a nitrogen base to the sugar molecule are basic intra-structure of nucleotides. Note that an oriented property of a single-stranded sequence of DNA has yet to be realized because inter-structures of nucleotides are not specified. If we verify chemical linkages among nucleotides, we can define the primary sequence of DNA as a category. In the following discussion, if we assign ordered natural numbers to individual subunits within a nucleotide based on the order to be synthesized, then we know the oriented order of the backbone.

4.2 Spatial Organization 1: Orientation

To capture the orientational (or directional) feature of the sequential structure of biological macromolecules, *a partially ordered set* (often abbreviated as poset) is introduced. A poset relation on a set X is a binary relation on X, denoted with an infix notation $\leq$, such that (i) $x \leq x$ and (ii) if $x \leq y$ and $y \leq z$, thus $x \leq z$. The first condition is called reflexivity, and the second is transitivity. If $x \leq y$ and $y \leq x$, we write $x \cong y$ and state that x and y are equivalent. We call a pair $(X, \leq)$ consisting of a set equipped with a preorder relation a preorder (see Louie 2009). One of the typical examples of a preorder is the natural number with the order given by a usually sized ordering such as $0 \leq 1$, $5 \leq 100$, and so on. Thus, if $x \leq y$, then x is considered less than or equal to y, and if y is greater than or equal to x, then it is written as $y \geq x$ (Awodey 2010). This asymmetrically ordered property corresponds to the directional linkage of two backbones of DNA because each strand is linearly formed from C-5' of sugar within nucleic acid to the C-3' hydroxyl group of sugar within another nucleic acid such that a strand has an ordered configuration of nucleic acids. If we assign four types of nucleotides as elements in a set, then a poset abstracts the

oriented sequential structure of DNA. Proteins are also satisfied with the same structural features of poset-like DNA.

As noted, single-stranded polynucleotides are synthesized by condensation reactions. Through these reactions, individual nucleotides are associated with phosphodiester bonds (red arrows in Fig. 5) *from* a monophosphate attached to C-5' of sugar within a nucleotide *to* C-3' hydroxyl group of sugar within another nucleotide. Interestingly, the chemical reactions to constructing a DNA backbone are the same as that used to construct the linkage between phosphate and C-5' of sugar within a nucleoside. Phosphodiester bonding is an essential linkage of intra- and inter-structures within nucleic acids. Notice that when phosphodiester bonding becomes the infrastructure of nucleotides, the backbone of the single chain of DNA can be formed. Similarly, peptide bonding is also an inter-structure of amino acids because it forms the primary sequence of proteins. Orange dotted lines in Fig. 5 indicate the backbone of DNA.

When a single polymer is synthesized, the polymer can become a category because individual subunits within the polymer satisfy the fundamental properties of categories, associativity as a compositional rule. In the case of polynucleotides, both phosphate groups and sugar molecules are associated by compositional rules. Recall that ordered natural numbers can represent the oriented order of the backbone (or the primary sequences of subunits). Assuming that the left nucleotide in Fig. 5 is the first ordered subunit, we have a sequential chain among four nucleotides. In the first ordered nucleotide, a phosphate group (P_1) is linked to the first sugar molecule (S_1) by an intra-phosphodiester bond ((p)) so that it can be notated by $S_1 = p(P_1)$, where a subscript number means the order. Assuming that inter-phosphodiester bonding is expressed by a bolded Capital $\mathbf{P}$ instead of p, the second ordered phosphate group can be linked to the first ordered sugar molecule like $S_2 = p(\mathbf{P}(S_1)) = p(\mathbf{P}(p(P_1)))$, which means that two individual nucleotides are synthesized. Additionally, when the third-ordered phosphate group is linked to the second-ordered sugar molecules, $S_3 = p(\mathbf{P}(S_2)) = p(\mathbf{P}(p(P_2))) = p(\mathbf{P}(p(\mathbf{P}(S_1)))) = p(\mathbf{P}(p(\mathbf{P}(p(P_1)))))$. We can formally represent a directional chain of polynucleotides by doing so repeatedly. Here, associativity among morphisms constructs a category that consists of two objects, such as phosphate groups (P_k) and sugar molecules (S_k), where k is an order of the natural number, and of two morphisms such intra- (p) and inter-structured ($\mathbf{P}$) phosphodiester bondings.

4.3 Compositional Hierarchy

We implicitly assume that a nucleotide consists of three internal sub-molecules, a phosphate group, a sugar molecule, and a base. That means that the three sub-molecules are components of a nucleotide. Furthermore, we knew that chemical bondings spatially organize the three sub-molecules. If all chemical ingredients of a single strand are collected in ordinary surroundings, then poly nucleic acid can be synthesized. Based on the above definitions of chemical units of DNA, a single strand can also be formally represented by the concept of colimit. Simply speaking, a colimit is an assembly or gluing of objects together. A colimit includes a simple coproduct,

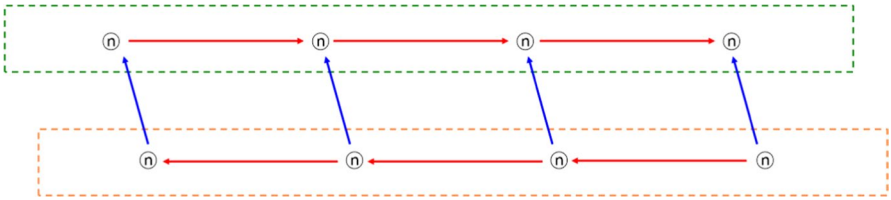

Fig. 6 Two strands of DNA at the middle level

a set of elements, similar to the sum or disjoint union in set theory. If we deal with a single strand of DNA as a coproduct, the colimit of the single strand includes only four types of nucleotides. If the entire single strand of DNA is a coproduct, we can focus solely on elements and nucleotides. However, we must include organized chemical linkages among the nucleotides, phosphodiester bonds, and glycosidic bonds. A simple sum of units cannot abstract the chemical bonding and directional orientation. For this reason, the term colimit is used at the higher level to represent both sub-component parts and their organized structures at the lower levels (see Fig. 5). In short, a colimit in category theory is qualified as the whole, including the component parts. It contains sub-components and their organized activities. Thus, each nucleotide (n) is a colimit of the three sub-molecules that are chemically linked together. Consequently, we acquire a type-level abstraction of the compositional hierarchy relating to a chain among nucleotides based on their lower-level structure of them.

As seen in Fig. 5, individual nucleotides (n_k) at the higher level are orderly linked to each other by inter-phosphodiester bond. Simultaneously, internal connections between a phosphate group and a sugar molecule ($S_k = p(\mathbf{P}(P_{k-1}))$) are also orderly linked to the following internal connection ($S_{k+1} = p(\mathbf{P}(P_k))$) at the lower level. Although the higher-level category of nucleotides includes different objects from the lower-level category of units of them, the morphism, phosphodiester bond ($\mathbf{P}$), is commonly *invariant* between them. The property of $\mathbf{P}$, the ordered orientation, is maintained in both levels. Thus, we formally stipulate the compositional hierarchy by saying that a single strand of nucleotides comprises its subunits and that the strand is a higher object than its subunits.

4.4 Spatial Organization 2: Shape

At the highest level, DNA is a double-stranded helix where each strand within the DNA is complementarily paired through hydrogen bonds. For example, each strand is a unit of DNA as a component. Fig. 6 shows two strands (green and orange dotted lines) associated together through hydrogen bonds (blue arrows) at the middle level. At the middle level of DNA, a strand is shaped as a chain of nucleic acids consisting of nucleotides. A chain of nucleic acids is formed by polymerizing each nucleotide as a monomer. Again, at the lower level of nucleic acids, a nucleotide is a compound in which a nucleoside is bound to phosphate. A nucleoside consisting

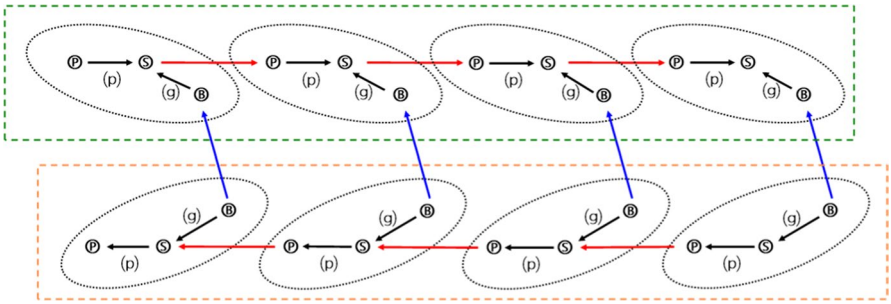

Fig. 7 Two strands of DNA at the lowest level

of sugar and bases through a covalent bond is hierarchically an essential subunit of DNA. When we consider the hierarchical aspects of DNA based on the category theory, examining the molecular structure of a nucleoside is the starting point. The diagram is like Fig. 7, which shows how two strands of nucleotides are composed of their units by revealing the compositional hierarchy of DNA molecules at the lowest level. A monotone morphism between two different pre-ordered categories is a type-level abstraction of hydrogen bonds. Consequently, each level of DNA is structurally preserved because two inter-structured chemical linkages, including phosphodiester bonds and hydrogen bonds, are invariant based on monotone morphism.

Note that categorical diagrams such as Fig. 7 at the middle level and Fig. 6 at the highest level are structurally identical to the double helical shape of DNA illustrated by two diagrams such as (C) and (D) in Fig. 1. It means that categorical formalism helps in demonstrating the visual shapes of biological molecules. In the diagrams, individual instances of kinds of nucleotides are ignored, and types such as 'circle B' or 'round bracket g', (g) in two Figs. 6 and 7 replace with the individual tokens. Hence, categorical diagrams provide a graphic toolkit to formally abstract biological molecules at the type level.

5 Advantages of Categorical Representations of Biological Structures

5.1 An Abstract Toolkit for Representing Structures

The previous section revealed how structural features of biological objects could be represented through category theory by focusing on a case of nucleic acids. Contrary to axiomatic methods, categorical abstractions as a mathematical method allow us to vividly represent the type-level molecular structures of biological objects. However, traditionally, set theory has been a predominant language in mathematics. Why would we pursue categorical abstractions of biological objects? Furthermore, why should we choose to map the structural features of biological objects into mathematical structures in category theory rather than set theory? To answer those questions, a distinction between a term *objects* (or *entities*) and a term *structures* will be mainly discussed at first.

Biological macromolecules such as nucleic acids, lipid acids, and amino acids are molecular objects (or entities) because they are spatiotemporally individualized. Set theory is a widely used way to indicate *part-whole* relations among objects *across* levels. A set of macromolecules generally includes three types of acids like {nucleic acids, lipid acids, amino acids}. All elements of this set share standard structural features, as noted in the previous section. Furthermore, individual objects include their sub-objects. A nucleic acid molecule, for example, consists of {phosphates, sugars, bases}. An amino acid molecule consists of {amino acid, carboxylic acid, α -carbon, side-chain groups}.

A sharp difference between objects and structures is that the latter requires one more ingredient than the former, that is, *connective* relations among objects *within* a level. An object includes its sub-objects (or parts). The connective relations include chemical bondings, such as covalent and hydrogen bonds. A structure of a nucleic acid molecule is composed of two set-theoretical expressions. In the case of nucleic acids, one is a set including sub-objects of it, {phosphates, sugars, bases}, and the other is another set indicating connective relations among elements of the set, { phosphodiester bonds, glycosidic bonds, hydrogen bonds}. As noted before, phosphodiester bonds link a phosphate molecule to a sugar molecule. Glycosidic bonds associate the sugar molecule with a nitrogenous base. Hydrogen bonds connect to two bases of the individually separated strands based on Chargaff's rule. A structural feature of nucleic acid molecules, orientation, depends on two relations: phosphodiester and glycosidic bonds. Another structural feature of them a ladder shape also depends on three relations, phosphodiester bonds, glycosidic bonds, and hydrogen bonds. An understanding of the structures of biological objects needs connective relations and compositional part-whole relations.

I concede that set theory is partially applicable to reveal the structural features of biological objects. Above, I employ set-theoretical expressions to indicate not only *inter*-level part-whole relations but also *intra*-level connective relations. However, each relation is demonstrated separately as a different set. Also, it may not be easy to graphically figure out the structural features of biological objects with set-theoretical expressions. Based on set-theoretical information about biological objects, imagining orientation, compositional hierarchy, and shape is difficult. As shown in the previous section, category-theoretical expressions provide an integrated way to express both types of relations simultaneously. That is, elements about molecular connections as well as sub-objects of macromolecules can be represented graphically (see Figs. 5, 6, and 7 again.). For example, nucleic acids as a category can be defined in terms of compositional hierarchy from nucleoside to nucleotide and nucleic acids as follows.

- A category of nucleoside

1. objects: {a class of sugars, a class of bases}
   (a) a class of sugars = {deoxyribose, ribose}
   (b) a class of bases = {a class of purines, a class of pyrimidines}
   (c) a class of purines = {adenine, guanine}
   (d) a class of pyrimidines = {cytosine, thymine, uracil}

2. arrows: morphisms from a class of sugars to a class of bases (phosphodiester bond)
    (a) a morphism from a ribose to a class of bases
    (b) a morphism from deoxyribose to a class of bases

Based on two components of a category of nucleoside, two classes of nucleosides can be abstracted as follows: one is a class for RNA, and the other is a class for DNA. Next, if a condensation reaction synthesizes nucleosides, then a category of nucleotides is formed like Fig. 4.

- A category of nucleotide

1. objects: {a monophosphate, a class of nucleosides}
2. arrows: morphisms from a monophosphate to a class of sugars (phosphodiester bond)
    (a) a morphism from a monophosphate to a ribose
    (b) a morphism from a monophosphate to a deoxyribose

Based on two components of a category of nucleotide, two classes of nucleotides can be abstracted as follows: one is a class for RNA such as {adenosine, guanosine, cytidine, uridine}, and the other is a class for DNA such as {deoxy-adenosine, deoxy-guanosine, deoxy-cytidine, deoxy-thymidine}. Next, if a condensation reaction synthesizes nucleotides, then a category of nucleic acids as a single strand is formed like Fig. 5.

- A category of nucleic acids

1. objects: {nucleotides}
2. arrows: morphisms from a monophosphate to a class of sugars (phosphodiester bond)
    (a) a morphism from a monophosphate to a ribose
    (b) a morphism from a monophosphate to a deoxyribose

Based on two components of a category of nucleic acids, a single strand of DNA and RNA are abstracted. Mainly, if a single strand of DNA is linked to another strand of DNA under Chargaff's rule, then DNA is synthesized like Figs. 6 and 7.

- A category of DNA

1. objects: {a single sequence of nucleic acids}
2. arrows: morphisms from a base to another base (hydrogen bond)

Note that Fig. 7 not only reveal connective relations among sub-objects but also indicate the two-dimensional shape of DNA. If we adopt category theory as a toolkit to represent their structural features, this categorical framework provides a helpful

way to visualize those features. The discovery of the double helical shape of nucleic acids was one of the significant achievements in molecular biology. Of course, set-theoretical descriptions of the nucleic acids are necessary to figure them out. But, category-theoretical descriptions are a graphical way to imagine structures directly.

5.2 A Mathematical Understanding of a Heuristic Inference from Structure to Function

How can biologists infer the functional properties of biological objects from their structural features? The inference of functions from molecular structures is widely acknowledged in molecular biology as a heuristic method for discoveries. This inference is not deductive because no deductive rules exist in logic to derive sentences about biological functions from sentences about molecular structures. Traditionally, logical empiricists had no interest in methodological procedures for scientific discoveries. For example, Hans Reichenbach argued that the context of discovery must be distinguished from the context of justification. The former context is subjectively psychological. Philosophical investigations must concentrate on how hypotheses are justified based on empirical evidence. Induction is another fundamental method to find biological knowledge based on repeatedly observed patterns. However, no matter how we collect empirical facts about the structural features of biological objects, further rules seem to request to infer their functional properties of them. That is because structural facts are ontologically distinguished from functional facts. Furthermore, in the absence of theoretical templates in molecular biology, such as Newton's mechanics in physics or Darwin's principles of natural selection in evolutionary theory, the inter-relationship between structure and function has been justly regarded as a heuristic strategy without further methodological elucidations. I will show that categorical abstractions of molecular structures of biological objects are necessary for the mathematical understanding of how functional properties of biological objects can be inferred from their structural features. For this, we will deal with a historically typical case: Watson and Crick's prediction about genetic replication.[7]

Since Watson and Crick published their first paper in April of 1953, they could confirm that their model was consistent with X-ray data from Rosalind Franklin.

[7] The structure proposed by Watson and Crick has two properties of central importance to the role of DNA as the hereditary material (Stryer et al. 2017). First, the structure is compatible with any sequence of bases. While the bases are distinct in structure, the base pairs have essentially the same shape and, thus, fit equally well into the center of the double-helical structure of any sequence. Without any constraints, the sequence of bases along a DNA strand can efficiently store information. Indeed, the sequence of bases along DNA strands is how genetic information is stored. The DNA sequence determines the sequences of the ribonucleic acid (RNA) and protein molecules that carry out most of the activities within cells. Second, because of base pairing, the sequence of bases along one strand ultimately determines the sequence along the other strand. As Watson and Crick so shortly wrote: "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material" (Watson and Crick 1953a, p.737). Thus, if the DNA double helix is separated into two single strands, each strand can act as a template for generating its partner strand through the specific base-pair formation. The three-dimensional structure of DNA beautifully illustrates the close connection between molecular structure and function.

After that, they published an additional paper to discuss the genetic implications of their model. In their second paper, they said:

> "Until now, no evidence has been presented to show how DNA might carry out the essential operation required of genetic material, that of exact self-duplication. We have recently proposed a structure for the salt of deoxyribonucleic acid, which, if correct, *immediately* suggests a mechanism for its self-duplication. (...) Though the structure will not be wholly proved until a more extensive comparison has been made with the X-ray data, we now feel sufficient confidence in its general correctness to discuss its genetical implications" (Watson and Crick 1953b, pp.964-965, *emphasis added*).

In the above quotation, Watson and Crick said again that a model of DNA replication mechanism could be "immediately" inferred from the structure of DNA. They emphasized that their suggestion was supported by X-ray data supplied by Wilkins at King's College and that the data contradicted Pauling's triple model. At the time, there was no competing model of DNA structure being supported by empirical data. This seemed to boost their confidence to publish genetic implications from their discovery (see Watson 2001; Crick 1988).

How could Watson and Crick infer from the double helical structure of DNA as genetic material to replication function? Interestingly, Watson and Crick divided the chemical linkages of DNA into regular and irregular parts. They referred to the linkage between deoxyribose and a phosphate group as a *regular* part, while the sequence of bases along the chain was considered an *irregular* part.

> "If the sequence of bases on one chain is *irregular*, it is difficult to explain these analytical results except by the sort of pairing we have suggested. The phosphate–sugar backbone of our model is completely *regular*, but any sequence of the pairs of bases can fit into the structure. It follows that in a long molecule many different permutations are possible, and it therefore seems likely that the precise sequence of the bases is the code which carries the genetical information" (Watson and Crick 1953b, pp.965-966, *emphases added*).

Based on their discovery of the DNA structure, they separated the part storing genetic information from the part sustaining the structure itself. Watson and Crick also said, "the first feature of our structure which is of biological interest is that it consists not of one chain, but of *two*" (Watson and Crick 1953b, p.965, *emphasis added)*). Fig. 8, shown in Watson and Crick's second paper in 1953, indicates their dimensional analysis of the DNA structure. This dimensional classification of DNA structure depends on the spatial organizations of compositional sub-objects. It stems from the difference between the two linkages, a collection of all morphisms within a category of nucleic acids and that of all morphisms within a category of DNA. The former is a morphism between sugars and phosphates for forming a linear nucleic acid chain, and the domain and codomain are homogenous. Thus, there is no possibility of storing various genetic information. On the contrary, the latter is a morphism between bases within each nucleotide

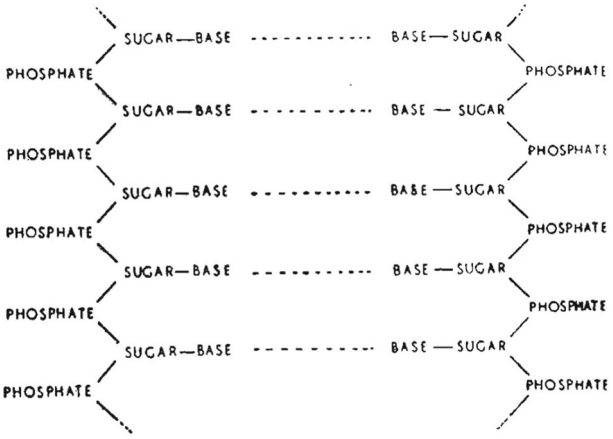

Fig. 8 Watson and Crick's dimensional interpretation of DNA (Watson and Crick 1953b, p.965)

to represent hydrogen bonds; as we know, there is no constraint that only one kind of base among purines and pyrimidines should be placed within a chain. A collection of all morphisms within a category of DNA to represent the hydrogen bonds allows four bases to be arranged freely in a chain. Then, the configuration of bases can be regarded as part of genetic information. Watson and Crick also said, "the other biologically important feature is how the two chains are held together. This is done by hydrogen bonds between the bases, as shown schematically", which is revealed in this paper at Fig. 8 (Watson and Crick 1953b, p.965). In other words, there are one-to-one mappings from the two sorts of linkages of DNA structure to two dimensions. One is a mapping *from* the covalent linkage synthesized by phosphodiester bonds between phosphate groups and deoxyriboses (or sugars) *to* the regular backbone part of DNA. The other is a mapping *from* the hydrogen linkage supported by glycosidic bonds between deoxyriboses and bases *to* the irregular part of DNA. This interpretation from structural linkages to two dimensions can be formalized in terms of products such as Fig. 9.

Based on the mathematical interpretation of the molecular structure of DNA as a product in a categorical abstraction, we understand how Watson and Crick were able to conclude that the DNA structure is a unique genetic material formally. The DNA structure consisting of two dimensions, regular backbones and irregular bases, can be decomposed into two elements, phosphate–sugar linkage and sugar-base linkage. There are two projection morphisms, π_1, and π_2, from the DNA structure to each element, respectively. As I discussed above, the phosphate–sugar linkage represents not only *backbone* (*B*) of DNA but also the size of genetic information. Hence, a morphism σ_1 from a gene to the phosphate–sugar linkage exists. Simultaneously, the sugar-base linkage represents a *genetic* information (*I*) which can be indicated by a morphism σ_2 from a gene to the sugar-base linkage. As a consequence, from the universal mapping property based on associativity, the DNA structure can be a gene since there

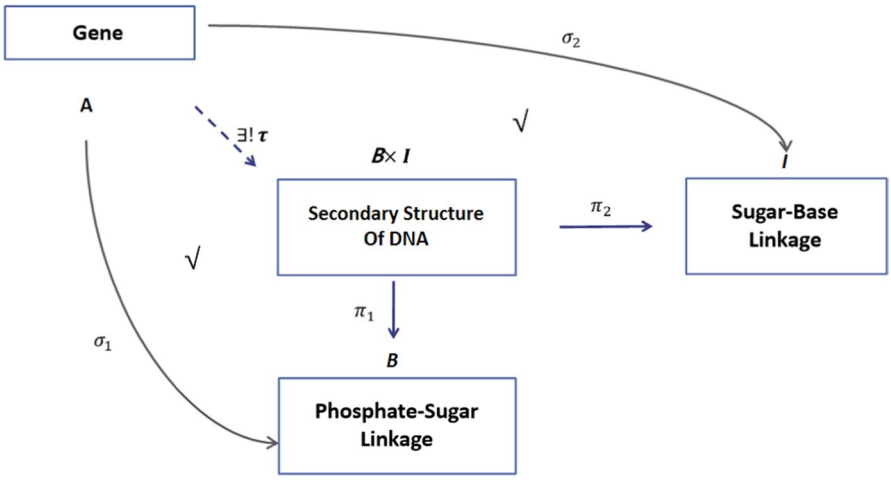

Fig. 9 Two strands of DNA

is a unique linear map from the gene into $B \times I$ for which $\sigma_1 = \pi_1 \circ \tau$ and $\sigma_2 = \pi_2 \circ \tau$. Fig. 9 represents this formal inference diagrammatically, and the following definition of the Cartesian product to represent the DNA structure helps in understanding the inference from the molecular structure of DNA to two dimensions of genes.

Consequently, DNA becomes a genetic material because of the following algebraic reasons: The linear structure of a chain is a product $B \times I$, the phosphate–sugar linkage (B) as the backbone of the chain and the sugar-based linkage (I) as the types of genetic information, where $B \times I = \{(b, i) \mid b \in B, i \in I\}$, and there are two projection maps $\pi_1 : B \times I \to B$ and $\pi_2 : B \times I \to I$ by $\pi_1 : b \times i \to b$ and $\pi_2 : b \times i \to i$. Besides, there are two maps σ_1 : gene $\to$ backbone as *length* or *size* of genetic information and σ_2 : gene $\to$ information as the *types* of genetic information. Therefore, there is a unique linear map τ : gene $\to B \times I$ for which $\sigma_1 = \pi_1 \circ \tau$ and $\sigma_2 = \pi_2 \circ \tau$.

Until now, we have seen how Watson and Crick interpreted the double helical structure of DNA as genetic material. They regarded the regular phosphate-sugar linkage as the backbone of genes and the irregular sugar-base linkage as the genetic information of genes. Then, how could the functional roles of DNA be inferred from the two dimensions of genes? As they mapped two chemical linkages of DNA structure into two dimensions of genes by one-to-one correspondence, they also mapped the two dimensions of genes into two functions. They said,

> "Now our model for deoxyribonucleic acid is, in effect, a *pair* of templates, each of which is complementary to the other. We imagine that prior to duplication the hydrogen bonds are broken, and the two chains unwind and separate. Each chain then acts as a template for the formation on to itself of a new companion chain, so that eventually we shall have *two* pairs of chains, where we only had one before. Moreover, the sequence of the pairs of bases will have been duplicated exactly" (Watson and Crick 1953b, p.965, *emphases original*).

Table 1 Mapping from Structures to Functions

Molecular structures of DNA	→	Dimensions of genes	→	Functions
Phosphate-sugar linkage	→	Backbone	→	Template
Sugar-base linkage	→	Genetic information	→	Polymerization

According to this quotation, we can firmly notice that Watson and Crick mapped the two backbones of genes as templates for replication. For each strand of DNA to become a template, Watson and Crick pointed out a prerequisite that hydrogen bonds rather than phosphodiester bonds must be cleaved. Then, how could a new sequence be formed from a template? Watson and Crick said,

> "We imagine that free nucleotides are available in quantity. From time to time the base of a free nucleotide will join up by hydrogen bonds to one of the bases on the chain already formed. We now postulate that the polymerization of these monomers to form a new chain is only possible if the resulting chain can form the proposed structure" (Watson and Crick 1953b, p.965).

Watson and Crick suggested a procedure to form a new sequence on the basis of a well-already established chemical reaction, polymerization, which is a common feature of macromolecules. Genetic information can be determined by attaching free nucleotides to a template. Hydrogen bonds are relatively weaker than covalent bonds, including phosphodiester bonds of the backbones of genes, so diverse genetic sequences can be formed through polymerization (Table 1).

The table gives us conceptual correspondences from the molecular structure of DNA to the dimensions of genes and the biological functions of replication. Those mappings are qualitatively transitive. Phosphate-sugar linkage is the backbone of genes, which functions as a replication template, so the phosphate-sugar linkage is thus the template. Sugar-base linkage is the genetic information, which is synthesized by polymerization, so then sugar-base linkage becomes a basic unit of polymerization as free nucleotides. In category theory, mapping is a fundamental property when analyzing mathematical structure case of the discovery of DNA replication; mapping from molecular structure to functional roles provides us with a mathematical ground to understand how biological functions can be inferred from molecular structural features. When type-level categorical abstractions formalize structural features of biological objects, those kinds of qualitative mappings could be revealed. Without categorical abstractions, we do not have any formal ways to capture the connection between structures and biological functions. But, more than categorical abstractions are required to understand the connection completely. Cartesian product is also an additional concept for understanding the dimensional features of genes in the middle of the structure of DNA and genetic replication. Consequently, the categorical framework gives a clue to mathematically understand a critical inference in molecular biology without regard to any logical rules.

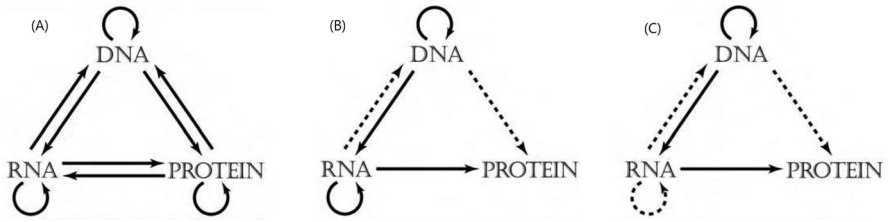

Fig. 10 Crick's the central dogma (Crick 1958)

5.3 A Mathematical Understanding of Mechanism Schemata

Biological mechanisms are temporal processes to produce biological phenomena. There are lots of mechanisms in all of the biological fields. Particularly, a mechanism of protein synthesis in molecular biology is a process from DNA to proteins via mRNA. Crick (1958) discussed all possible transfers of sequential information, such as (A) in Fig. 10. Crick argued that the sequential information of DNA only flows from nucleic acids to proteins, not vice versa, so-called the *central dogma* in molecular biology. In Fig. 10, (B) indicates all possible flows of sequential information under the dogma, which Crick imagined. The missing arrows at (B), compared with (A), including arrows from protein to DNA, RNA, and proteins themselves, imply impossible information transfers. (C) at Fig. 10 shows a tentative classification in Crick's 1970 paper (Crick 1970). Solid arrows indicate general transfers and dotted arrows indicate special transfers (Darden 2006, p.243). Temin suggested a provirus hypothesis that sequential information is transferred from RNA to DNA, and ultimately Temin's hypothesis was proven by the discovery of reverse transcriptase in the 1970s.

Most biologists, as well as proponents of the New Mechanisms in the philosophy of science, may believe that Fig. 10 is a representation of the mechanism of protein synthesis as a *mechanism schema* (see Machamer et al. 2000). The important thing is that concrete sequential information among macromolecules within the mechanism is ignored and just shown by molecular *names* such as 'DNA,' 'RNA,' and 'protein.' That is, the representation of the mechanism is a highly abstract diagram indicating that genetic information flows from nucleic acids into amino acids. Furthermore, diverse enzymes engage in every transition from one step to another. Still, those enzymatic roles in the mechanism are also neglected in Fig. 10.[8] For this reason, Fig. 10 is too simple to explain the full mechanistic steps. Mechanistic

[8] In the translational transition from mRNA to protein, for example, a small ribosomal subunit attaches itself to the 5' end of a messenger RNA sequence and moves along the mRNA until it searches for a start codon within the mRNA. Before the large ribosomal subunit synthesizes a sequence of amino acids, aminoacyl tRNA synthetase attaches amino acids to their corresponding tRNA molecules in advance. As soon as the small ribosomal subunit finds out the start codon, the large ribosomal subunits join the first tRNA together. Subsequently, other tRNAs with anticodons matching the mRNA codons bind to the growing peptide chain of amino acids by the large ribosomal subunits. Until the ribosome reaches a stop codon, the synthesis continues repeatedly.

explanations must reveal spatially and temporally organized characteristics among parts (or entities) and their operations (or activities) within the mechanism to exhibit how explanandum phenomena occur (Machamer et al. 2000; Darden 2006). No matter how we can regard Fig. 10 as a mechanism schema of protein synthesis, this figure oversimplifies complex steps into an arrow from RNA to protein. Highly abstract representations of biological mechanisms are not directly identified with complete explanations of the mechanisms.

Although the central dogma such as DNA → RNA → Protein never includes fully complete details, why is this simple representation taken into account as a schema *of* a mechanism? Advocates of the New Mechanism, particularly Machamer, Darden, and Craver, think that mechanism schemata allow for varying degrees of abstraction depending on how much detail is included. According to their historical analyses, Crick's hypothetical diagram, such as Fig. 10 as was a research guideline in molecular biology.[9] Historically, after the mid-twentieth century, many molecular biologists and biochemists have filled in a lot of biological discoveries from DNA to protein to achieve the mechanism's productive continuity. When the mechanism schemata are specified by instantiating concrete components and fleshing out arrows with detailed spatiotemporal descriptions, the mechanism schemata become a type-level abstraction for mechanistic explanations. A type-level abstract scheme played a roughly demonstrated representation of biological mechanisms.

Then how is an abstract mechanism schema related to a mechanistic explanation? If mechanism schemata are distinguished from mechanistic explanations, then is it implied that the mechanism schemata are non-explanatory? Of course, I do not deny that detailed descriptions of spatiotemporal organizations of mechanisms provide us with an understanding of how proteins are synthesized. However, the explanatory forces of mechanistic explanation are not solely given by spatiotemporal organizations. Note that a biological mechanism is a system including temporal causal chains among parts. An understanding of how different entities are causally connected is also a critical explanatory goal of mechanistic explanation. *Invariance* is a crucial feature to represent biological mechanisms as causal processes. Suppose invariant properties are shared among different types of entities within a mechanism. In that case, those properties give a clue for understanding how independent individual macromolecules are causally connected.

Using Crick's abstract diagram, we can judge that heterogeneous macromolecules are causally associated with each other only when we know that the spatially organizational features of molecular objects, such as orientation and sequential structures, are constantly maintained through synthesizing proteins. As discussed in Sect. 4.2, a partially ordered set represents the orientational feature of the sequential structure of macromolecules. The orientation from the 5'-end to the 3'-end of nucleotides is mapped into the orientation from the N-terminal end to the C-terminal end of amino acids homogeneously. That is, the partially ordered feature of arrangements of components within macromolecules is invariant.

[9] See Machamer et al. (2000) and Darden (2006) ch. 3.

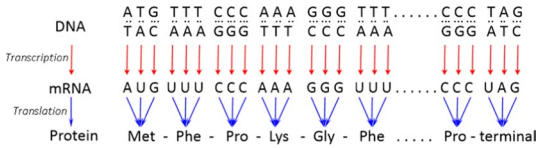

Fig. 11 Flows of sequential information

Most significantly, the arrangements of three bases within nucleic acids are mapped into a single target among twenty kinds of amino acids, too (see Fig. 11). The mapping between DNA and mRNA is invariant based on Chargaff's rule. The mapping between mRNA and proteins is also invariant based on the genetic codes discovered by Nirenberg and Matthaei in 1961 (Nirenberg and Matthaei 1961).

Notice that the mapping from DNA to mRNA differs from the mapping from mRNA to protein. The former is a bijective or one-to-one correspondence, whereas the latter is injective. Bijection implies an inverse mapping, but injection does not. This mathematical understanding of relationships among biological objects within the mechanism of protein is directly associated with a fundamental feature of the unidirectional flow of genetic information from DNA to the protein via mRNA. Recall that Crick's hypothetical diagram shows the impossibility of the reverse informational flow from protein to RNA. This feature indicates a significant aspect of mechanistic explanation related to a mechanism's temporal organizations. Temporal organizations include order, rate, and duration of operations or chemical reactions. The sequential ordering of bases within DNA can be preserved at that of RNA. And Temin proved that the reverse transformation from RNA to DNA is also possible. However, the sequential ordering of bases within nucleic acids cannot be transformed from that of amino acids. That is, a mathematical understanding of structural aspects among DNA, mRNA, and protein with the mapping concept indirectly gives abstract evidence to Crick's hypothesis.

Crick's diagrams also symbolize a temporal order of the mechanism of protein synthesis from the viewpoint of genetic information. Fig. 10 is a mechanism schema of protein synthesis to demonstrate that genetic information of nucleic acids flows into proteins, not vice versa. Mapping relations and invariance are also essential concepts in category theory in mathematics. For this reason, categorical abstractions of structural features of biological objects can help in understanding Crick's main ideas of central dogma in molecular biology. With a mechanism schema of protein synthesis we can implicitly imagine such that the genetic information of DNA is structurally mapped into that of RNA, which is subsequently mapped into sequential information of amino acids. Type-level objects, not individual token sequences, can illustrate this internal imagination. Also, individual names, including DNA, RNA, and proteins, can be more specified by each ordered sequential structure at the lower level. For example, a DNA molecule is a colimit as an object having a sub-pattern such as Fig. 6. Individual nucleotides within the DNA molecule in Fig. 6 are also colimits as objects having their sub-patterns shown by each circle in Fig. 7. Three ordered nucleotides within a single strand of RNA are mapped into a type of twenty amino acids through transfer RNA and a genetic code. That is, categorical representations of biological structures shed light on why a simple mechanism schema such

as Fig. 10 explains the phenomenon of protein synthesis without detailed descriptions of transcription and translation, such as Fig. 11. I argue that categorically interpreted mechanism schemata are explanatory because those diagrams show a constantly invariant property of informational and structural features in the mechanism of protein synthesis. In the following subsection, we will see that invariance in category theory plays a vital role in generalizing higher-order types of biological objects.

5.4 Generalizing Higher-Order Types of Biological Molecules

Nucleic acids include not only DNA but also RNA, which consists of messenger RNA (mRNA), transfer RNA (tRNA), ribosomal RNA (rRNA), and so forth. Among them, mRNA mediates genetic information from sequences of DNA to those of proteins. Interestingly, all sequential information of mRNA transcribed from DNA in eukaryotes is not employed when amino acids are synthesized. All initial transcription products, called precursor transcripts of mRNA (pre-mRNA), are spliced and ligated. Pre-mRNA consists of two parts: exons, whose sequences are used in synthesizing proteins, and introns, which are excised from the pre-mRNA. When Crick suggested the central dogma such as Fig. 10, he did not imagine pre-mRNA's editing processes. Since the 1970s, it has been revealed that some enzymes regulate splicing patterns. A remarkable discovery was that some pre-RNAs are spliced by themselves without any enzymatic proteins. A new concept *ribozyme* refers to as auto-catalytic pre-RNA to excise its introns and to rejoin exons again (Kruger et al. 1982). Since the 1980s, self-splicing reactions have been found in species as widely dispersed as bacteria and eukaryotes.

Auto-catalytic pre-RNAs are classified into two groups (Saldanha et al. 1993). Group I introns are found in fungal and plant mitochondrial DNAs (mtDNAs), bacteria such as nuclear rRNA genes of *Tetrahymena*, and even eukaryotes. Group II introns are found only in fungal and plant mitochondria and chloroplasts. Both groups are commonly ribozymes but differ concerning two types of type-level features, structural and functional reactions.

Note that a categorical framework for formalizing molecular structures of biological objects depends upon a fundamental assumption that the structural features of an object are *invariant* at those of another object. Nucleic acids are the higher-order types, such as DNA and diverse RNA, and their lower-order typed biological objects share the phosphate–sugar linkage so-called backbone. Similarly, type-level abstraction of biological structures helps in classifying group I and II molecules. In the case of the ribozyme, biologists pick up commonly shared sequences among individual introns by comparing ordered arrangements of bases in secondary structures. Group I introns catalyze their splicing based on their highly conserved secondary structures. In (A) of Fig. 12, P, Q, R, and S, indicated by heavy lines, are conserved sequences of group I introns. Thin lines simplify non-conserved sequences. And dashed lines indicate base pairs and a dot indicates wobble base pair at the 5' splice site. In (B) of Fig. 12, I to VI indicate conserved sequences of group II introns.

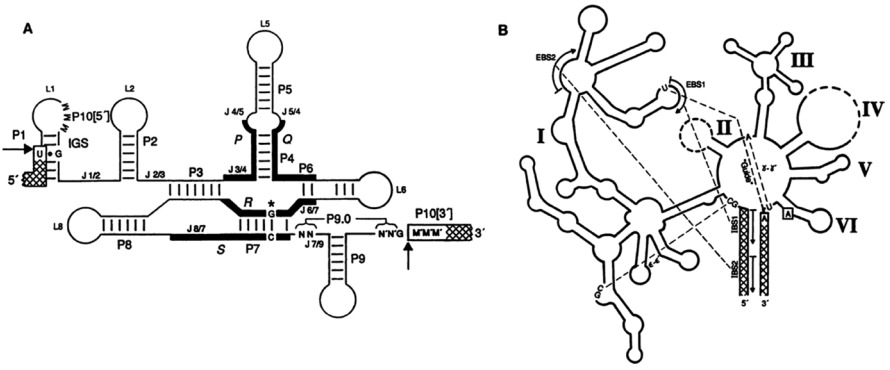

Fig. 12 Group I (left) and II (right) introns conserved structures (Saldanha et al. 1993, pp.17–19)

The dotted lines indicate interactions of two binding sites between exon and intron. Group I and group II introns have different invariant structural portions, respectively. That is, structural invariance at the type-level is a key epistemic norm to sort out types of molecules.

As protein synthesis can be represented categorically, splicing reactions are also represented by type-level abstraction. Recall that genetic orders of nucleic acids are invariantly maintained with orders of amino acids. Also, two steps, including transcription and translation, are common functional reactions across diverse species. Similarly, the group I and II introns are excised by two common transesterification reaction steps. In the case of group I introns, the splicing reactions begin when the 3' hydroxyl nucleophilic guanosine cofactor attacks the 5P splice site. And the 3' hydroxyl nucleophilic uracil terminal within a spliced 5' exon attacks the G site between the intron and 3' exon. In the end, the group I intron is separated from ligated exon. On the other hand, the splicing reactions of group II introns initiate when the branch site's 2' hydroxyl nucleophilic adenosine attacks the 5' splice junction. Subsequently, group II introns form a cyclized structure. Simultaneously, the 3' hydroxyl terminal within the 5' exon attacks the binding site between the intron and 3' exon. Finally, the group II intron is separated from ligated exon. Group, I and group II introns share common type-level functional reactions, whose specific individual processes differ (Fig. 13). That is, structural invariance at the type-level is also a helpful epistemic norm to generalize common types of functional processes in different categories.

In summary, the group I and II introns are generalized as different higher-order objects that are type-level abstractions of individual sub-objects. This classification can be possible because of the commonly invariant property of structural and functional aspects. A categorical understanding of type-level abstraction of molecular structures provides a representational clue to figuring out the emergence of higher-order concepts and classification rules of biological objects.

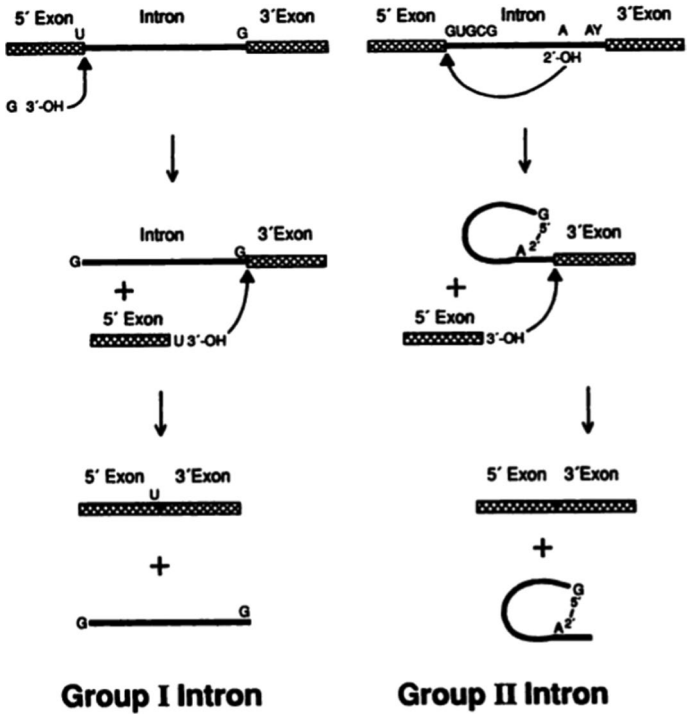

Fig. 13 Group I and Group II introns (Saldanha et al. 1993, p.16)

6 Conclusion

In this paper, I pursue a formal framework to represent molecular structures of biological objects in molecular biology abstractly. I specify three features of biological structures, so-called spatial organizations, such as directional orientation and shape, compositional hierarchy, and functionality. Those types of knowledge about biological objects are typically common features among different macromolecules, including nucleic acids, polysaccharides, proteins, and so on. Thus, an abstract formal toolkit is required to figure out type-level structural aspects of objects regardless of token-level ontological diversity.

I suggest mathematical abstractions to represent molecular structures with category theory graphically. In the philosophy of science, logical empiricists have traditionally used an axiomatic method with formalizing scientific knowledge as linguistic axioms and deductively derived theorems. No matter how we can denote molecular structures with theoretical terms, including chemical concepts such as various molecular references and predicates as to chemical connections, linguistic axiomatizations have limitations in illustrating visual aspects of molecular structures. I focus on category theory as a promising alternative to an axiomatic method. That is because category theory gives us to demonstrate intra-connective relations

among sub-objects of an object, such as various covalent bonds and hydrogen bonds. Of course, set-theoretical abstractions allow us to refer to the names of sub-objects and chemical linkages separately. But those abstractions also need to be supported when visualizing their structural features of them. By contrast, category-theoretical abstractions provide graphical ways to denote sub-objects as dots and internal relations as arrows among them simultaneously. As category theory is widely used when discussing mathematical structures, it is also expected to be applied when visualizing molecular structures of biological objects.

I demonstrate how category theory can be used to graphically represent molecular structures of nucleic acids by emphasizing either some philosophical advantages or biological strategies of a categorical framework of biological objects. Contrary to axiomatic methods, which are traditionally dealt with as essential formal tools to analyze scientific theories and methods, category theory is the most promising formal language to represent structural relationships among units of macromolecules. In Sect. 5.1, I suggest a sharp distinction between two terms, entities and structures, and define that structures include objects (or entities) and internal relations among the entities. Molecular structures of biological objects are also composed of both of them. In category theory, categories are defined as structures consisting of objects and morphisms (or arrows). When we think of diagrammatic representations as a frequently used type of scientific knowledge in molecular biology, category theory is a successful language to demonstrate biological structures formally. Additionally, categorical abstractions illuminate the compositional hierarchy from sub-objects at the lowest level to objects at the highest level.

Categorical abstractions of biological objects are necessary when understanding (i) the inference of the functional properties from structural features of biological objects, (ii) an explanatory role of highly abstract schema such as the central dogma in molecular biology, and (iii) generalization of higher-order types of biological molecules. That is, we cannot comprehend the above three topics at this point in the absence of categorical abstractions. Some mathematical concepts such as product, mapping, and invariance are also required to understand the three topics fully.

If molecular structures of biological objects are stipulated through category theory, we can understand how molecular biologists could infer the functional properties of biological objects from their structural features. As a mathematical language, category theory gives several significant concepts to analyze structures. With a concept of product, we can prove why the double helical material of DNA is interpreted as an abstract gene. It was well-known that a gene was self-replicated before Watson and Crick's discovery. With a qualitative mapping analysis among structures, the phosphate-sugar linkage plays a template role in replicating a new DNA. Furthermore, as a single sequence of nucleotides is synthesized by a polymerization reaction in the artificial tubes, the sugar-base linkage can be formed by the same reaction. Categorical abstractions about molecular structures of biological objects and mathematical toolkits give us a proper understanding of a heuristic inference from structure to function.

Another critical concept of category theory is invariance. The universal mapping properties are a central concept in category theory when figuring out invariant relationships among different mathematical structures. I admit that a highly abstract

mechanism schema such as Crick's hypothetical idea of the central dogma in molecular biology differs from a mechanistic explanation because Crick's diagram omits spatiotemporally organizational features. However, I show that an explanatory force of mechanistic explanation can be supported by revealing how different types of objects are causally connected and argue that invariant features are causal marks to organize them. Categorical abstractions of molecular structures of biological objects are prerequisites to identifying the invariant features of the objects. I also show that an invariance analysis allows for generalizing higher-order types of biological molecules by discussing a case of the ribozyme.

This proper understanding of the biological structures of individual objects is essentially required when we scrutinize the formal framework of biological mechanisms. Mechanisms consist of entities and activities; particularly activities perform to produce explanandum phenomena. When we explain the phenomena by revealing their mechanisms, we must show what states are temporally changed into another, what makes the changes of states, and how they are spatially organized. Commonly shared features of biological structures among biological objects indicate biochemical states within biological mechanisms. Additionally, spatially organized structures of proteins are significant when we search for what makes transitions from one state to another because proteins play an enzymatic role in biochemical reactions. Without a proper understanding of the biological structures of individual objects, no mechanistic explanations become complete and correct. Our discussion of the formal representation of biological structures of individual objects is a cornerstone of further formal analyses of biological mechanisms and the nature of mechanistic explanation. Furthermore, invariance, the essential mathematical property in category theory, is expected to be applied significantly when formally discussing biological mechanisms in the future.

References

Albert B et al (2014) Essential cell biology, 4th edn, Garland Science
Awodey S (2010) Category theory. Oxford University Press, Oxford
Bechtel W, Abrahamsen A (2005) Explanation: a mechanist alternative. Stud History Philos Biol Biomed Sci 36(2):421–441
Bechtel W, Abrahamsen A (2012) Diagramming phenomena for mechanistic explanation. Proc Annual Meet Cognitive Sci Soc 34(34):102–107
Carnap R (1953) Testability and meaning. In Feigl, Brodbeck (eds) Reading in the philosophy of science, Appleton, Century, Crofts, pp. 47–92
Craver C, Darden L (2013) Search of mechanisms: discoveries across the life sciences. The University of Chicago Press, Chicago and London
Craver C (2001) 'Structures of Scientific Theories,' in P. Machamer and M. Silberstein (eds.), The Blackwell guide to the philosophy of science. Blackwell Publishers. 55-79
Crick F (1958) On protein synthesis. Symp Soc Exp Biol 12:8
Crick F (1970) Central dogma of molecular biology. Nature 227:561–563
Crick F (1988) What mad pursuit: a personal view of scientific discovery. Basic Books
Culp S, Kitcher P (1989) Theory structure and theory change in contemporary molecular biology. Br J Philos Sci 40(4):459–483
Costa Da, Newton CA, French Steven (2003) Science and partial truth: a unitary approach to models and scientific reasoning. Oxford University Press, New York

Darden L (2006) Reasoning in biological discoveries: essays on mechanisms, interfield relations, and anomaly resolution. Cambridge University Press, Cambridge
Ehresmann AC, Vanbremeersch JP (2007) Memory evolutive systems: hierarchy, emergence, cognition, Elsevier
Ehresmann A (2018) Applications of categories to biology and cognition. In Landry E (ed) Categories for the working philosopher, Oxford University Press, pp. 358–380
Ehresmann A, Vanbremeersch JP (2019) MES: a mathematical model for the revival of natural philosophy. Philosophies 4(1):9
Glennan S, Illari P (2018) The Routledge handbook of mechanisms and mechanical philosophy, Routledge
Gómez-Ramirez J (2014) A new foundation for representation in cognitive and brain science: category theory and the hippocampus, Springer
Kruger K, Grabowski P, Zaug A, Sands J, Gottschling D, Cech T (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. Cell 31:147–157
Louie AH (2009) More than life itself: a synthetic continuation in relational biology. Ontos Verlag, Frankfurt
Machamer P, Darden L, Craver CF (2000) Thinking about mechanisms. Philos Sci 67:1–25
McKinsey JCC, Sugar AC, Suppes Patrick (1953) Axiomatic foundations of classical particle mechanics. J Ration Mech Anal 2(2):253–272
Nirenberg M, Matthaei J (1961) The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. Proc Natl Acad Sci 47:1588–1602
Rosen R (1991) Life itself: a comprehensive inquiry into the nature, origin, and fabrication of life, Columbia University Press, New York
Saldanha R, Mohr G, Belfort M, Lambowitz A (1993) Group I and group II introns. FASEB J 7:15–24
Stryer L et al (2017) Biochemistry. Freeman & Company
Suppe F (1977) The structure of scientific theories. University of Illinois Press
Suppes P (2002) Representation and invariance of scientific structures. Center for the Study of Language and Information Stanford Publications, California
van Fraassen BC (1980) The scientific image. Clarendon Press, Oxford
van Fraassen BC (1989) Laws and symmetry. Clarendon Press, Oxford
Watson J, Crick F (1953a) A proposed structure for deoxyribose nucleic acid. Nature 171:737–738
Watson J, Crick F (1953b) General implications of the structure of deoxyribonucleic acid. Nature 171:964–967
Watson J (2001) The double helix. Weidenfeld and Nicolson, London