A symmetry model for genetic coding via a wallpaper group composed of the traditional four bases and an imaginary base E: Towards category theory-like systematization of molecular/genetic biology

Jitsuki Sawamura1*, Shigeru Morishita2 and Jun Ishigooka1

Abstract

Background: Previously, we suggested prototypal models that describe some clinical states based on group postulates. Here, we demonstrate a group/category theory-like model for molecular/genetic biology as an alternative application of our previous model. Specifically, we focus on deoxyribonucleic acid (DNA) base sequences.

Results: We construct a wallpaper pattern based on a five-letter cruciform motif with letters C, A, T, G, and E. Whereas the first four letters represent the standard DNA bases, the fifth is introduced for ease in formulating group operations that reproduce insertions and deletions of DNA base sequences. A basic group \( Z_5 = \{ r, u, d, l, n \} \) of operations is defined for the wallpaper pattern, with which a sequence of points can be generated corresponding to changes of a base in a DNA sequence by following the orbit of a point of the pattern under operations in group \( Z_5 \).

Other manipulations of DNA sequence can be treated using a vector-like notation \( \mathbf{D}_j \) corresponding to a DNA sequence but based on the five-letter base set; also, \( \mathbf{D}_j \)s are expressed graphically. Insertions and deletions of a series of letters \( \mathbf{E} \) are admitted to assist in describing DNA recombination. Likewise, a vector-like notation \( \mathbf{R}_j \) can be constructed for sequences of ribonucleic acid (RNA). The wallpaper group \( B = \left( Z_5^{\times \infty}, \bigcirc \right) \) (an \( \infty \)-fold Cartesian product of \( Z_5 \)) acts on \( \mathbf{D}_j \) (or \( \mathbf{R}_j \)) yielding changes to \( \mathbf{D}_k \) (or \( \mathbf{R}_k \)) denoted by \( \mathbf{D}_j \circ B_{(j \rightarrow k)} = \mathbf{D}_k \) (or \( \mathbf{R}_j \circ B_{(j \rightarrow k)} = \mathbf{R}_k \)). Based on the operations of this group, two types of groups—a modulo 5 linear group and a rotational group over the Gaussian plane, acting on the five bases—are linked as parts of the wallpaper group for broader applications. As a result, changes, insertions/deletions and DNA (RNA) recombination (partial/total conversion) are described. As an exploratory study, a notation for the canonical “central dogma” via a category theory-like way is presented for future developments.

Conclusions: Despite the large incompleteness of our methodology, there is fertile ground to consider a symmetry model for genetic coding based on our specific wallpaper group. A more integrated formulation containing “central dogma” for future molecular/genetic biology remains to be explored.

Keywords: DNA (RNA) bases, Imaginary base, Wallpaper group, Operation, Cartesian vector, Category, Central dogma
Background

Group theory is the cornerstone in classifying and studying abstract concepts involving symmetry [1,2]. In general, when group theory is used in various fields of natural sciences, it plays an important role in describing geometrical or dynamical symmetries of phenomena under consideration; examples include mathematics [3,4], physics [5-8], chemistry [9], molecular/genetic biology [10-22], and anthropology [23]. Moreover, much fertile ground still exists where group theory can display its versatility from a multitude of viewpoints. To our knowledge, one such candidate is molecular/genetic biology where group theory has already provided great contributions [10-22].

Deoxyribonucleic acid (DNA) is a nucleic acid containing genetic instructions coded in ordered sequences of four bases located in genes that determine specific genetic characteristics of an organism. In the canonical Watson-Crick DNA base pairing, adenine (A) forms a base pair with thymine (T) and guanine (G) forms a base pair with cytosine (C) [24-26]. Similarly, ribonucleic acid (RNA), which has various biological roles, is a molecule that has a much shorter chain of nucleotides. The sequence of DNA consisting of bases ‘A, C, T and G’ is transcribed into RNA, composed of bases ‘A, C, U and G’; the sets differ in that ‘U (uracil)’ replaces ‘T (thymine)’.

Over the latter half of the 20th century, the nature of the genetic code became fairly well established. As for the coding sequences of DNA into nucleotide units, one needs to build up more general, sophisticated, rationally functionalized systematics concerning DNA base sequences that will enable genes to be understood at the molecular biology level in more optimized form. Indeed, many approaches have been undertaken to describe gene characteristics from various viewpoints within the participating disciplines [24-42]. In particular, the concept of ‘symmetry’ for DNA sequences plays an important role in understanding their characteristics.

However, each has its advantages and disadvantages in terms of utility and convenience in applications. To our knowledge, so far, if we intend to incorporate a sequence of bases into another sequence and/or exclude certain bases from that substitution, we need to look further afield because normally, sequencing and inserting-deleting operations cannot help in distinguishing one from the other. That means that multiple types of operations are necessary if features of DNA containing exceptional sequences are to be treated.

Previously, we suggested prototypal models that describe some clinical states based on group postulates [43]. In this article, we demonstrate a group/category theory-like model for molecular/genetic biology as an alternative application of our previous model. Specifically, focusing on DNA base sequences, we present a simple model where not only changes in sequences of DNA bases but also insertion, deletion, and recombination (partial/total conversion) of DNA bases are treatable within some simple rules via the combination of a set and a group defined over some specific wallpaper pattern. Moreover, a category theory-like formalism, where a description of the DNA bases and their transcription to RNA bases can be made, is attempted from which a category theory-like framework is constructed requiring as few and as simple rules as possible. As an example, by assimilating the canonical “central dogma” [26], we hope to provoke more interactivity among those interested branches of natural science, if possible. The methodology consists of eight parts, the content of which is built-up step-by-step as scope is enlarged to encompass the more advanced themes.
A preliminary setting describing a wallpaper pattern used as a symmetry model for DNA sequences

First, we consider a certain wallpaper pattern that helps us to visualize the operations of the present model (see Figure 1) [2,44-47]. There, the pattern comprises repetitions of a cruciform motif with each motif consisting of five letters E, C, A, T, and G with the latter four letters equally spaced at the points of a cross about a central E. The motif generates the pattern through a translation specified as a knight’s move in chess—two steps out and one step right. In this way, the grid-points in this regular wallpaper pattern can be obtained uniquely and be extended indefinitely. Note also that each horizontal line is generated by repetitions of the sequence E-C-A-T-G. Moreover, the line above is a displaced copy of the one below with letter A placed directly above letter E. This preserves the condition that any cruciform is composed of one each of the five letters.

The wallpaper pattern as an array of cruciforms is capable of being constructed as stacks of a unit cell (the 5 × 5 square enclosed in the dotted line in Figure 1) by horizontal and vertical translations [2,44-47]. The positions of the bases of the cruciform motif are so determined to make it easier to determine the complementary base of each base; the practical applications are clarified later. We introduce the letter ‘E’ to indicate an ‘empty’ base which is treated in the same way as the other bases at least for display purposes. This five-base scheme is adopted to aid the notion of group composition in our model.

In addition, we focus on a point ‘P’ on the wallpaper pattern (i.e., the grid-point array in Figure 1), to compose a certain DNA base sequence. In accordance with this, we shall always adjoin a series of letters that are determined as a trajectory of the point ‘P’—also called the ‘orbit of P’—over the wallpaper pattern. For instance, when we identify or recognize some changes of DNA bases with ‘P’ moving from ‘A → C → E’
over the wallpaper pattern, then this represents a series of changes to one base located at a specific position of a DNA sequence in the manner ‘ACE...’ or ‘...A...’→ ‘...C...’→ ‘...E...’.” The ‘orbit of P’ can describe series of sequences of DNA bases, or series of changes of each letters in the same places, although, in this article, we focus mainly on the latter case, without provisory context.

With these postulates, we consider the set $C_5 = \{C, A, T, G, E\}$. If the point ‘P’ moves onto an ‘E; ‘E’ must be included and identified in the series of letters, as in ‘ACGTE’; for example. This is interpreted as the series of DNA bases ‘ACGT’. Thus, ‘E’ depends on context; that is, ‘E’ can be inserted or removed from any series where we would like to include or eliminate ‘E’s so long as these are recognized/tracked in the entire process. When read from left to right, the place number of each letter in the series is subscribed, as in ‘A1C2G3T4’. After insertions/deletions, the place number is augmented/diminished depending on initial and final positions; hence following three insertions ‘A1C2G3T4’→ ‘A1E2C3G4E5E6T7’; this means the point ‘P’ takes the place ‘E’ once between A1 and C3, and twice between G4 and T7 over the wallpaper pattern in Figure 1. More details are to be given later.

As a further refinement, the orbit of ‘P’ can be stated as a sequence of shift operations as follows; let ‘r’ denote a move one step to the right corresponding to say A→ T, T→ G or G→ E. Similarly, we denote ‘l’: move one step to the left as for C→ E, and E→ G; ‘u’: move one step up; and ‘d’: move down. We include ‘n’ to designate a ‘no move’ (remain at the same point). A sequence of ‘r’, ‘u’, ‘d’, ‘l’, and ‘n’ then provides a position-independent means to describe the orbit of ‘P’; any of these five operations can be applied to any of the five letters. We denote their operations on ‘P’ in the following way. If point ‘P’ moves from ‘E’ to ‘C’ (step to the right), we write ‘E→ r→ C’ where ‘→ r’ signifies apply ‘r’ to ‘E’ (see Figure 1). In a similar way, ‘E: l→ G; ’ ‘E: u→ A; ’ ‘E: d→ T’ and ‘E= n→ E’. Note though that each operator means a change of one base to another base within these five bases; the meaning of ‘=’ is not the degree of translation but equivalence to the resultant base from the wallpaper pattern.

To shorten multiple applications of the operations, we introduce ‘●’ to denote the composition of two operations, for example, ‘((E: r): u) = E: r ● u’. From Figure 1, we find ‘E: d = T’ yields the same change as ‘E: r ● u = T’. As other examples, ‘r ● r = u’ results in ‘r ● r = u’, and ‘d ● d = n’ results in ‘d ● d = r’, because from Figure 1, ‘E: r ● r = E: u = A’; and ‘E: d ● d = E: r = C’. All possible one-step changes between letters ‘C, A, T, G, E’ and operators ‘r, u, d, l, n’, and all possible compositions of operators for the wallpaper pattern of Figure 1 are presented in Figures 2 and 3, and Appendix A.

The binary compositions among the five operations ‘r, u, d, l and n’ can be shown to satisfy the Abelian group postulates (wallpaper group/plane symmetry group/plane crystallographic group [2-4,44-47]). Indeed, let $Z_5 = \{r, u, d, l, n\}$, then $\{Z_5, \bullet\}$ is the Abelian group of order five. That is, for all elements $x \in Z_5$, we have:

1) Associativity: $x \bullet (y \bullet z) = (x \bullet y) \bullet z$, (x, y or z being arbitrary elements belonging to $Z_5$);
2) Identity: ‘n’ is an identity element such that $x \bullet n = n \bullet x = x$;
3) Inverse: a unique element $x^{-1}$ exists such that $x \bullet x^{-1} = x^{-1} \bullet x = n$ (‘$x^{-1}$ is called the inverse element of x’);
4) Commutativity: $x \bullet y = y \bullet x$,
5) Closure: any combination of operations between $x \bullet y$ belongs to $Z_5$. 

### Figure 2

Cayley tables for the five bases and five operations of linear/rotational groups. Any of the five operations on any of five bases yields a base in a cyclic order.

|     | r   | u   | d   | l   | n   |
|-----|-----|-----|-----|-----|-----|
| O   | C   |     |     |     |     |
|     | (→0₁) |     |     |     |     |
| E   |     | A   | T   | G   | E   |
|     | (→0₂) |     |     |     |     |
| C   |     |     | G   | E   | C   |
|     | (→0₃) |     |     |     |     |
| A   |     |     |     | E   | A   |
|     | (→0₄) |     |     |     |     |
| T   |     |     |     |     | E   |
|     | (→0₅) |     |     |     |     |
| G   |     |     |     |     |     |
|     | (→0₆) |     |     |     |     |

*→*: bijection

### Figure 3

Cayley tables for the linear group, rotational group and wallpaper group for five bases. This confirms the bijection between the wallpaper group and rotational group.

|     | r   | u   | d   | l   | n   |
|-----|-----|-----|-----|-----|-----|
| .   |     |     |     |     |     |
|     | (→0₁) |     |     |     |     |
| r   | C   |     |     |     |     |
|     | (→0₂) |     |     |     |     |
| u   |     | A   | T   | G   | E   |
|     | (→0₃) |     |     |     |     |
| d   |     |     | G   | E   | C   |
|     | (→0₄) |     |     |     |     |
| l   |     |     |     |     |     |
|     | (→0₅) |     |     |     |     |
| n   |     |     |     |     |     |
|     | (→0₆) |     |     |     |     |

*→*: bijection
Therefore, $Z_5$ is an Abelian group [2-4,44-47]. The inverses for each of the elements are:

$$r^{-1} = 1, l^{-1} = r, u^{-1} = d, d^{-1} = u, n^{-1} = n,$$

which can be used to complete the composition table—also known as the Cayley table of the group.

We further stipulate that when we perform these operations, then we always assume/identify the coding of the sequence of DNA bases in accordance with these operations, and vice versa. This is because the action of ‘u’ on ‘E’ yields base ‘A’, that of ‘d’ on ‘E’ yields base ‘T’, that of ‘l’ on ‘E’ yield base ‘G’, and naturally that of ‘n’ on ‘E’ results in the same ‘E’. For a more complex example, we might insert a certain series of ‘A’s in ‘ACCGT’ between the 3rd and 4th base. To begin, we decide to write this manipulation as follows: ‘ACC( )GT’ is transformed into ‘ACC (E)GT’ by inserting ‘E’. Next, because the operation ‘u’ to the new 4th component ‘E’ yields ‘E→A’ (ACCGT→ACCAGT), and vice versa, that is, ‘d’ operating on the 4th component ‘A’ produces ‘A→E’ (ACCAGT→ACCGT). In this way, appropriate use of ‘E’s through the adequate combination of operators of $Z_5$ enables to express inclusion and/or exclusion of any base between bases in a DNA sequence. To indicate this, we adopt a vector-like description with an infinite number of ‘E’s being assumed to be present at the end of any given base sequence. This means the point ‘P’ takes ‘E’s an infinite number of times over the wallpaper pattern (Figure 1); i.e.,

$$D_j = [C|T|G|A|T|A|A|C|E|E|E|E|E|E|E|E|E|E]$$

$$(2a)$$

$$= [C_1|T_2(G_3)|A_4|T_5(A_6)|A_7|C_8|E_9|E_10|E_11|E_12|E_13|E_14|E_15|E_16|E_17|E_18|E_19|E_20|E_21|E_22|E_23|E_24]$$

$$(2b)$$

(j: the number of the sequence, N: the number of single-stranded DNA bases of $D_j$’s except for the infinite tail of ‘E’s; in the above case, N = 8)

In the last expression (2b), ‘E_3’ is inserted before the 3rd component ‘G_3’ and 6th component ‘A_6’ marked by ‘( )’ in formula (2a), and the place numbers of all components to the right of the 3rd component are all incremented by ‘1’, those to the right of the 6th component; by ‘1 + 2’. Likewise, we assume that the deletion of any ‘E’s that are already displayed in $D_j$ is always permissible according to need with the place numbers being decreased by the necessary size.

Essentially, we regard the subscripted place number of a component of $D_j$, e.g., ‘3’ of ‘A_3’, as a convenient place mark to help in recognizing and counting the order of sequences. Place numbers remain fixed when performing operations within a series of operations during code recognition of bases. However, for an operation, another place number is always permissible in principle, from where indexing of a specific DNA base sequence starts.

Alternatively, we use the following notation to describe various cases:

1) we denote by ‘$\{D_j\}$’ a sequence ‘$D_j$’ where specified ‘E’s other than the trailing series of ‘E’s are implicitly implied but the place number indexing is retained; i.e.,

$$\{D_j\} = [C_1|T_2|G_3|A_5|T_6|A_9|A_{10}|C_{11}|E_{12}|E_{13}|E_{14}|E_{15}|E_{16}|E_{17}|E_{18}|E_{19}|E_{20}|E_{21}|E_{22}|E_{23}|E_{24}]$$

$$(3a)$$
Here, the explicitly indicated place numbers are the same as in (2b) and missing subscripted place numbers indicate omitted 'E's. Hence, (3a) without trailing 'E's and subscripts represents an ordinal/conventional DNA sequence.

2) we denote by '<Dj>' a sequence 'Dj' where specified explicit 'E's other than the trailing 'E's are deleted (changed into implicit 'E's) and the base sequence is re-indexed with sequential place numbers, i.e.,

\[
<Dj> = [C_1|T_2|G_3|A_4|T_5|A_6|A_7|C_8|E_9|E_{10}|E_{11}|E_{12}|E_{13}|…].
\]

Note that 'E's other than the trailing 'E's are not recognized as explicit components and hence are not indexed. Additional insertions/deletions of 'E's are permitted after deletions of 'E's; therefore, apart from the trailing 'E's, (3b) signifies an ordinal/conventional DNA sequence.

Although equivalent to 'CTGATAAC' as an actual DNA sequence expressions, related expressions \{Dj\} and <Dj> differ from each other; the former retains all information regarding inserted 'E's and place numbers whereas the latter does not.

In an extension of the notation, a multiple sequence of deletions of 'E's (say t-times) can be written as a t-tuple of '<<Dj>>' (t-tuple) = <Dj>. The final expression is without explicit 'E's other than those trailing at the end, and thus formulates a genuine DNA sequence after the appearance of indels. (Short for insertion/deletion markers, the idels are strings of mutated base pairs.) Similarly for the operation \{\}, we have '({{Dj}}) (t-tuple) = {Dj}'. The operations '{ }' and '<>' can be performed freely when necessary; if further indels occur at say 'G3' and 'A7' in

\[
<Dj> = [C_1|T_2|G_3|A_4|T_5|A_6|A_7|C_8|E_9|E_{10}|E_{11}|E_{12}|E_{13}|…],
\]

then <Dj> changes into

\[
<Dj_1> = [C_1|T_2|E_3|A_4|T_5|A_6|E_7|C_8|E_{9}|E_{10}|E_{11}|E_{12}|E_{13}|…],
\]

and subsequently into

\[
<<Dj_1>> = [C_1|T_2|A_3|T_4|A_5|C_6|E_7|E_8|E_{9}|E_{10}|E_{11}|…].
\]

The sequence <Dj_1> contains implicit 'E's aside from the trailing 'E's, and can be written as

\[
{<Dj_1>>} = [C_1|T_2|A_4|T_5|A_6|C_8|E_{10}|E_{11}|E_{12}|E_{13}|…].
\]

Naturally, \{<Dj_1>>\} and <Dj_1> are equivalent, but <<Dj_1>> and <Dj_1> differ. Moreover, as long as place numbers are recognized/traced precisely, combinations of manipulations '{ }' and '<>' are allowed; e.g., \{{<<Dj_1>>>\}). Hence, with appropriate use, we could treat (read, interpret, describe, record) conventional sequences of DNA via '{Dj}' or '<Dj>'. However, below we shall focus on simple sequences 'Dj'.

Looking at the beginning of a base sequence as in the following:

\[
D_j = [C_1|G_2|A_3|C_4|…|T_i|…|A_{N-1}|T_N|E_{N+1}|E_{N+2}|E_{N+3}|…],
\]

(i: i-th component of D_j; N: the number bases D_j)

a directionality for any D_j can be imposed;

\[
D_j(5→3) = [C_1|G_2|A_3|C_4|…|T_i|…|A_{N-1}|T_N|E_{N+1}|E_{N+2}|E_{N+3}|…],
\]

and

\[
D_j(3→5) = [T_1|A_2|…|T_{N+1−i}|…|C_{N−3}|A_{N−2}|G_{N−1}|C_N|E_{N+1}|E_{N+2}|E_{N+3}|…].
\]
The notation, ‘(5 → 3)’ and ‘(3 → 5)’, is simply an additional label representing the two possible types of endings of single-stranded DNA. Nonetheless, when the number of bases is finite, two sequences can be equivalent, as for example

\[ D_j(5 \rightarrow 3) = [C_1|G_2|A_3|C_4|T_5|A_6|T_7] \]

and

\[ D_j(3 \rightarrow 5) = [T_1|A_2|T_3|C_4|A_5|G_6|C_7], \]

unless the prime endings <5’(five prime) → 3’(three prime) > or <3’ → 5’> accompanies the sequence designation.

In accordance with these postulates, we now can define the set \( D = \{D_j (j = 1,2,3,\ldots) | D_j \in C_5 \times C_5 \times C_5 \times \ldots \text{ (N times, } N \leq \infty) \} \) as the set of all possible sequences of recognized N-tuple single-stranded DNA bases. We can regard N to be a positive integer or infinity.

An analogous definition is clearly possible for the set \( R \) of RNA sequences; with ‘T’ substituted by ‘U’, operations of group \( Z_5 \), are similarly definable because all results obtained for DNA pertain to RNA under the base substitution. Thus, set \( R = \{R_j (j = 1,2,3,\ldots) | R_j \in C_5 \times C_5 \times C_5 \times \ldots \text{ (N times, } N \leq \infty) \} \) is the set of all possible sequences of recognized N-tuple single-stranded RNA bases with \( C_5 = \{C, A, U, G, E\} \).

§2 Group composition that yields changes in DNA bases via a Cartesian vector

Next, we can consider \( B = \{B_m (m = 1,2,3,\ldots) | B_m \in Z_5 \times Z_5 \times Z_5 \times \ldots \text{ (an N-fold product, } N = \infty) \} = \{Z_5^N, \bullet\} \), where elements of \( B \) act on any \( D_j \). This means that \( D_j \) covers all possible sequences of the DNA bases, and this situation is the same for \( R_j \) of sequences of RNA bases.

Because \( B \) is a Cartesian product of the same Abelian group, it is also Abelian, where composition of any two elements of \( B \) is denoted by ‘\( \bullet \)’ [4]. Details are shown in Appendix B and Figure 3. Accordingly, its formulation as a group \( B = \{Z_5^N, \bullet\} \) is confirmed.

In a more general context, a Cartesian vector that is composed of the respective operators ‘\( b_{(j \rightarrow k)} \)’ that effects the change \( D_j \) into \( D_k \) is definable in the following way:

\[
B_{(j \rightarrow k)} = [b_{(j \rightarrow k)}^1|b_{(j \rightarrow k)}^2|b_{(j \rightarrow k)}^3|\ldots|b_{(j \rightarrow k)}^{(N-1)}|b_{(j \rightarrow k)}^N]|n_{N+1}|n_{N+2}|n_{N+3}|\ldots),
\]

(\( N \) : the number of components).

Hence,

\[ 'D_j \cdot B_{(j \rightarrow k)} = D_k'. \]

Clearly, for arbitrary ‘j’ and ‘k’, there exists a unique ‘m’ such that \( 'B_{(j \rightarrow k)} = B_m (m = 1,2,3,\ldots)' \); despite the difference in notation, the two are identical in practice.

Here, we present a simple example that consists of a multiple product of ‘\( B_{(j \rightarrow k)} \)’.

Consider the scenario that a certain sequence of a single strand (or one side of a double-strand) of DNA transitions from \( D_1 \) to \( D_3 \) in stepwise fashion,

\[
D_1 = [A_1|C_2|C_3|G_4|T_5|E_6|E_7|\ldots] = [A_1|C_2|C_3( )G_4|T_5|E_6|E_7|\ldots],
D_2 = [A_1|C_2|C_3(E_4|E_5)|G_6|T_7|E_8|E_9|\ldots],
D_3 = [A_1|C_2|C_3(E_4|E_5)|T_6|G_7|E_8|E_9|\ldots],
D_4 = [C_1|T_2|G_3(T_4|C_5)|G_6|A_7|E_8|E_9|\ldots].
\]
We next consider the change ‘$D_1 \rightarrow D_2$’. There exists an operator ‘$B_{(1 \rightarrow 2)} = [n_1|n_2|n_3 (r_4|u_5)|n_6|n_7|n_8|n_9|\ldots]$’ that is able to produce this change, specifically, the insertion of two ‘E’s between ‘C_3’ and ‘G_4’ yields the change ‘$D_1 \rightarrow D_2$’. However, this sort of manipulation can be troublesome. Hence, in our model, insertion/deletion of ‘E’s are instead ascribed to the way the vector $D_j$ is interpreted. This is preferable as this avoids easier manipulations. Next, we construct the operator ‘$B_{(2 \rightarrow 3)}$’ that maps ‘$D_2 \rightarrow D_3$’ (the details are shown in Appendix C). With reference to Figures 1, 2 and 3, we find

$$B_{(2 \rightarrow 3)} = [n_1|n_2|n_3 (u_4|d_5)|n_6|n_7|n_8|n_9|\ldots].$$

In a similar manner,

$$B_{(3 \rightarrow 4)} = [l_1|u_2|d_3 (r_4|u_5)|n_6|l_7|n_8|n_9|\ldots].$$

Naturally, the final $D_4$ is obtained from $D_1$ recursively,

‘$D_1 \rightarrow D_2$’,

‘$D_2 \bullet B_{(2 \rightarrow 3)} \bullet B_{(3 \rightarrow 4)} = D_4$’.

From the decomposition

‘$B_{(j \rightarrow k)} = B_{(j \rightarrow 0)} \bullet B_{(0 \rightarrow k)} = B_j^{-1} \bullet B_k$’, we obtain

‘$D_j \bullet B_{(j \rightarrow k)} = D_j \bullet B_{(j \rightarrow 0)} \bullet B_{(0 \rightarrow k)}$

$= D_0 \bullet B_{(0 \rightarrow k)} = D_k$,

(7)

where

$$D_0 = [E_1|E_2|E_3|\ldots|E_i|\ldots|E_{N-1}|E_N|E_{N+1}|E_{N+2}|E_{N+3}|\ldots]$$

(8)

denotes the identity element of $D$.

Note that the group operations can act on $D_j$ irrespective of whether the ‘E’s are explicit or implicit as defined in §1. Moreover, any sequence ‘$D_j$’ can be presented as a polygonal line; as an example, the evolution of changes ‘$D_1 \rightarrow D_3$’ is displayed in Figure 4.

§3 Integration of a linear group and a rotational group as a wallpaper group

Looking at the definitions of groups $Z_5$, $D$, and $B$, another approach is possible. The five bases can be represented by five equispaced phasors with a ‘$2\pi/5$’ angular phase separation located on the unit circle on the Gaussian plane, as depicted in Figure 5.

Herein, in the Gaussian plane, if ‘$\omega$’ is defined to be the counterclockwise rotational angle ‘$\omega = 2\pi/5$ (rad)’ and composition of ‘$\omega$’ is denoted ‘$\bullet$’, then assuming ‘$\omega$’ obeys the ‘right translation rule’, we have

$$\begin{align*}
\omega &= \omega_1, \\
\omega \bullet \omega &= 2\omega = \omega_2, \\
\omega \bullet \omega \bullet \omega &= 3\omega = \omega_3, \\
\omega \bullet \omega \bullet \omega \bullet \omega &= 4\omega = \omega_4, \\
\omega \bullet \omega \bullet \omega \bullet \omega \bullet \omega &= 5\omega = \omega_5 = \omega_0 = 0 \text{ (= no rotation).}
\end{align*}$$

(9)

The general form of an arbitrary base is expressed as ‘$X_m \leftrightarrow \text{Exp}(m \cdot \omega \cdot \text{i})$’ (here, ‘$\text{i}$’ is the ‘imaginary unit’, $\omega = 2\pi/5$ (rad), $m = \{0, 1, 2, 3, 4, 5\}$). With ‘{*}’ meaning one of the
bases among 'C, A, T, G and E' we construct the following map. Denoting composition by '◦', \( \omega_m \) acts on the identity trivially and hence yields the correspondences

\[
\begin{align*}
\text{Exp } 0 & : i(\cdot) \mapsto \text{Exp } 0 \\
\text{Exp } 2 & : \pi(i(\cdot)) = 5 \mapsto \text{Exp } 1 \\
\text{Exp } 4 & : \pi(i(\cdot)) = 5 \mapsto \text{Exp } 2 \\
\text{Exp } 6 & : \pi(i(\cdot)) = 5 \mapsto \text{Exp } 3 \\
\text{Exp } 8 & : \pi(i(\cdot)) = 5 \mapsto \text{Exp } 4 \\
\end{align*}
\]

Expanding the operations for \( \omega_1, \omega_2, \omega_3, \ldots \)' on bases 'C, A, T, G and E', we establish for instance:

\[
E \cdot \omega_1 = C, \quad C \cdot \omega_1 = A, \quad A \cdot \omega_2 = G, \quad T \cdot \omega_3 = C, \quad G \cdot \omega_1 = E.
\]
In continuance, the set $P_\omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_0 (= \omega_5)\}$ is readily confirmed to form group $\{P_\omega, \cdot\}$ where the identity element is $\omega_0$ and the inverse of $\omega_m$ is $\omega_m^{-1}$:

$$\begin{align*}
\omega_0^{-1} &= \omega_0, \\
\omega_1^{-1} &= \omega_4, \\
\omega_2^{-1} &= \omega_3, \\
\omega_3^{-1} &= \omega_2, \\
\omega_4^{-1} &= \omega_1, \\
\omega_5^{-1} &= \omega_0 = 0.
\end{align*}$$

(11)

Closure and associativity follow from (9) and (10).

Here, if we turn our attention to the wallpaper pattern, a further bijection obeying the postulates of the wallpaper group can be confirmed. Corresponding to Figures 3 and 6 a bijection between the Cayley Tables for translational and rotational operations can be established:

$$\begin{align*}
r &\leftrightarrow \omega = \omega_1, \\
u &\leftrightarrow 2\omega = \omega_2, \\
d &\leftrightarrow 3\omega = \omega_3, \\
l &\leftrightarrow 4\omega = \omega_4, \\
n &\leftrightarrow 5\omega = \omega_5 = \omega_0 = 0.
\end{align*}$$

(12)

Naturally, inverses (e.g., $\omega_2^{-1} = \omega_3$) are preserved in accordance with the inverses for ‘r, u, d, l, and n’. Any right translation of the horizontal line in Figures 1 and 6
is also expressible as a rotation over the fivefold phasor diagram in Figure 5 (rotational group). Thus, these are able to be regarded as a synthesized form of the wallpaper style (wallpaper group) from which expressions such as

\[ A^\circ r = A^\circ \omega_1 = T = E^\circ d \]

and

\[ A^\circ l = A^\circ \omega_4 = C = E^\circ \omega_0 \]. Transitions in the four directions of the cruciform are expressible as rotations of the fivefold phasor diagram in the Gaussian plane. In other words, the linear operation and the rotational operation of the five bases are synthesized into a unique scheme (wallpaper pattern).

### §4 Methods to obtain complementary sequences from primary DNA

Suppose, from among 'C, A, T, G and E' a base 'X_m' is given; its complementary base 'X_m\text{'}' to 'X_m' is defined as follows; for 'X_m = \{\exp(m \cdot \omega \cdot i)\}', m = \{0, 1, 2, 3, 4, 5\}, then 'X_m\text{'}' is obtained by 'X_m\text{'} = \{\exp((5 - m) \cdot \omega \cdot i)\}', where '\{\exp(5 \cdot \omega \cdot i)\} = \{1\} = E\text{'}'. In this regard,

\[ X_3\text{'} = X_0 = X_0\text{'} = X_5 \]  

(13)

The procedure yields specifically 'A\text{' = T' and 'C' = G'.

Clearly, the complement of 'E' is 'E' itself; 'E\text{'} = E'.

Figure 6 Scheme for an accessorized wallpaper pattern synthesized from the linear group and rotational group for the five bases. A bijection exists between the primitive operations of both groups: 'r ↔ \omega_1, u ↔ 2\omega, d ↔ 3\omega, l ↔ 4\omega, n ↔ \omega_0 = 0'. Transitions in the four directions of the cruciform are expressible as rotations of the fivefold phasor diagram in the Gaussian plane.
Another notation for the ‘\(X_m\)’ expressed as a base can be given. We introduce the one-value function ‘\(\omega X(m)\)’ that provides the same results,

\[
'X_m=\omega X(m) = E \cdot m = (\omega \bullet \omega \bullet \ldots \bullet \omega \text{ (m times)})' (m = 0, 1, 2, ..., 5).
\]

(14)

As for ‘\(m\)’ in (14), both positive and negative integers are permissible. Thus, ‘\(X_m\)’ is expressible as

\[
X_m^+ = 'X(5-m) = E \cdot (5-m) = (\omega \bullet \omega \bullet \ldots \bullet (5-m) \text{times})' (m = 0, 1, 2, 3, 4, 5).
\]

(15)

A simple example is illustrated below.

Suppose ‘\(D_1\)’ = [\(A_1|T_2|C_3|E_4|G_5|T_6|\ldots\] = \( [^\omega X(2)^\omega X(3)^\omega X(4)^\omega X(5)] \ldots \), then,

\[
'D_1' = [^\omega X(5-2)^\omega X(5-3)^\omega X(5-4)^\omega X(5-3)] \ldots ,
\]

= \( [^\omega X(3)^\omega X(2)^\omega X(4)^\omega X(0)^\omega X(1)^\omega X(2)] \ldots ,
\]

= \( [T_1|A_2|G_3|E_4|C_5|A_6|\ldots\] in accordance with the wallpaper group in Figure 1, the translations in one direction (e.g., right) over a horizontal line form a cyclic group \(P_r\) that contains only \(\{r, r^2, r^3, r^4, r^5\}\). This group is isomorphic with group \(P_x\) = \(\{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\ (= \omega_0)\}\), as is the group similarly generated over a vertical line.

Similar to ‘\(\omega X(m)\)’, ‘\(X_m\)’ can be expressed using another one-value function ‘\(X(m) = E \cdot r^m\)’:

\[
X_m = 'X(m) = E \cdot r^m = (r \bullet r \bullet \ldots \bullet r \text{ (m times)}) (m = 0, 1, 2, 3, 4, 5).
\]

(17)

Hence, ‘\(X_m^+\)’ (the complementary base of ‘\(X_m\)’) is written as

\[
X_m^+ = 'X(5-m) = E \cdot r^ {-m} = (r \bullet r \bullet \ldots \bullet r \text{ (5-m times)}).
\]

(18)

Extension to vertical translations is straightforward;

\[
'X_m=\omega X(5) = E \cdot u^5 = (u \bullet u \bullet \ldots \bullet u \text{ (m times)}).
\]

(19)

and its complementary base ‘\(X_m^+\)’ can be identified similarly although the order of letters are somewhat different.

Consider the following simple example in identifying ‘\(D_1\)’ using ‘\(X(m)\)’;

for ‘\(D_1\)’ = [\(A_1|T_2|C_3|E_4|G_5|T_6|\ldots\] = \( [X(2)|X(3)|X(1)|X(0)|X(4)|X(3)] \ldots \),

by replacing ‘\(\omega X(m)\)’ by ‘\(X(m)\)’ in formula (15), the same result is obtained.

According to these rules, ‘\(X_m^+ \bullet b[X_{m1} \rightarrow X_{m2}]\)’ = \(X_1^r r^3 = E \cdot r^4 = G\). In general, when the i-th component ‘\(b[X_{m1} \rightarrow X_{m2}]\)’ of ‘\(B_{(m1 \rightarrow m2)}\)’ changes \(X_{m1}\) (= \(E \cdot r^m = E \cdot (m_1 \omega)\)) to \(X_{m2}\) (= \(E \cdot r^{m_2} = E \cdot (m_2 \omega)\)).
Hence, the highlighted form of the operator vector is expressed as

\[ B_{(m_1 \rightarrow m_2)} = [...|b_{Xm_1 \rightarrow Xm_2}|...] = [...|r^{m_2-m_1}|...] = [...|(m_2-m_1)\omega|...]. \]  

(20)

For a further example, given the operator ‘\(B_{(j \rightarrow k)}\)’ that changes \(D_j\) to \(D_k\),

\[ D_j = \{C_1|A_2|E_3|...|C_i|...|T_{N-1}|A_N|E_{N+1}|E_{N+2}|E_{N+3}|...\}, \]

\[ = [E\cdot r_1]^1[E\cdot r_2]^2[E\cdot r_3]^3[...][E\cdot r_1]^1[E\cdot r_2]^3[E\cdot r_3]^N\cdot E\cdot r_0^N+1\cdot E\cdot r_0^N+2\cdot E\cdot r_0^N+3|...], \]

\[ D_k = \{G_1|T_2|C_3|...|G_i|...|C_{N-1}|T_N|E_{N+1}|E_{N+2}|E_{N+3}|...\}, \]

\[ = [E\cdot r_1]^1[E\cdot r_2]^2[E\cdot r_3]^3[...][E\cdot r_1]^1[E\cdot r_2]^N\cdot E\cdot r_0^N+1\cdot E\cdot r_0^N+2\cdot E\cdot r_0^N+3|...]. \]

With details shown in Appendix D, ‘\(B_{(j \rightarrow k)}\)’ takes the form

\[ B_{(j \rightarrow k)} = [r_1^3|r_2^3|...|r_3^3|...|r_{N-1}^1|r_N^0|r_{N+1}^0|r_{N+2}^0|r_{N+3}^0|...]. \]

Naturally, the state \(D_k\) is obtained through recursively applying the operations, \(D_j \cdot B_{(j \rightarrow k)} = D_k\) (Details are presented in Appendix D).

Whereas ‘\(D_j\)’ might have components in reverse order in terms of sense (5’ or 3’), there exists however certain ‘\(D_\text{pr}\)’ such that ‘\(D_k = D_\text{pr}\)’, (\(j, k = 0, 1, 2, 3, 4,...\)’). With this, ‘\(D_j\)’ is one of the ordinal elements belonging to the same set D. Thus, the symbol ‘\(\text{‘}\)’ need only be present when elements are distinct.

§5 Further unifying notation to describe the wallpaper group operation

Consider Figure 6; we assume that the number of right translations ‘\(r\)’ in group \(P_n\) (or \(\omega\) in group \(P_\omega\)) is ‘a’ and the number of up translations ‘\(u\)’ in group \(P\) (or 2\(\omega\) in group \(P_\omega\)) is ‘b’ with a, b = ..., 2, -1, 0, 1, 2,... Similarly, with ‘\(d \leftrightarrow 3\omega\)’, ‘\(l \leftrightarrow 4\omega\)’, the total change can be summarized as ‘\(x[a, b]\)’. We can confirm that there exists at least a pair of ‘a, b’ that satisfies

\[ 'X = E\cdot x[a, b]' , \]  

(21)

because any base in Figure 6 can be obtained by a finite number of transitions from ‘E’.

For instance, A can be expressed as; ‘\(A = E\cdot x[0, 1] = E\cdot x[2, 0] = E\cdot x[1,3] = E\cdot x[2,3] = E\cdot x[4,4]\)’. However, we remark that ‘\(x[a, b]\)’ means changes of bases from one to another prescribed by the wallpaper pattern. In practice, ‘\(x[a, b] = r^a\cdot u^b\) constitutes a multiple composition of elements of group \(Z_5\). In addition, ‘\(x[-a, -b] = r^a\cdot u^{-b} = l^b\cdot d^{-b}\) or \(\leftrightarrow (-a)\omega\cdot(b)(2\omega) = (-a-2b)\omega\)’. E.g., ‘\(x[-3, -2]\) means ‘\(r^{-3}\cdot u^{-2} = l^3\cdot d^{-2}\) or \(\leftrightarrow (-3)\omega\cdot(-2\cdot 2)\) (\(\omega = (-3-4)\omega = (-7)\omega = (-2)\omega = 3\omega = \omega_3\)’.

For the wallpaper group, the ‘\(a\)’ and ‘\(b\)’ should be interpreted in modulo 5 addition. The Cayley table for the wallpaper group are presented in Appendix A.

Within ‘the square unit cell’ in Figure 1 or 6, there are five pairs of ‘\(a, b\)’ for each base, as for ‘\(A\)’. Under modulo 5 addition, ‘\(x[a+5, b+5] = x[a, b]\)’ holds. Moreover, if ‘\(X\)’ is obtained from ‘\(X\)’ using ‘\(X = E\cdot x[a, b]\)’, ‘\(X\)’ can be determined as

\[ 'X' = E\cdot x^{-1}[a, b] = E\cdot x[-a, -b]' , \]  

(22)

or

\[ 'X' = E\cdot x[5-a, 5-b]' , \]  

(23)

because ‘\(X\)’ and ‘\(X\)’ are symmetrically disposed with respect to ‘\(E\)’ over the wallpaper pattern that would be selected as a standard for the definition of ‘\(a, b\)’. In practice, for
an arbitrary ‘X’, ‘X’ can be obtained via (22) or (23) by making use of an arbitrary ‘E’ as the reference point for the symmetry.

For example, if \( G = E\cdot x[1,2] \), then according to (22) \( G' = E\cdot x^{-1}[1,2] = E\cdot x[-2, -1] = C \) or according to (23) \( G' = E\cdot x[1,2] = E\cdot x[3,4] = C' \). There are an infinite number of identifications for the complementary base for an arbitrary base ‘X’.

Moreover, if we define the composition for the ‘x[a, b]’s as

\[
\{x[a_1, b_1]\cdot x[a_2, b_2]\} = x[a_1, b_1] + x[a_2, b_1] = x[a_1 + a_2, b_1 + b_2],
\]

we can confirm descriptions (20), (22) and (23). As for the operators ‘B_{m}’, with ‘D_{j}’ expressed as \( D_{j} = [...|E\cdot x[a_i, b_i]|...] \), one of the candidates of the appropriate ‘B_{(j\rightarrow j)}’s that produces ‘D_{j}\cdot B_{(j\rightarrow j)} = D_{j}' \) is identified:

\[
B_{(j\rightarrow j)} = [...|x[-2a_i, -2b_i]|...] = [...|r^{-2ai}\cdot u^{-2bi}|...] = [...|l^{2ai}\cdot d^{2bi}|...],
\]

(25a)

or \[
[...|r^{-2ai}\cdot (r^2)^{-2bi}|...] = [...|r^{-2ai-4bi}|...] = [...|l^{2ai+4bi}|...]. \quad \text{(using ‘u = r^2, ‘d = l^2’),}
\]

(25b)

or \[
[...|(-2a_i)\cdot (-2b_i)\cdot (2a_i)|...] = [...|(-2a_i-4b_i)\omega_i|...]. \quad \text{(using ‘r^m\leftrightarrow mo’, ‘u^m\leftrightarrow 2mo’).}
\]

(25c)

The exponents ‘-2a_i -4b_i’ in (25b) are permitted to take positive or negative integer values.

In these expressions, the rules for the wallpaper group (25a) can also be expressed as either for the linear group or for the rotational group (25b or 25c).

More generally,

‘D_{j} = [...|E\cdot x[a_{(j)}, b_{(j)}]|...]’ is changed into ‘D_{k} = [...|E\cdot x[a_{(k)}, b_{(k)}]|...]’; and

‘B_{(j\rightarrow k)}’s that provides ‘D_{j}\cdot B_{(j\rightarrow k)} = D_{j}' \) is identified as

\[
B_{(j\rightarrow k)} = [...|x[a_{(k)}-a_{(j)}], b_{(k)}-b_{(j)}]|...] = [...|x^{a(k)-a(j)}\cdot iy^{b(k)-b(j)}]|...],
\]

(26a)

Also, \[
[...|r^{a(k)-a(j)}\cdot (r^2)^{b(k)-b(j)}]|...] = [...|r^{a(k)-a(j)+2b(k)-2b(j)}]|...].
\]

(26b)

or else \[
[...|a_{(k)}-a_{(j)}|\omega|2(b_{(k)}-b_{(j)})\omega_k|...] = [...|a_{(k)}-a_{(j)}+2b_{(k)}-2b_{(j)}|\omega_k|...].
\]

(26c)

As mentioned in §1, if a certain sequence ‘X’ has sense \(<5' \rightarrow 3'>\), the complementary sequence ‘X’ of a certain sequence ‘X’ is reversed to \(<3' \rightarrow 5'>\).

To aid understanding, we present the following examples: Given
\[ D_j = [A_j \mid T_j \mid E_3 \mid \ldots \mid G_{N-1} \mid A_N \mid E_{N+1} \mid E_{N+2} \mid E_{N+3} \ldots] \]
\[ = [E \cdot x[0, 1] \mid E \cdot x[0, -1]_j \mid E \cdot x[0, 0] \mid \ldots \mid E \cdot x[1, 0] \ldots] \]
\[ \ldots [E \cdot x[-1, 0]_{N-1} \mid E \cdot x[0, 1]_N \mid E \cdot x[0, 0]_{N+1} \mid E \cdot x[0, 0]_{N+2} \mid E \cdot x[0, 0]_{N+3} \ldots] \]

then, according to (22), \( D_j^* \) is simply
\[ D_j^* = [E \cdot x[-1, 0]_{N-1} \mid E \cdot x[0, 1]_N \mid E \cdot x[0, 0]_{N+1} \mid E \cdot x[0, 0]_{N+2} \mid E \cdot x[0, 0]_{N+3} \ldots] \]
\[ = [E \cdot x[0, -1]_j \mid E \cdot x[0, 0]_j \mid E \cdot x[0, 0]_j \ldots] \]
\[ \ldots [E \cdot x[-1, 0]_{N-1} \mid E \cdot x[1, 0]_N \mid E \cdot x[0, -1]_N \mid E \cdot x[0, 0]_{N+1} \mid E \cdot x[0, 0]_{N+2} \mid E \cdot x[0, 0]_{N+3} \ldots] \]
\[ = [T_j \mid A_2 \mid E_3 \mid \ldots \mid G_{N-1} \mid T_N \mid E_{N+1} \mid E_{N+2} \mid E_{N+3} \ldots] \]

If we use the optional formula (25a – c), the relation \( \text{‘} D_j \text{‘} B_{\langle j \rightarrow k \rangle} = D_j^* \) is derived. Details are given in Appendix E.

Apart from these examples, additional identities for the wallpaper group can be verified using Figure 1 or 6; e.g.,
\[ x[1, 0]_m x[0, 1] = x[1, 0] + x[0, 1] = x[1, 1] = x[0, -1] \]
\[ x[2, 0] \]}
\[ = r \cdot u \cdot (r \cdot u) = u \cdot u \cdot u = 1 \]

We develop various general formulas:
\[ x[a + 5, b] = x[a, b] \]
\[ x[2a, -a] = x[0, 0] \]
\[ x[2a, 2a] = x[0, 0] \]
\[ x[-2a, a] = x[0, 0] \]
\[ x[-a, -2a] = x[0, 0] \]
\[ (27) \]

Other unknown rules might underlie the wallpaper pattern.

Concerning style in treating the wallpaper group, examples \( X_m = X(m) \) in (16, 17, 19), and \( \text{‘} X_m \text{‘} \) (m) in (14, 15, 20) could be regarded as a specific combination that are displayed as
\[ X_m = X(m) = X \cdot x[a, 0] \]
\[ (a = \ldots, -2, -1, 0, 1, 2, \ldots, 5, 6, \ldots; \text{integer}) \]
\[ (28) \]

**§6 Treatment of changes of sequences and the insertion/deletion of DNA bases via an optionally generalized operation**

Below, we demonstrate, using several examples containing \( E \)'s, changes and inclusion/exclusion of DNA bases using a more generalized scheme.

For definiteness, let \( D_j \) be the sequence ‘CGTAT...C...TA’; we consider the change of its ‘1–3’ components ‘\( C_1G_2T_3 \)’ into ‘\( G_1T_2A_3 \)’, and moreover the insertion of two bases ‘GC’ between ‘\( T_3 \)’ and ‘\( A_4 \)’ denoted ‘(‘)’;
\[ D_1 = [C_{\langle j \rangle} \mid G_{\langle j \rangle} \mid T_{\langle j \rangle} \mid (A_{\langle j \rangle} \mid \ldots \mid C_{\langle j \rangle} \mid T_{\langle j \rangle} \mid A_{\langle j \rangle} \mid E_{\langle j \rangle} \mid E_{\langle j \rangle+1} \mid E_{\langle j \rangle+2} \mid E_{\langle j \rangle+3} \ldots] \]

We denote the result of this transformation as \( D_k \),
\[ D_k = [G_{\langle k \rangle} \mid T_{\langle k \rangle} \mid A_{\langle k \rangle} \mid G_{\langle k \rangle} \mid C_{\langle k \rangle} \mid A_{\langle k \rangle} \mid \ldots \mid C_{\langle k \rangle+2} \mid T_{\langle k \rangle} \mid A_{\langle k \rangle} \mid E_{\langle k \rangle+1} \mid E_{\langle k \rangle+2} \mid E_{\langle k \rangle+3} \mid E_{\langle k \rangle+4} \mid E_{\langle k \rangle+5} \ldots \]
\[ (29) \]

The procedure from \( D_j \) to \( D_k \) is described recursively to find operator ‘\( B_{\langle j \rightarrow k \rangle} \)’. 

Sawamura et al. Theoretical Biology and Medical Modelling 2014, 11:18  
http://www.tbiomed.com/content/11/1/18 
Page 16 of 32
First, two ‘E’s are inserted after the 3rd component (this change is denoted ‘$D_j \rightarrow D_h$’ in preparation for insertion of ‘GC’):

$$D_j \rightarrow D_h = [C_{(h)1}]G_{(h)2}T_{(h)3}[E_{(h)4}E_{(h)5}]A_{(h)4+2}…[C_{(h)i+2}]…$$

$$…T_{(h)N-1+2}[A_{(h)N+2}[E_{(h)N+2+2}E_{(h)N+3+2}]…].$$

Thus, ‘$b_{(j \rightarrow h)}$’

$$= [b_{[C \rightarrow C]}[b_{[G \rightarrow G]}]b_{[T \rightarrow T]}[b_{[E \rightarrow E]}[b_{[C \rightarrow C]}])b_{[A \rightarrow A]}[b_{[G \rightarrow G]}]…]b_{[C \rightarrow C]}[b_{[T \rightarrow T]}[b_{[A \rightarrow A]}[b_{[G \rightarrow G]}]…].$$

This change is in accordance with those rules for vector-like ‘$D_j$’s dependent upon ‘E’s.

Hence, the operator $b_{(h \rightarrow k)}$ that produces the change from $D_h$ to $D_k$ is described as:

$$b_{(h \rightarrow k)} = [b_{[C \rightarrow C]}[b_{[G \rightarrow G]}]b_{[T \rightarrow T]}[b_{[A \rightarrow A]}[b_{[E \rightarrow E]}[b_{[C \rightarrow C]}])b_{[A \rightarrow A]}[b_{[E \rightarrow E]}[b_{[A \rightarrow A]}]…]b_{[C \rightarrow C]}[b_{[T \rightarrow T]}[b_{[A \rightarrow A]}[b_{[E \rightarrow E]}[b_{[E \rightarrow E]}[b_{[E \rightarrow E]}[b_{[E \rightarrow E]}[b_{[E \rightarrow E]}[b_{[E \rightarrow E]}]…]]]]]]]]]]]]]]]].$$

Thereby,

$$D_h \cdot b_{(h \rightarrow k)} = [C_{(k)1}]b_{[C \rightarrow C]}[G_{(k)2}b_{[G \rightarrow G]}[T_{(k)3}b_{[T \rightarrow T]}[A_{(k)3}b_{[E \rightarrow E]}[E_{5}b_{[E \rightarrow E]}[C_{(k)5}b_{A \rightarrow A}]…]A_{(k)6}b_{[A \rightarrow A]}[b_{[E \rightarrow E]}[b_{[C \rightarrow C]}])…]b_{[C \rightarrow C]}[b_{[T \rightarrow T]}[b_{[A \rightarrow A]}[b_{[G \rightarrow G]}]…][T_{N+1}b_{[T \rightarrow T]}[b_{[E \rightarrow E]}[b_{[N+2]}b_{[E \rightarrow E]}[b_{[E \rightarrow E]}[b_{[E \rightarrow E]}[b_{[E \rightarrow E]}[b_{[E \rightarrow E]}]…]].$$

With reference to Figure 1, 2, 6 or Appendix A, (32)

$$= [C_{(k)1}d_{(k)1}[G_{(k)2}d_{(k)2}]T_{3}d_{(k)3}E_{(k)4}L_{(k)4}E_{5}r_{5}A_{(k)5}n_{6}…[C_{(k)1+2}n_{k+2}]…$$

$$…T_{N+1}n_{N+1}A_{N+2}n_{N+2}[E_{N+3}n_{N+3}E_{N+4}n_{N+4}E_{N+5}n_{N+5}]…]].$$

$$= [G_{(k)2}T_{2}A_{3}[G_{(k)4}C_{(k)5}A_{(k)6}…[C_{(k)2]}…]T_{N+1}A_{N+2}[E_{N+3}E_{N+4}E_{N+5}…]].$$

$$= D_k(29).$$

This indicates a code change of the ‘1–3’ components and a ‘GC’ insertion after the 3rd as described via the two steps: 1) $D_j \rightarrow D_h$ (inserting two ‘E’s after the ‘3rd’ component), and 2) $D_h \cdot b_{(h \rightarrow k)} = D_k$. Note that the exclusion of the ‘4–5’ components ‘GC’ from $D_h$ and the transformation of the ‘1–3’ components from ‘GTA’ to ‘CGT’ constitute the recursive procedure for the inverse operator

$$b_{(k \rightarrow h)} = b_{(h \rightarrow k)}^{-1}. \quad (35)$$

Alternatively, ‘$D_h \rightarrow D_j$’ is obtained by deleting the two ‘E’s from the ‘4–5’ components of $D_h$ to yield the initial state ‘$D_j$’ in accordance with the characteristics of the vector-like ‘$D_j$’s.

In summary, essentially, all transitions (changes and inclusion/exclusion) of a certain sequence within the same single-stranded DNA, whether it has finite or infinite length, can be described in principle within a single operation using only the unique operator $b_{( \rightarrow \rightarrow \rightarrow \rightarrow )} \in \text{group } B$. 
§7 Synthesis of changes, insertion/deletion, and recombination of DNA bases

As a further development, to demonstrate recombination, take two finite sequences ‘GETAGT (= Dc1)’ and ‘ATAGCTA (= Dd1)’. These have vector expressions

\[ Dc1 = [G1E2T3A4G5T6 | E7E8E9…], \]
\[ = [G(c1)_1]E(c1)_2 [T(c1)_3]A(c1)_4G(c1)_5[T(c1)_6]E(c1)_7|E(c1)_8|E(c1)_9…], \]
\[ Dd1 = [A1T2A3G4C5T6A7 | E8E9E10…], \]
\[ = [A(d1)_1]T(d1)_2|A(d1)_3|G(d1)_4C(d1)_5|T(d1)_6|A(d1)_7|E(d1)_8|E(d1)_9|E(d1)_10…]. \]

To illustrate for the pair Dc1 and Dd1, we consider recombination to take place between the sequence ‘T3A3G3T6’ of the (3–6)-th component of ‘Dc1’ and the ‘AGCTA’ of the (3–7)-th component of ‘Dd1’ at the same instant.

First, in the pair of sequences, a series of ‘E’s of complementary size is inserted in ‘Dc1’ just before the sequence to be converted, and in ‘Dd1’ just after the sequence to be converted. For example, for ‘Dc1’, five ‘E’s, ‘EEEE’, of size equivalent to that of ‘A3G4C5T6A7’ of ‘Dd1’, are inserted just before ‘T3’ in ‘Dc1’ where ‘A3G4C5T6A7’ is to be located, that is, the interval between the 2nd ‘E2’ and 3rd ‘T3’ within ‘Dc1’. Under this procedure, Dc1 changes into Dc2:

\[ Dc2 = [G1E2(EEEE)T3+5A4+5G5+5T6+5 | E7+5E8+5E9+5…], \]
\[ = [G1E2(EEEE)T8A9G10T11 | E12E13E14…]. \]

Here, we assume that ‘EEEE’ is changed into ‘A3G4C5T6A7’ (originally, the (3–7)-th component of ‘Dd1’). In addition, ‘T8A9G10T11’ is transformed into the same number of ‘E’s, ‘EEEE’, at the same time. By this process, ‘Dc2’ changes in ‘Dc3’:

\[ Dc3 = [G1E2(A3G4C5T6A7)E8E9E10E11 | E12E13E14…]. \]

Note that bold type and underline are here merely pedagogical aids to help identify sequence changes. Meanwhile, four ‘E’s ‘EEEE’ equivalent in size to ‘T8A9G10T11’ of ‘Dc3’ would be inserted after ‘A3’ of ‘Dd1’, where ‘T3A3G3T6’ of ‘Dc1’ is to be located within ‘Dd1’. That is, ‘T3A3G3T6’ is inserted into the interval between the 7th ‘A7’ and 8th ‘E8’ within ‘Dd1’. In this procedure, Dd1 changes into Dd2:

\[ Dd2 = [A1T2A3G4C5T6A7(EEEE)E8+4E9+4E10+4…], \]
\[ = [A1T2A3G4C5T6A7(EEEE)E12E13E14…]. \]

Furthermore, we change ‘EEEE’ into the equivalent-sized ‘T8A9G10T11’ (originally, the (3–6)-th components of ‘Dc1’) while ‘A3G4C5T6A7’ is transformed into the equivalent-sized ‘EEEE’. Through this procedure, ‘Dd2’ changes in ‘Dd3’:

\[ Dd3 = [A1T2E5E6E7E8(T8A9G10T11)E12E13E14…] \]

As a result, if we omit the infinite series of ‘E’s from right end, we have the recombination (partial conversion between this pair of sequences from ‘Dc1’, ‘Dd1’) with ‘Dc1’ = ‘G1E2T3A4G5T6’ being transformed into ‘Dc2’ = ‘G1E2A1A3G4C5T6A7’ and ‘Dd1’ = ‘A1T2A3G4C5T6A7’ being transformed into ‘Dd3’ = ‘A1T2T8A3G5T6’. We
define the manipulation of the recombination (partial/total conversion) between
‘D_{c1}, D_{d1}’ in this way.

In the initial stage in the previous illustration, we inserted different sizes of ‘E’ se-
quences in each line; however, processes ‘D_{c1} → D_{d2}’ and ‘D_{d1} → D_{d2}’ are preferred to
be regarded as ‘E’ insertions/deletions (see comments prior to equation (5)) and this
rule depends upon the characteristics of these vectors (e.g., ‘D’/’s).

As previously explained, the operations can be performed in any of the three equivalent
linear group, rotational group, and wallpaper group. Choosing the wallpaper group,

\[
\begin{align*}
D_{c2} \cdot B_{(c2 \rightarrow c3)} &= [G_1 \cdot n_1; E_2 \cdot n_2; (E_3 \cdot u_3); E_4 \cdot l_4; E_5 \cdot r_5; E_6 \cdot d_6; E_7 \cdot u_7; T_8 \cdot u_8; |A_9 \cdot d_9; G_{10} \cdot r_{10}; |T_{11} \cdot u_{11}; |E_{12} \cdot n_{12} ] \\
&= [G_1; E_2(A_3); G_4; C_5; T_6(A_7); E_8; E_9; E_{10}; E_{11}; E_{12}; E_{13}; E_{14}; ...], \\
&= D_{c3}, \\
&\text{where } B_{(c2 \rightarrow c3)} = [n_1; n_2; (u_3; l_4; r_5; d_6; u_7); u_8; r_{10}; u_{11}; n_{12}; n_{13}; n_{14}; ...].
\end{align*}
\]

Also,

\[
\begin{align*}
D_{d2} \cdot B_{(d2 \rightarrow d3)} &= [A_1 \cdot n_1; T_2 \cdot n_2; A_3 \cdot d_3; G_4 \cdot r_4; C_5 \cdot l_5; T_6 \cdot u_6; A_7 \cdot d_7 \]
&= [E_2 \cdot d_8; E_9 \cdot u_9; E_{10} \cdot l_{10}; E_{11} \cdot d_{11}; E_{12} \cdot n_{12}; E_{13} \cdot n_{13}; E_{14} \cdot n_{14}; ...], \\
&= [A_1; T_2; E_3; E_4; E_5; E_6; E_7; (T_8; A_9; G_{10}; T_{11}); E_{12}; E_{13}; E_{14}; ...], \\
&= D_{d3}, \\
&\text{where } B_{(d2 \rightarrow d3)} = [n_1; n_2; d_3; r_4; l_5; u_6; d_7; (d_8; u_9); l_{10}; d_{11}; n_{12}; n_{13}; n_{14}; ...].
\end{align*}
\]

With respect to (42) and (44), the inverse identities are confirmed:

\[
B_{(c2 \rightarrow c3)}^{-1} = B_{(d2 \rightarrow d3)}. \tag{46}
\]

Generally, \(B_{(\rightarrow \rightarrow \rightarrow)}\) giving transition ‘\(D_{c2} \rightarrow D_{c3}\)’ automatically produces an inverse
change for ‘\(D_{d2} \rightarrow D_{d3}\)’, as stated in (46) and reduces troublesome manipulations,
even if only partially.

§8 Further applications of the composition category-like prototypal model
using additional ribonucleic acid (RNA)

We next comment on other possible applications of the model. The category theory-
like construction for treating DNA transcription to RNA might be conceivable, and the
combination of the set and the group can comprise a category when these satisfy cat-
egory theory postulates [48,49]. That is because we believe that in future developments
the discussion should embrace category theory as one of the important options.

To begin, according to our description for handling ‘E’ s, it seems difficult to define inverse
elements in a group theoretical way when there are deletions of ‘E’ s from any place in a se-
quence because we cannot find sufficient numbers of ‘E’ s in the target component of \(D\).

Thus, we consider the ‘morphism \(f\)’ that transforms the sequence of DNA bases
within set \(D\) as follows [48,49].

morphism \(f : X \rightarrow X\), \(\text{dom}(f) = D_p, \text{cod}(f) = D_k\). Object ‘\(X\)’ is the set of ‘\(D\)’/’s. There
exists a morphism ‘\(1_X\)’ such that ‘\(1_X \cdot f = f = f \cdot 1_X\)’ for every ‘morphism \(f\)’, when ‘\(1_X =
\left[ n_1; |n_2; n_3; |n_4; ... |n_{N-1}; n_N; ... \right] \in \text{group } B\). If supplemented, the ‘morphism \(f\)’
comprise ‘group B’ (see reference list in Figure 7). The group composition for ‘f1’ and ‘f2’ is denoted ‘f1 • f2’.

As mentioned earlier, sequences of DNA consisting of bases ‘C, A, T, and G’ are transcribed into RNA consisting of ‘C, A, U, and G’. This process can be regarded as the combination of two manipulation; I) transcription from the original DNA sequences (Dj) to those of its complement (Dj†), and II) alternation from ‘C, A, T, and G’ to ‘C, A, U, and G’ (Dj† → Rj†) (both are illustrated in Figure 8).

**Step I**
The transformation from the original DNA bases ‘Dj’ into the complementary sequence ‘Dj†’ (e.g., ‘TCATEAGCTGA...’ → ‘AGTAETCGACT...’) (for transcription to pre-messenger RNA (pre-mRNA) before splicing) can be performed via the manipulation (13–15, 17, 18, 20, 22, 23, 25a–c) in §3 and §4. ‘Dj†’ can be obtained via the linear group (17, 18, 20), the rotational group (14, 15, 20) and also the wallpaper group (21–23, 25a–c). Thereby, morphism ρ : X → Y, dom(ρ) = Dj, cod(ρ) = Dj†. Object ‘Y’ is the set of ‘Dj†’s (essentially equivalent to the set of ‘Dj’s).

There exist morphisms ‘1X’ and ‘1Y’ such that ‘1Y • ρ = ρ = ρ • 1X’ for ‘morphism ρ’, where

\[1X = 1Y = \{n_1, n_2, n_3, ..., n_i, ..., n_{N-1}, n_N\}^T\]  

(47)

However, in practice, morphism ρ is one of the ‘Bm’s ∈ group B (see Figures 7 and 8).

---

**Figure 7 Definition of category C.** A simplistic definition of category C treating the traditional “central dogma” is presented. In practice, any selection of operations (morphisms) is permissible because composition within the category C using any of the three operations belonging to the ‘linear group’, ‘rotational group’ or ‘wallpaper group’ is considered possible. The morphisms f and h correspond to elements ‘Bm’s ∈ group B, the only difference being the ‘T’s and ‘U’s. Actually, morphism j is regarded as part of morphism h (Rj†’s and ‘set R’ are also substitutable for ‘Rj’s and ‘set R’); all morphisms except for ‘r’ an ‘g’ satisfy the group postulates, and are treated as operations of group B.
Step II

Next, we define manipulations that change the above \( D_j^\dagger \) into \( R_j^\dagger \) where all \( T \)'s are converted into \( U \)'s; e.g., 
\[
D_j^\dagger = [A|G|T|A|E|T|C|G|A|C|T|C|G|A|C|T|…] \rightarrow (R_j^\dagger = [A|G|U|A|E|U|C|G|A|C|U|…]).
\]

This process can also be expressed in a similar way as transcription.

\[
\tau : Y \rightarrow Z, \quad \text{dom}(\tau) = D_j^\dagger, \quad \text{cod}(\rho) = R_j^\dagger.
\]

Object \( Y \) is in set \( D \), and object \( Z \) is in set \( R \); however, object \( Z_j \) is in set \( R_j \), which is a part of set \( Z \). Over set \( R_j \), group operations are not definable at present.

\[
\text{morphism } \tau : Y \rightarrow Z, \quad \text{dom}(\tau) = D_j^\dagger, \quad \text{cod}(\rho) = R_j^\dagger.
\]

The only difference between \( D_j^\dagger \) and \( R_j \) is the appearance of \( T \) and \( U \) in the sequences.
Naturally, for RNA base sequences, similar treatments are possible in the single group B: 

\[
\text{morphism } h : Z \rightarrow Z, \quad \text{dom}(h) = R_j^*, \quad \text{cod}(h) = R_s^* \quad \text{(Figures 7 and 8)}.
\]

There exists morphism \(1_Z\) such that \(1_Z \bullet h = h \bullet 1_Z\). 

Ordinarily, in prokaryotic cells, the DNA sequences are transcribed along their entire length. For eukaryotic cell, a splicing process is needed using nascent pre-messenger RNA (pre-mRNA) where introns of DNA bases are removed and exons are joined before producing a correct protein through translation, resulting in the mature messenger RNA (mRNA). Thus, the previous procedure was about the prokaryotic cell or the pre-translation of pre-mRNA in the eukaryotic cell. Therefore, to treat the products after RNA (mRNA). Thus, the previous procedure was about the prokaryotic cell or the pre-translation of pre-mRNA in the eukaryotic cell. Therefore, to treat the products after RNA (mRNA) where introns of DNA bases are removed and exons are joined before producing a correct protein through translation, resulting in the mature messenger RNA (mRNA). Thus, the previous procedure was about the prokaryotic cell or the pre-translation of pre-mRNA in the eukaryotic cell. Therefore, to treat the products after RNA (mRNA). 

For eukaryotic cell, a splicing process is needed using nascent pre-messenger RNA (pre-mRNA) where introns of DNA bases are removed and exons are joined before producing a correct protein through translation, resulting in the mature messenger RNA (mRNA). Thus, the previous procedure was about the prokaryotic cell or the pre-

If ‘GUA’ is removed from ‘A(GUA)EUCGACU...’ to become ‘A( )EUCGACU...’, this procedure can be described as; \(R_j^* \rightarrow R_s^*\);

\[
R_j^* = [A_1(G_2|U_3|A_4)E_5|U_6|C_7|G_8|A_9|C_{10}|U_{11}|E_{12}|E_{13}|E_{14}...],
\]

\[
R_s^* = [A_1(E_2|E_3|A_4)E_5|U_6|C_7|G_8|A_9|C_{10}|U_{11}|E_{12}|E_{13}|E_{14}...].
\]

The \(R_s^*\) form a set \(R_s = \{R_s^* (j = 1,2,3,...)\}\) that is a part of set \(R\) (see Figures 7 and 8).

Hereon, we admit \(E\)’s in the sequences of RNAs (as elements of set \(R_s\)) during the operations before morphism \(\Gamma\) and after morphism \(\psi\) to maintain theoretical consistency. Thus, if the result of a series of these maps is \(R_s^* = A_1E_2E_3E_4E_5U_6C_7G_8A_9C_{10}U_{11}E_{12}E_{13}E_{14}...\), then the actual RNA sequence should be interpreted as ‘AUC...’. Specifically, an equivalent-sized substitution of some bases in pre-mRNA with \(E\)’s can be written morphism \(\psi : Z \rightarrow Z_s\), \(\text{dom}(j) = R_j^*\), \(\text{cod}(j) = R_s^*\). There exists a morphism \(1_{Z_s}\) such that \(1_{Z_s} \bullet j = j \bullet 1_{Z_s}\).

\(\psi\) changes some series of bases from ‘C, A, U, G, E’ to an equivalent-sized series of ‘E’s within the partial operations of the group B. However, morphism \(\psi\) fails the group axioms, as inverse might not be definable.

Finally, as in §1, we apply the simultaneous deletions of all explicit \(E\)’s of mRNA other than the trailing \(E\)’s, the state after these deletions being denoted with ‘< >’; for

\[
R_s^* = [A_1E_2E_3E_4E_5U_6C_7G_8A_9C_{10}U_{11}E_{12}E_{13}E_{14}...]
\]

the description ‘<R_s^*> = [A_1U_2G_3G_4A_5C_6U_7...]’ is specified without explicit non-trailing ‘E’s. In this regard, as in §1, if some indels (insertions/deletions) occur at certain bases of ‘<R_s^*>’, as for ‘<R_s^*>> = [A_1U_2(E_3)G_4(U_5)C_6U_7...]’ (with the deletion of \(C_3\) and \(A_5\)), we state the result as ‘<<R_s^*>> = [A_1U_2G_3G_4A_5C_6U_7...]’. ‘R_j’s include <R_s^*> > s and ‘<<R_s^*>> > s from the set \(R\) and both still satisfy the postulates of group B. This rule is a relative postulate and explicit \(E\)’s are not absolutely forbidden in ‘<R_s^*>’s or ‘<<R_s^*>>’s, hence further indels of ‘E’s into ‘<R_s^*>’s or ‘<<R_s^*>>’s are not forbidden.

Also, omissions of explicit ‘E’s are considered as in ‘<R_0^*> = [A_1U_2G_4C_6U_7...], where place numbers ‘3’ and ‘5’ are absent indicating implicitly their presence in the vector. (Note that all products belong to group B.) Similarly, t-tuples of ‘< >’s are denoted ‘<<<<R_s^*>>>>’(t-tuple) = <R_s^*>, representing multiple deletions of ‘E’s (t-times).
Combinations of symbols ‘{’ and ‘< >’ are also allowed when necessary, as for example `{<[(<RS1>)]>}, as long as the subscripted place numbers are adequately recognized/traced.

Nevertheless, the multiple use of ‘< >’ to remove all ‘E’ in the vector ‘RS’ should have a unique meaning with regard to protein synthesis. As a result, the subsequent reading/translation in line with ‘codon’-like ‘[AUC|GAC|U...]’ leads in an ordinal way to a description of protein synthesis. Through the use of ‘{’ and/or ‘< >’, the concept ‘E’ may have benefits, although this may need to be intensely explored in future studies.

The procedure reversing transcription, found for example in retrovirus, is also describable if additional options are added to the scheme. However, these options are omitted at this stage to keep the model simple.

In summary, suppose we have a ‘category C’ with objects ‘X, Y, Z, Zs’ and morphisms ‘f, f′, f′′, g, h and ‘j’. We affirm that these definitions satisfy the postulates of category. A list is given in Figure 7. Indeed, morphisms other than ‘f′ and ‘g′ are simple group-theoretical products. One of the reasons we have introduced the concept ‘category’ is that the translation from single-strand DNAs to RNAs is difficult or impossible to systematize as a group structure. Therefore, if we identify the differences, we can treat all manipulations, except for ‘f′ and ‘g′, based simply on group B.

The expression ‘hom(X, Y)’ denotes all morphisms f: ‘from X to Y’. Likewise, ‘hom(X, Y)’ denotes all morphisms ρ: ‘from X to Y’ In addition, ‘hom(Y, Z)’ denotes all morphisms τ: ‘from Y to Z, and hom(X, Z) denotes all morphism g: from X to Z. Then, hom (Z, Z) denotes all morphisms h: ‘from Z to Z’. Finally, hom(Y, Z) denotes all morphism h: from Y to Z. (Details are displayed in Appendix F)

As is explained in §3 and §4, the rotational group can be regarded as a specific bijection of the vector group [2,44-47], so, we can describe this relationship naturally in a category theory-like way where two categories C1 and C2 are linked.

First, we consider two categories C1 and C2 with a ‘functor F’ from C1 to C2 written ‘F: C1 → C2’. For example, the pre-category C is denoted C1 and the product of functor F on category C1 is denoted C2 [48,49]. Note that the only difference between C1 and C2 is assumed to be the nature of its expression; morphism f1 = B1 (∈ category C1) is based on the rotational group in Figure 1 or 6; e.g.,

\[ B_1 = [r_1] [l_2] [u_3] ... [n_i] ... [r_{N-1}] [d_N] ... = \ldots [x[a_i, b_i, i, \ldots] = \ldots [r^{\text{ub}_i}] \ldots \in \text{group B}_1] \].

(54)

Additionally, morphism f2 = B2 (∈ category C2) is based on the rotational group over the Gaussian plane in Figure 5; e.g.,

\[ B_2 = \omega_1 \omega_2 \omega_3 \omega_0 \omega_1 \omega_{N-1} \omega_N \ldots = \ldots [\omega_0] \ldots \in \text{group B}_2] \).

(55)

With regard to the identity morphisms, we have

\[ \begin{align*}
1_{X_1} &= 1_{Y_1} = 1_{Z_1} = 1_{Z_1} = [n_1] [n_2] [n_3] \ldots [n_i] \ldots [n_{N-1}] [n_N] \ldots, \\
1_{X_2} &= 1_{Y_2} = 1_{Z_2} = 1_{Z_2} = \omega_1 \omega_2 \omega_3 \omega_0 \omega_1 \omega_{N-1} \omega_N \ldots. 
\end{align*} \]

(56)

Herein, we view ‘functor F: C1 → F(C1) (= C2)’ in following way [48,49].

(Details are shown in Appendix G)
Note that a similar definition like the composition of $C_1$ based on the linear group and $C_2$ based on rotational group is possible, being linked with ‘functor $F$’.

This is satisfied provided an adequate definition of ‘Functor $F$’ is given, and we presume that the morphisms described previously formulates a model that renders one of forms of the canonical “central dogma” proposed by Crick in 1958 [26].

In the transcription of RNA bases, the ‘RNA splicing’ process is well-known, whereby ‘intron sequences’ are excised, and ‘exon sequences’ are combined to condense effectual information for further interpretation in protein synthesis. However, as for further processing of the triplets of bases e.g., ‘ACG’ and ‘AUG’, a considerable number of models have been reported e.g., [12,18-21,33,34,50]. We refrain from pursuing this issue at present.

**Results**

We added an imaginary base ‘$E$’ to the set of actual DNA bases, and composed group $Z_5$ of basic translational operations on grid-points of a cruciform wallpaper pattern constructed of the five base letters. Moreover, using the same five letters, we integrated the wallpaper group as the combination of linear group over the horizontal line and the rotational group based on symmetries of a fivefold phasor diagram on the unit circle in the Gaussian plane. Additionally, changes in the sequences of the DNA bases are treated using set $D$, the set of all possible sequences of DNA bases that also contain ‘$E$’. Also, ‘$D_j$’s are drawn as polygonal lines graphically. Moreover, by combining group $Z_5$, the operators that rearrange bases of DNA sequences constitute the group $B$. Using these results, simple changes of sequences, insertions/deletions, and recombination of DNA bases are also treatable via a synthesis of group-theoretical operations between sets $D$ and group $B$. Together with this, all results obtained for DNA pertain to RNA by replacing T with U. Using these tools, category theory-like language is introduced to describe the canonical “central dogma” that is expected to integrate DNA-based processes, although the overall profile and range of applicability is unclear at this stage. Alternatively, by introducing the manipulations ‘$\{\}$’ and ‘$<>$’, operations on states of ‘$E$’s in ‘$D_j$’s’/’$R_j$’s, whether explicit or not, can be performed in parallel with the conventional description for DNA/RNA sequences.

**Discussion**

The issue treated in this article is, roughly speaking, the combination of two ideas: one is the wallpaper pattern in the context of DNA sequencing §1 – §7, and the other is the tentative development towards systematization of molecular/genetic biology in the style of some category-theory-like description §8. Essentially, the two are different topics although strongly connected. The former is an independent study on symmetry modeling of DNA sequences, whereas the latter can be re-expressed using different material as long as the basic elements can be treated within a category-theory-like model satisfying group theory-like postulates.

In this article, we considered a group/category theory-like treatment devising an expedient ‘$E$’, grid-point array, and group operations to move over the array. We discussed whether and how a more synthesized description can be constructed, using simplistic postulates of group. Next, we take the basics of category theory to describe processes, although this is only at a preliminary stage.
For an application of our ideas, we have chosen DNA sequencing from the perspective of not only coding sequences of DNA bases but also describing insertions/deletions of DNA sequences using a single operation that is an element of a group. The expedient \( E \) permits inserting and/or deleting sequences depending on the purpose. Specific notation was introduced so that vector-like DNA sequences and operations can be composed as a set and a group.

Ordinarily, a method to describe DNA sequences is often limited in scope by focusing on only one aspect such as recognizing each base sequentially (e.g., A, G, T, A, C… from ‘AGTAC…’) [13,33-35], where an operation like ‘rotation,’ ‘transition’ or ‘conversion’ based on a certain solid is often used. Another focus of attention is the rules for interpretation of codons in synthesizing proteins from DNA sequences. The rules are defined to capture the specific activity from the viewpoint of group-like operations [18-22].

It also enables us to treat three manipulations as one type of operation in the group, with easily-imaginable graphic displays such as Figure 4, although it is only an accessorial tool at this stage. Increasing the degree of freedom by one and integrating changes of coding and sequence recombination might yield some polysemous utility.

When inserting/deleting sequences of bases into the main DNA sequence, even if the endpoints of the base series are identified precisely, it appears that manipulations via \( E \)'s are not always necessary. Nonetheless, to determine the final order of the bases in these cases, we must track base changes from one to the next (including \( E \); even when lost or deleted). If we use the \( E \)-assisted manipulations for coding, we need only to examine the inclusions; the rest remains unchanged in order. Additionally, we assume that when any operation is performed, the position and number of \( E \)'s should be fixed so that the order of any component of ‘\( B \)' or ‘\( D \)' is not changed, at least, during operations (e.g., (A.7)). The exception is specifically the insertion/deletion of \( E \)'s such as in (29–34) and (36–45).

We briefly point out the notational benefits of the imaginary \( E \)'s. These are three; 1) to adjust the sorts (number) of bases in DNAs and RNAs (from ‘4’ to ‘5’), thus enabling group-theoretical composition over the (two-dimensional) plane; 2) to link the notational sequences of DNAs and mRNAs in a single format that can be used in a more compact database to record and analyze genetic information; and 3) to express sequences of DNAs/RNAs as a vector in three different ways: a) with explicit \( E \)'s in the vector, b) with implicit \( E \)'s in the vector, and c) with all \( E \)'s omitted in the vector except the trailing ‘E’s. The last offers flexibility in storing world-wide genetic data in a single set. We suggest that exhaustiveness is one of the potentialities of the model adding versatility in addressing the possibilities of certain behaviors of DNA/RNA sequences. While that might be far from practical applications at this moment, a more rigorous methodology in the near future may yield a means.

Regarding style of the grid-point/cruciform/wallpaper pattern (Figure 1) in defining the group postulates, one of its advantages is that each base is surrounded by the four others. This symmetrical simplicity is absent in the linear group and the rotational group (Figures 1, 5 and 6), where the relative position of the five bases is fixed and thereby restrictive. Also, it might be crucial that the number ‘5’ is key in enabling composition of the sort provided by the wallpaper group using the cruciform, and an identity element necessary to satisfy the group postulates. Being a prime, ‘5’ will be convenient in further developments of the model exploiting algebraic structures such as rings or fields.
A similar synthesis might be possible between a modulo 7 additive rotational group based on a sevenfold phasor diagram with a space group depending upon six ‘forward/backward,’ ‘up/down’ and ‘left/right’ directions. In practice, a space group is formed that consists of three orthogonal cruciforms comprising the six directions (±x, ±y, ±z) with seven elemental operations \{m_u(up), m_d(down), m_r(right), m_l(left), m_f(forward), m_b(backward) and m_n(no movement)\}. These determine the operations of the group, which permute seven bases (prime number) or seven letter-like constituents. Analogous to the fivefold phasor diagram, we draw equispaced elements on the unit circle over the Gaussian plane; suppose \(\phi = 2\pi/7\) (rad), then the set \{\phi_1(= \phi), \phi_2(= 2\phi), \phi_3(= 3\phi), \phi_4(= 4\phi), \phi_5(= 5\phi), \phi_6(= 6\phi), \phi_7(= \phi_7 = 0)\} parameterizes the rotational group [51, and both are, at least, in partial correspondence. We presume that, in extension, bringing together an n-dimensional space group (using the \(2n+1\) elements associated with the \(\pm n\)-directions and E) and a rotational group based on the n-fold phasor diagram on the unit circle (with \(2n+1\) elements as points of the vertex of a polygon) might be possible. For this article, we have just focused on \(n=2\) in §1–§8.

Apart from the above, the model based on the wallpaper pattern might have a close relationship with ‘cellular automata’ [52]. Appropriate definitions of the wallpaper pattern for the five bases might find an expression between groups and cellular automata [53].

One consideration concerns whether a more integrated/synthesized style to describe biomolecular processes is possible using only simple, primitive defining rules, in particular, when describing genetic processes such as DNA transcription and RNA synthesis of proteins. Whereas the group postulates might be too restrictive to define molecular behavior, category postulates might enable such schemes to proceed because its postulates are weaker than those defining a group. If the interpretation of DNA by messenger RNA is definable within category theory, and protein synthesis is expressible within the same theory, there might be advantages in having the molecular system classified and treated in a reduced size in the database. At least, we conjecture that these ideas might be valid when clarifying impossible phenomena associated with changes of DNA sequences, resulting in reducing unnecessary, recrementitious efforts or roundabout paths that might encroach on researchers’ limited time for investigation. That issue might be avoided if the impossibility of certain themes was known beforehand. From this standpoint, we believe that a mathematical systematization (in a general and unexceptional manner) is crucially important for future molecular/genetic biology.

The limitations of the present model should be noted. First, the wallpaper pattern drawn in Figure 1 is one example of various patterns. In general, the wallpaper groups have been classified into seventeen categories [2,44-47]. There could be other types of patterns like Figure 1 and groups upon which to compose this sort of model. For instance, if we exchange all ‘A’s for all ‘C’s, and all ‘G’s for all ‘T’s in the model presented in this article, an almost equivalent model ‘§1–§8’ is constructed. Other arrangements might provide still unknown advantages that enable models like ours to be treated in a more rational manner. It remains unclear how to construct an optimal method to determine models yielding the wallpaper pattern of Figure 1 and the bijection given in Figure 3, and to develop the categories presented in Figures 7 and 8. The best positions of the five bases should be examined under a rigorous methodology.

Second, a Cartesian vector is defined as a combination of components on which operations are conducted independently. Indeed, we can perform operations on the
i-th component of $D_j$ of set $D$ using the $(i + 1)$-th component of $B_j$ by adding an ‘E’ any place before the (i-1)-th component of $D_j$. This is because the components after the i-th of $B_j$ shift to the right within the vector $B_j$. Therefore, in appearance, the components at different positions are essentially spectators (e.g., a base ‘C’ cannot change into either ‘A’ or ‘G’ by any $B_j$ except via ‘E’-assisted manipulations). In this case, after the insertion of two ‘E’s between the ‘2’ and ‘3’ components, ‘C$_3+2$(= C$_5$)’ can become either ‘A’ or ‘G’ by acting appropriately on $B_j$ at C$_5$. However, that might raise some confusion. For ‘E’-assisted operations (such as 29–34, 38, 40), the results might change according to the place number of inserted/deleted ‘E’s that yields the mis-matches between the ‘i-th’ component of $D_j$ and that of ‘B$_j$’. We believe further studies are warranted to find a descriptive format for the model.

Third, as for the graphical displays of ‘D’s in Figure 4, although the sequences of ‘D$_m$’s (m = 1, 2, 3) are in reality the same, the respective expressions are not always unique because the presence of the imaginary ‘E’s changes the shape of each sequence; e.g., $D_1 = [A_1C_2C_3G_4T_5E_6E_7E_8...]$ and $D_2 = [A_1C_2C_3(E_4E_5)G_6T_7E_8E_9E_{10}...]$ are different over the wallpaper pattern despite being equivalent as real sequences. Although by use of electronic tools, these graphics might be of versatility for detection or identification of DNA sequences, these might produce other confusions in the present form. We hope that more appropriate devices would be performed in future study.

Fourth, DNA transcription to RNA and/or mRNA and translation of RNA and/or mRNA into proteins at the ribosomes are performed using a grammar rule based on a three-base set called a ‘codon’. Codons have information to synthesize twenty types of proteins; for example, ‘CAG’ codes for ‘glutamine’. As mentioned before, a number of approaches have been proposed exploit group-theoretic methods. These cover the rules for composition of triplet of bases ‘XXX’, the ways of reading codons, and models to compose geometric solids such as the tetrahedron and hexahedron, [12,18–21,33,34,50]. The rules for treating this aspect (transcription and translation of DNA bases information) are not established in the present article. In addition, there are specific types of codon, such as ‘TAA’, ‘TGA’, and ‘TAG’, which are presently classified as ‘stop’ or ‘halt’ commands. Aside from this, there are various rules related to biogenetic activities such as DNA repair, alternative splicing, transposition, and translocation. These specific characterizations are lacking in our model, so, further improvements on this issue are desirable.

Fifth, the traditional symmetry model of DNA bases often is based on the chemical types ‘purine/pyrimidine’, ‘amino/keto’, and ‘strong/weak hydrogen binding’ using biomolecular characteristics, which often have advantages for their treatments where three-dimensional graphics aid the imagination, and ‘matricized’ expressions are possible [29,35,36]. In our model, we merely use a rule for complementary pairing in §4 and §5. No restriction on couplings between ‘C, A, T, G and E’ is postulated in the present article. There might occur a number of combinations where non-realistic pairings of bases (e.g., ‘A-G’, ‘C-C’, and ‘T-E’) produce futileities and wastefulness in applications. We hope that future studies can solve this problem.

Sixth, there might be too many speculative conjectures with hypothetical situations those should be used to prove scientific facts using verified methods. Thus, a more
rigorous examination for a rational style with a more effective methodology is necessary.

Our model is far from a complete systematization. However, we believe that it is necessary that some principal breakthrough should be pursued if we intend to systemize a descriptive model, and that if appropriate definitions are devised, that might help to systemize biomolecular/genetic biology in a more optimized manner with greater sophistication to make a significant contribution to the field.

Conclusions
Within the large limitations of our methodology, it is considered that there is fertile ground where variants of the symmetry model for genetic coding based upon a specific wallpaper group are constructible. By integrating the linear group and rotational group over a specific wallpaper pattern, a more integrated formulation based on a group/category theory-like description is open to exploration in applications to a number of topics from molecular/genetic biology.

Appendix A
According to Figures 1, 3 and 6, the following relationships are confirmed straightforwardly between any bases and independently of the type of bases:

\[ \begin{align*}
  d \circ d &= u \circ u = r \circ n = n \circ r = r, \\
  r \circ r &= d \circ d = u \circ n = n \circ u = u, \\
  l \circ l &= r \circ u = u \circ r = d \circ n = n \circ d = d, \\
  u \circ u &= r \circ d = d \circ r = l \circ n = n \circ l = l, \\
  n \circ n &= r \circ l = l \circ r = u \circ d = d \circ u = n \\
\end{align*} \] (A.1)

Here the symbol ‘↔’ signifies ‘bijection’ and the meaning of ‘\([x[-1, 0]]\)’ is explained in §4. Hence, operators that are regarded to effect changes from one base to another can be re-expressed as illustrated in the following examples for various types of component operations:

\[ b_{[E \rightarrow C]} = b_{[C \rightarrow A]} = b_{[A \rightarrow G]} = b_{[G \rightarrow T]} = b_{[T \rightarrow G]} = b_{[G \rightarrow E]} = r \ (\leftrightarrow \omega_1) = x[1, 0], \] (A.2)

\[ b_{[E \rightarrow A]} = b_{[A \rightarrow G]} = b_{[G \rightarrow C]} = b_{[C \rightarrow T]} = b_{[T \rightarrow E]} = u \ (\leftrightarrow \omega_2) = x[0, 1], \] (A.3)

\[ b_{[E \rightarrow T]} = b_{[T \rightarrow C]} = b_{[C \rightarrow G]} = b_{[G \rightarrow A]} = b_{[A \rightarrow E]} = d \ (\leftrightarrow \omega_3) = x[0, -1], \] (A.4)

\[ b_{[E \rightarrow G]} = b_{[G \rightarrow T]} = b_{[T \rightarrow A]} = b_{[A \rightarrow C]} = b_{[C \rightarrow E]} = l \ (\leftrightarrow \omega_4) = x[-1, 0], \] (A.5)

\[ b_{[E \rightarrow E]} = b_{[C \rightarrow C]} = b_{[A \rightarrow A]} = b_{[T \rightarrow T]} = b_{[G \rightarrow G]} = n \ (\leftrightarrow \omega_0 (= \text{no rotation}) = \omega_5) = x[0, 0]. \] (A.6)

Appendix B
As for B,

1) Associativity: ‘\((B_j \bullet B_k) \bullet B_l = B_k \bullet (B_j \bullet B_l)\)’ holds for all positive integers j, k and l.

2) Identity: ‘\(B_0 = [n_1 \mid n_2 \mid n_3 | \ldots | n_i | \ldots | n_{(n-1)} | n_n | n_{n+1} | n_{n+2} | n_{n+3} | \ldots]’ is an identity element that satisfies ‘\(B_0 \bullet B_m = B_m \bullet B_0 = B_m\)’. (i = 1, 2, 3, ...; ‘\(n_i (= n)\)’ is an element of \(Z_5\) (no movement of the point P))
3) Inverses: there exists a unique ‘B\text{-m}^{-1}\text{'} that satisfies ‘B\text{-m}^{-1}\text{'} \bullet B_\text{m} = B_\text{m} \bullet B\text{-m}^{-1} = B_0\text{'}.

Actually, the components of the inverse are the inverses of each individual component.

4) Commutativity: ‘B_\text{m} \bullet B_k = B_k \bullet B_\text{m}’.

5) Closure law: any ‘B_\text{m} \bullet B_k’ belongs to the set B.

Appendix C

\[
D_2 B_{2 \rightarrow 3} \bullet B_{3 \rightarrow 4} = [A_1 \circ C_2 \circ C_3 \circ E_2 \circ E_3 \circ G_6 \circ T_7 \circ A_N \circ E_8 \circ \ldots]
\]
\[
\circ \{n_1 \circ n_2 \circ u_1 \circ d_1 \circ n_6 \circ n_7 \circ n_8 \circ n_9 \circ \ldots \} \circ \{l_1 \circ u_2 \circ d_3 \circ (r_4 \circ d_5 \circ n_6 \circ l) \circ n_8 \circ n_9 \circ \ldots \}.
\]
\[
= [A_1 \circ n_1 \circ l_1 \circ C_2 \circ n_2 \circ u_2 \circ C_3 \circ n_3 \circ d_3 \circ (E_4 \circ u_4 \circ r_4 \circ E_5 \circ d_5 \circ d_5) \circ G_6 \circ n_6 \circ n_7 \circ l_2 \circ E_8 \circ n_8 \circ n_9 \circ \ldots \] (A.7)

Again, with reference to Figure 1 or Appendix B,
\[
= [A_1 \circ l_1 \circ C_2 \circ u_2 \circ C_3 \circ d_3 \circ (E_4 \circ d_4 \circ E_5 \circ r_2 \circ E_6 \circ n_6 \circ E_8 \circ n_9 \circ \ldots \] (A.8)
\]
\[
= [C_1 \circ T_2 \circ G_3 (T_4 \circ C_5) \circ G_6 \circ A_7 \circ E_8 \circ \ldots ] = D_4.
\]

Appendix D

Naturally, the series D_k is generated through the following sequence of operations:

\[
B_{(i \rightarrow k)} = [r^4 \circ l_1 \circ r^3 \circ l_2 \circ n_0 \circ i_3 \circ \ldots \circ n_0 \circ n_1 \circ n_3 \circ t^3 \circ \ldots \circ n_3 \circ n_1 \circ n_3 \circ t^3 \circ \ldots \]
\[
\circ \{l_1 \circ n_1 \circ r_0 \circ n_2 \circ l_2 \circ n_3 \circ \ldots \} \circ \{l_1 \circ n_1 \circ r_0 \circ n_2 \circ l_2 \circ n_3 \circ \ldots \}.
\]
\[
= [E^t \circ r^4 \circ r^3 \circ r^3 \circ r^3 \circ n_0 \circ n_1 \circ n_3 \circ t^3 \circ \ldots \circ n_3 \circ n_1 \circ n_3 \circ t^3 \circ \ldots \]
\[
\circ \{l_1 \circ n_1 \circ r_0 \circ n_2 \circ l_2 \circ n_3 \circ \ldots \} \circ \{l_1 \circ n_1 \circ r_0 \circ n_2 \circ l_2 \circ n_3 \circ \ldots \}.
\]
\[
= [E^t \circ r^4 \circ r^3 \circ n_0 \circ n_1 \circ n_3 \circ t^3 \circ \ldots \circ n_3 \circ n_1 \circ n_3 \circ t^3 \circ \ldots \] (A.9)
\]

Then,

\[
D_2 \circ B_{(i \rightarrow k)} = [C_1 \circ A_2 \circ \ldots \circ C_i \circ \ldots \circ T_{N-1} \circ A_N \circ E_{N+1} \circ E_{N+2} \circ E_{N+3} \circ \ldots \circ l_1 \circ n_1 \circ r_0 \circ n_2 \circ l_2 \circ n_3 \circ \ldots \]
\[
\circ \{l_1 \circ n_1 \circ r_0 \circ n_2 \circ l_2 \circ n_3 \circ \ldots \} \circ \{l_1 \circ n_1 \circ r_0 \circ n_2 \circ l_2 \circ n_3 \circ \ldots \}.
\]
\[
= [E^t \circ r^4 \circ E^t \circ r^3 \circ \ldots \circ E^t \circ r^3 \circ \ldots \circ E^t \circ r^3 \circ \ldots \circ E^t \circ r^3 \circ \ldots \]
\[
\circ \{l_1 \circ n_1 \circ r_0 \circ n_2 \circ l_2 \circ n_3 \circ \ldots \} \circ \{l_1 \circ n_1 \circ r_0 \circ n_2 \circ l_2 \circ n_3 \circ \ldots \}.
\]
\[
= [E^t \circ r^4 \circ E^t \circ r^3 \circ \ldots \circ E^t \circ r^3 \circ \ldots \circ E^t \circ r^3 \circ \ldots \] (A.9)
Appendix E

\[ D_1 \cdot B_{i(p)} = [\text{.hom(X, Y)} \times \text{hom(Y, Z)} \rightarrow \text{mor}(X, Z)] \]

Appendix F

The axioms are:

I) A binary operation and closure law: the combination of two morphisms satisfies \( \text{hom}(X, X) \times \text{hom}(X, Y) \rightarrow \text{hom}(X, Y) \). Moreover, \( \text{hom}(X, X) \times \text{hom}(Y, Z) \rightarrow \text{mor}(X, Z) \) hom(Y, Z) \rightarrow \text{mor} \) and \( \text{hom}(X, X) \times \text{hom}(Y, Z) \rightarrow \text{mor}(X, Z) \) both hold.

II) Associativity: If \( f: X \rightarrow X \), \( \rho: X \rightarrow Y \), \( \tau: Y \rightarrow Z \), \( g: X \rightarrow Z \), and \( h: Z \rightarrow Z \), then \( f \cdot (\rho \circ \tau) = (f \cdot \rho) \circ \tau \). \( \rho \cdot (\tau \circ h) = (\rho \cdot \tau) \circ h \), \( \rho \cdot (\tau \circ h) = (\rho \cdot \tau) \circ h \) and \( \rho \cdot (\tau \circ h) = (\rho \cdot \tau) \circ h \) hold.

III) Identity: there exist morphisms \( 1_X, 1_Y, 1_Z \) such that \( 1_X = 1_Y = 1_Z = \text{[1]} \) satisfies these conditions.

\[ (A.10) \]

Appendix G

For Category \( C_1 \),

morphism \( f_1 = B_1(\text{[group B]}): X_1 \rightarrow X_1 \),
morphism \( \rho_1 = X_1 \rightarrow Y_1 \),
morphism \( \tau_1 = Y_1 \rightarrow Z_1 \),
morphism \( g_1 = B_1(\text{[group B]}): X_1 \rightarrow Z_1 \),
morphism \( h_1 = B_1(\text{[group B]}): Z_1 \rightarrow Z_1 \),
morphism \( j_1 = Z_1 \rightarrow Z_1 \).

\[ (A.12) \]
Similarly for category $C_2$, for each object $F(X_1) = X_2$, $F(Y_1) = Y_2$, $F(Z_1) = Z_2$, $F(Zs_1) = Zs_2$ ($\in C_2$), the following relationships also hold:

- morphism $F(\mathcal{f}_1) (= \mathcal{f}_2 = B_2(\in group B_2))$: $F(X_1) \rightarrow F(X_1)$,
- morphism $F(\mathcal{p}_1) (= \mathcal{p}_2)$: $F(X_1) \rightarrow F(Y_1)$,
- morphism $F(\mathcal{t}_1) (= \mathcal{t}_2)$: $F(Y_1) \rightarrow F(Z_1)$,
- morphism $F(\mathcal{g}_1) (= \mathcal{g}_2 = \mathcal{p}_2 \circ \mathcal{t}_2)$: $F(X_1) \rightarrow F(Z_1)$,
- morphism $F(\mathcal{h}_1) (= \mathcal{h}_2 = B_2(\in group B_2))$: $F(Z_1) \rightarrow F(Z_1)$,
- morphism $F(\mathcal{f}_1) (= \mathcal{j}_2)$: $F(Z_1) \rightarrow F(Zs_1)$.

Other than these, if relationships $F(\mathcal{f}_1 \circ \mathcal{p}_1) = F(\mathcal{f}_1) \circ F(\mathcal{p}_1)$, $F(\mathcal{p}_1 \circ \mathcal{t}_1) = F(\mathcal{p}_1) \circ F(\mathcal{t}_1)$, $F(\mathcal{t}_1 \circ \mathcal{h}_1) = F(\mathcal{t}_1) \circ F(\mathcal{h}_1)$, $F(\mathcal{g}_1 \circ \mathcal{g}_1) = F(\mathcal{g}_1) \circ F(\mathcal{g}_1)$, $F(\mathcal{g}_1 \circ \mathcal{h}_1) = F(\mathcal{g}_1) \circ F(\mathcal{h}_1)$, and $F(\mathcal{h}_1 \circ \mathcal{j}_1) = F(\mathcal{h}_1) \circ F(\mathcal{j}_1)$ are satisfied, the composition of $C_1$ and $C_2$ linked with ‘functor $F$’ is possible although the proof is omitted here.

Furthermore, the following postulates hold: for object $X$ ($\in C_1$), $F(1_X) = 1_{F(X)}$ ($\in C_2$)’ is true, for object $Y$ ($\in C_2$), $F(1_Y) = 1_{F(Y)}$ ($\in C_2$)’ and for object $Z$ ($\in C_2$), $F(1_Z) = 1_{F(Z)}$ ($\in C_2$), is true under the condition:

$$F(1_X) = 1_{F(X)} = F(1_Y) = 1_{F(Y)} = F(1_Z) = 1_{F(Z)} = [00_1 | 00_2 | 00_3 | \ldots | 00_4 | \ldots | 00_N | \ldots | 00_N | \ldots].$$

(A.14)

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
JS conceived of the main idea of this article and wrote the manuscript. SM revised the manuscript. JI gave advice on potential versatilities of the model to the biological science. In addition, all authors read and approved the final version of the manuscript.

Acknowledgments
The authors wish to acknowledge Katsuji Nishimura, Ken Inada, and Kaoru Sakamoto for providing us with very useful advice during this study.

Author details
1Department of Psychiatry, Tokyo Women’s Medical University, Tokyo, Japan. 2Depression Prevention Medical Center, Inariyama Takeda Hospital, Kyoto, Japan.

Received: 25 September 2013 Accepted: 9 February 2014
Published: 7 May 2014

References
1. Rosen J. Symmetry rules: How science and nature are founded on symmetry. 1st edition. New York: Springer-Verlag; 2008.
2. Armstrong MA: Groups and symmetry, undergraduate texts in mathematics. New York: Springer-Verlag; 1988.
3. Judson TW: Abstract Algebra: Theory and Applications. Virginia: PWS Publishing Company; 1997.
4. Hungerford TW: Abstract Algebra, An Introduction. 2nd edition. Philadelphia: Saunders College Publishing; 1997.
5. Tung WK: Group theory and physics. Singapore: World Scientific Pub. Co. Inc.; 1985.
6. Unger AA: The abstract Lorentz transformation group. Am J Phy 1992, 60:815–828.
7. Sexl RJ, Ulbertke HU: Relativity, Groups, Particles: Special Relativity and Relativistic Symmetry in Field and Particle Physics. Wien: Springer; 2001.
8. Hamermesh M: Group theory and its application to physical problems. New York: Dover Publications, Inc.; 2012.
9. Ladd M: Symmetry and group theory in chemistry. Cambridge: Woodhead Publishing Limited; 1998.
10. Deome J-R: Biological Similarity and Group Theory. J Theor Biol 1977, 65:369–378.
11. Wang J: A complete symmetrical group DNA sequences and symmetry of poly-codon sequences (I). J Biomathematics 2001, 16:129–136.
12. Wang J: A complete symmetrical group DNA sequences and symmetry of poly-codon sequences (II). J Biomathematics 2001, 16:257–265.
13. Chirikjian GS: Group theory and biomolecular conformation: I Mathematical and computational models. J Phys Condens Matter 2010, 22:S23103.
14. Chirikjian GS: Mathematical aspects of molecular replacement. I. Algebraic properties of motion spaces. Acta Crystallogr A 2011, 67:435–436.
15. Fischer M, Klaere S, Nguyen MAT, von Hae seler A: On the group theoretical background of assigning stepwise mutations onto phylogenies. Al gorithms Mol Biol 2012, 7:36.
16. Bashford JD, Tsou hanijs I, Jarvis PD: Codon and nucleotide assignments in a supersymmetric model of the genetic code. Phys Lett A 1997, 233:388–481.
17. Bashford JD, Tsohantjis I, Jarvis PD: A supersymmetric model for the evolution of genetic code. *Proc Natl Acad Sci U S A* 1998, 95:887–992.

18. Sánchez R, Morgado E, Grau R: Gene algebra from a genetic code algebraic structure. *J Math Biol* 2005, 51:431–457.

19. Sánchez R, Grau R: Vector space of the extended base-triplets over the Galois field of five DNA bases alphabet. *World Acad Sci Eng Technol*, Int J Comp, Inf Sci Eng 2007, 1:5.

20. Sánchez R, Grau R: An algebraic hypothesis about the primeval genetic code architecture. *Math Biosci* 2009, 221:60–76.

21. Sánchez R, Grau R: A novel Lie algebra of the genetic code over the Galois field of four DNA bases. *Math Biosci* 2006, 202:156–174.

22. Rietman EA, Karp RL, Tuszynski JA: Role of the amino-acid evolution. *Evolutionary changes in the genetic code*. *J Theor Biol* 1968, 23:359–365.

23. Korn F: Elementary Structures reconsidered: Lévi-Strauss on kinship. *London: Routledge*. 2004.

24. Crick FHC: Codon-anticodon pairing: The wobble hypothesis. *J Mol Biol* 1966, 19:548–555.

25. Crick FHC: The origin of the genetic code. *J Mol Biol* 1968, 38:367–379.

26. Crick FHC: On protein synthesis. *In Symp Soc Exp Biol* 1958, 12:138–163.

27. Bashford JD, Jarvis PD: The genetic code as a periodic table. *Biosystems* 2000, 57:147–161.

28. Beland P, Allen TF: A novel graphical and numerical representation for analyzing DNA sequences. *ARKIVOC* 2007, ix:211–238.

29. Beland P, Allen TF: A supersymmetric model for the evolution of genetic code. *J Theor Biol* 1997, 187:297–306.

30. Duplij D, Duplij S: DNA sequence representation by trianders and determinative degree of nucleotides. *J Zhongshan Univ (Sci)* 2005, 58:743–755.

31. Rushdi A, Tuqan J, Strohmer T: Map-invariant spectral analysis for the identification of DNA periodicities. *EURASIP J Bioinform Syst Biol* 2012, 2012:16.

32. Zupan J, Randić M: Algorithm for coding DNA sequences into “spectrum-like” and “zigzag” representations. *J Chem Inf Model* 2005, 45:309–313.

33. Lockwood EH, Macmillan RH: Transformation Geometry. *London: Routledge*. 2004.

34. Trainor LEH, Rowe GW, Szabo VL: Transformation Geometry. *Oxford Logic Guides*. 2nd edition. New York: Springer-Verlag; 1998.

35. Butler JT: A note on cellular automata simulations. *Inf Control* 1974, 26:286–295.

36. Ceccherini-Silberstein TG, Mochi A, Scarabotti F: Amenable groups and cellular automata. *Ann Inst Fourier (Grenoble)* 1999, 49(issue 2):673–685.

Cite this article as: Sawamura et al.: A symmetry model for genetic coding via a wallpaper group composed of the traditional four bases and an imaginary base E: Towards category theory-like systematization of molecular/genetic biology. *Theoretical Biology and Medical Modelling* 2014 11:18.