Scientific
Research
Publishing

On the Defining Equations of Protein's Shape from a Category Theoretical Point of View

Naoto Morikawa

Genocript, Zama, Japan
Email: nmorika@genocript.com

Abstract

This paper proposes a novel category theoretic approach to describe protein's shape, *i.e.*, a description of their shape by a set of algebraic equations. The focus of the approach is on the relations between proteins, rather than on the proteins themselves. Knowledge of category theory is not required as mathematical notions are defined concretely. In this paper, proteins are represented as closed trajectories (*i.e.*, loops) of flows of triangles. The relations between proteins are defined using the fusion and fission of loops of triangles, where allostery occurs naturally. The shape of a protein is then described with quantities that are measurable with unity elements called "unit loops". That is, protein's shape is described with the loops that are obtained by the fusion of unit loops. Measurable loops are called "integral". In the approach, the unit loops play a role similar to the role "1" plays in the set $\mathbb{Z}$ of integers. In particular, the author considers two categories of loops, the "integral" loops and the "rational" loops. Rational loops are then defined using algebraic equations with "integral loop" coefficients. Because of the approach, our theory has some similarities to quantum mechanics, where only observable quantities are admitted in physical theory. The author believes that this paper not only provides a new perspective on protein engineering, but also promotes further collaboration between biology and other disciplines.

Keywords

Differential Geometry, Discrete Mathematics, Protein Design, Triangular Flow, Algebra of Loops

1. Introduction

The interaction of proteins is mainly determined by their shape because they form protein-protein comlexes to perform their tasks. This paper considers the

defining equations of the shape of proteins, using the mathematical toy model proposed in [1]. Our focus is on "relations between proteins" rather than "proteins".

In the mathematical toy model, proteins are represented as closed trajectories (*i.e.*, loops) of flows of triangles. The formation of protein-protein complexes is then described as a "fusion" of loops (denoted by "+"). One of the features of the model is a simple mechanism of the enzyme/substrate/activator-type regulation, *i.e.*, the type of regulation called "allosteric regulation" in which a substrate cannot bind to an enzyme without an activator [2] [3].

The purpose of this paper is to describe the shape of a loop with "measurable" quantities, which are the shapes obtained by the "fusion" of unity elements called "unit loops". Unit loops are of length 6 and have a hexagonal shape. The unit loops play a role similar to the role "1" plays in the set $\mathbb{Z}$ of integers. We call the loops obtained by the fusion of unit loops "integral". Loops obtained by factoring an integral loop are called "rational".

Because of the approach, our theory has some similarities to quantum mechanics, where only observable quantities are admitted in physical theory. For example, we consider two sets of flows in the following. One is a set L of the "rational" flows, and the other is a set L^A of the "integral" flows. (Loops of flows of L and L^A are called "rational" and "integral", respectively.) In analogy to quantum mechanics, L corresponds to the quantum world of objects, and L^A corresponds to the world of observers on which the measurement is performed.

The motivation of the study is the search for a new (discrete) geometry with high affinity to biology. Despite of the significant progress in mathematics in recent decades, it is still too early to say that we can describe the shape of bio-molecules, such as proteins, mathematically.

When designing a new protein, there are two types of approaches. One is the design of proteins with a desired backbone structure. The other is the design of proteins with desired functions (*i.e.*, desired active sites or desired interacting surfaces). In both cases, target descriptions are usually given as a two-dimensional schematic diagram [4] [5]. In the diagram, three-dimensional backbone structures are represented as a sequence of local structural patterns (such as alpha-helices and beta-strands) with sets of pairwise spatial relationships between them.

In the previous studies, the author was thinking of specifying a given shape by the way it splits, *i.e.*, a fission-based approach [6] [7]. However, it is not yet clear how to describe the relation between various fissions of the given shape. (In the conclusion section, this issue is mentioned as one of the directions for future research. That is, for a given *L-spectrum*, find a flow U that has the same *L-spectrum*.) To obtain a novel mathematical representation of protein shape, the author now adopts a fusion-based approach rather than the fission-based approach using the notion of "integral loops".

In this paper the author shows that it is possible to describe protein-like

shapes using algebraic equations if one takes a category theoretical approach. Along the way, also provided is the basis for describing the relation between various fissions of a shape, *i.e.*, *L-spectrum* of a flow $U \in L$, albeit briefly. However, they are results for a simple model and the author considers this to be a starting point rather than an end point of the study. The ultimate goal is not only to understand the functional mechanisms of proteins, but also to design protein-like molecules with new functions. I hope that the mathematical description proposed will provide some insight into the realization of this goal.

Basic ideas of the approach adopted in this paper are:

- Loops are identified with the flow consisting of the loop and any open trajectories.
- A binary relation "<" between flows is defined using the fusion "+" of loops.
- The fusion of flows of L^A is regarded as an "addition".
- The fission of a flow of L^A into flows of L is regarded as a "factorization".
- Flows of L are specified as a "direct factor" of a flow of L^A.
- The "shadow" of a flow of L is a projection on L^A, and is measurable with unit loops (*i.e.*, hexagons).

Figure 1(a) shows the process for obtaining the defining equations of the shape of a loop b_1.
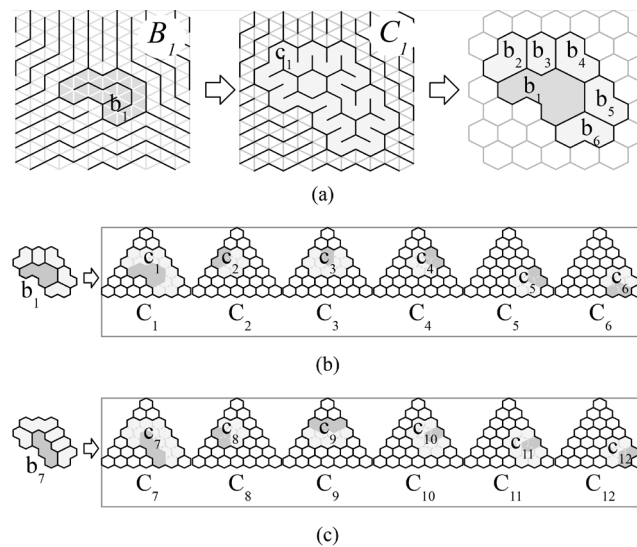

(a)

(b)

(c)

Figure 1. Approach for obtaining the defining equations of a loop b_1. (a) The shadow C_1 of a flow B_1. (left) A flow $B_1 \in L$ consisting of a rational loop b_1 and open trajectories. (middle) The flow $C_1 \in L^A$ consisting of an integral loop c_1 and open trajectories. (right) A fission of c_1 into rational loops $b_1, b_2, \cdots$ and $b_6 \in L$. The background unit loops (*i.e.*, hexagons) are shown for clarity. (b) Fission of c_1 into $b_1, b_2, \cdots, b_6$, where b_1 is colored dark grey. On the right are the shadows C_i of B_i, where B_i consists of a loop b_i (colored dark grey) and open trajectories ($i = 1, 2, \cdots, 6$). C_i consists of an integral loop c_i and open trajectories. The background unit loops are shown for clarity. (c) Fission of c_1 into $b_7, b_8, \cdots, b_{12}$, where b_7 is colored dark grey. On the right are the shadows C_i of B_i, where B_i consists of a loop b_i (colored dark grey) and open trajectories ($i = 7, 8, \cdots, 12$). C_i consists of an integral loop c_i and open trajectories.

1) First, b_1 is identified with $B_1 \in L$ consisting of b_1 and open trajectories.

2) Next, "a minimal flow $C_1 \in L^A$ that contains B_1" is computed. C_1 consists of an integral loop c_1 and open trajectories. C_1 is called the "shadow" of B_1.

3) C_1 is then factored into b_1 and other loops $b_2, b_3, \cdots, b_6$.

4) The shadows $C_i \in L^A$ of $B_i \in L$ ($i = 2, 3, \cdots, 6$) are also computed, where B_i consists of a loop b_i and open trajectories. C_i consists of an integral loop c_i and open trajectories.

5) B_1 is specified by a set of equations using the six shadows $C_1, C_2, \cdots, C_6$. (See Problem 1 below for the equation.)

Note that we can represent integral loops $c_1, c_1, \cdots, c_6$ as a fusion of unit loops (*i.e.*, hexagons).

To check uniqueness, let's consider another loop b_7 of **Figure 1(c)**. Let $B_7 \in L$ be the flow consisting of b_7 and open trajectories. Then, B_7 has the same shadow C_1 on L^A, where b_7 fuses with loops $b_8, b_9, \cdots, b_{12}$ to form c_1. In this case, $\{C_7, C_8, \cdots, C_{12}\}$ is different from $\{C_1, C_2, \cdots, C_6\}$, *i.e.*, $\{C_4, C_5\} \neq \{C_{10}, C_{11}\}$.

In summary, B_1 is a solution to the following problem.

Problem 1. For a given set of flows $\{C_1, C_2, \cdots, C_6\} \subset L^A$, find a flow $W_1 \in L$ such that there exists a set of flows $\{W_2, W_3, \cdots, W_6\} \subset L$ that satisfies

$$\begin{cases} cl(W_1) = W_1 + W_2 + W_3 + W_4 + W_5 + W_6 & (\text{fusion}), \\ cl(W_i) = C_i & (i = 1, 2, \cdots, 6), \end{cases}$$

where $cl(W_i)$ is the shadow of W_i ($i = 1, 2, \cdots, 6$).

We denote the set of all the solutions to the above type of problem by "$C_1/C_2C_3 \cdots C_m$". By analogy with quantum mechanics, C_1 is called the "output spectral flow", and $C_2, C_3, \cdots, C_m$ are called the "input spectral flows". In the case of **Figure 1**, we obtain

$$\begin{cases} C_1/C_2C_3C_4C_5C_6 = \{B_1\}, \\ C_1/C_2C_3C_{10}C_{11}C_6 = \{B_7\}. \end{cases}$$

In general, $C_1/C_2C_3 \cdots C_m$ consists of more that one flows. For example

$$\{B_5, B_6, B_{12}\} \subset C_5/C_5C_5.$$

Also shown in this paper is that the set $\{c_1, c_2, \cdots, c_6\}$ of integral loops are the unique (up to the rotational and mirror symmetries) solution to the following problem. For intuitive understanding, we use loops instead of flows.

Problem 2. Find integral loops $z_1, z_2, \cdots, z_6$ such that

$$\begin{cases} z_1 = x_4 + \cdots + x_{10} + x_{12} + \cdots + x_{15} + x_{19} + \cdots + x_{21} + x_{26} + x_{27}, \\ z_2 = x_4 + x_5 + x_7 + x_8 + x_9 + x_{12} + x_{13}, \\ z_3 = x_4 + \cdots + x_{10} + x_{12} + x_{13} + x_{14}, \\ z_4 = x_5 + x_6 + x_8 + x_9 + x_{10} + x_{13} + x_{14}, \\ z_5 = x_{14} + x_{15} + x_{19} + x_{20} + x_{21} + x_{26} + x_{27}, \\ z_6 = z_5, \end{cases}$$

where $x_1, x_2, \cdots, x_{28}$ are unit loops and satisfy

$$\begin{cases} y_L = x_1 + x_2 + \cdots + x_{28}, \\ y_1 = x_1 + x_2 + x_3, \\ y_2 = x_2 + x_5 + x_6, \\ \quad \cdots, \\ y_{21} = x_{21} + x_{27} + x_{28} \end{cases}$$

for some integral loops $y_L, y_1, y_2, \cdots, y_{21}$.

That is, combining Problem 1 and Problem 2, we obtain the defining equations of the shape of the given loop b_1.

In what follows, the author tries to present the approach outlined above in a systematic manner from scratch using simple examples. Knowledge of category theory is not required, as we define mathematical notions concretely. The rest of the paper is organized as follows. In Section 2, we give a brief review of previous results. After defining basic notions of loops of triangles in Section 3, we give the definitions of the set L of the rational flows in Section 4. The set L^A of the integral flows is defined in Section 5. Then, we consider the defining equations of flows of L in a step-by-step manner in Section 6. After a discussion in Section 7, the author concludes with some suggestions for future research directions in Section 8.

The author believes that this paper will open up a new perspective for the protein engineering and bring about further advances in collaboration between biology and other disciplines.

Finally, Genocript (http://www.genocript.com) is the one-man bio-venture started by the author in 2000 which is developing software tools for protein structure analysis. In particular, the author is not affiliated with any research institution.

2. Previous Works

Category theory has been offered as a common language for over 50 years across disciplines other than pure mathematics, such as theoretical computer science, mathematical physics, network theory, and others. In biology, the application of category theory began in 1958 when Robert Rosen proposed a category theoretic model of metabolic networks called the "M-R system" [8] [9] [10]. However, the major concern is mathematical modeling of biological systems and has not been applied to analysis of the shape of bio-molecules. (See also [11] for application of group theory to molecular systems biology.)

As for analysis of the shape of bio-molecules, the alpha shape theory provides an accurate and robust method for computing topological, combinatorial, and metric properties of the union of finitely many spherical d-balls in the d-dimensional Euclidean space $\mathbb{R}^d$ [12] [13]. The principle of graph theory is also adopted in the description of protein structure [14] [15]. However, the fusion and fission of bio-molecules are not considered there.

Finally, as for the mechanism of allosteric regulation, it has been a focus of much effort over 50 years and many models have been successfully formulated

[2]. Broadly, studies of allostery fall into three mainstream categories. The first is based on the principle of thermodynamic equilibrium. The second is based on the conceptual thermodynamic view such as conformational selection and induced fit. The third is based on the inferred structural coupling between the two sites (*i.e.*, the active and binding sites). On the other hand, our interpretation provides a purely geometrical explanation of allostery.

3. Basic Notions of Loops of Triangles

We begin by defining the two basic notions, "loops of triangles" and "fusion and fission" of loops. We then show how to quickly compute a loop with a given shape. The same method can be used for the computation of fusion and fission of given loops.

3.1. Loops of Triangles

To define a loop of triangles, we divide the two-dimensional Euclidean plane $\mathbb{R}^2$ into equilateral triangles as illustrated in **Figure 2(a)**.

Definition 1 (Flow of triangles). Let $TT = \{t_1, t_2, t_3, \cdots\}$ be the triangle tiling obtained by dividing $\mathbb{R}^2$ into equilateral triangles (**Figure 2(a)**). A flow U of triangles is a "reflective" assignment of an adjacent triangle to triangles of TT, *i.e.*,

$$U : TT \to TT \,, \quad U(t) \text{ is adjacent to } t \text{ and } U(U(t)) = t \,.$$

In the figures, we indicate the value of $U(t)$ ($t \in TT$) by thickening the edge shared by t and $U(t)$. *Trajectories* of U are then obtained by connecting triangles to the two adjacent triangles that are not assigned by U (**Figure 2(b)**). That is, $U(t)$ is the "normal vector" to the trajectory at t. The set of all flows on TT is denoted by $FL(TT)$.

For example, a loop of length 6 is obtained by connecting triangles one by one as shown in **Figure 2(b)**. The loop b_1 of **Figure 1(a)** is also obtained in the same way (**Figure 2(c)**).

Definition 2 (Loop of triangles). A (triangular) loop is a closed trajectory of triangles of finite length. Let $U \in FL(TT)$. The set of all the loops occurred in U is denoted by $lp(U)$. If $lp(U) = \{b\}$, then we identify a flow U with the loop b. We denote flows by upper case letters, and loops by lower case letters.
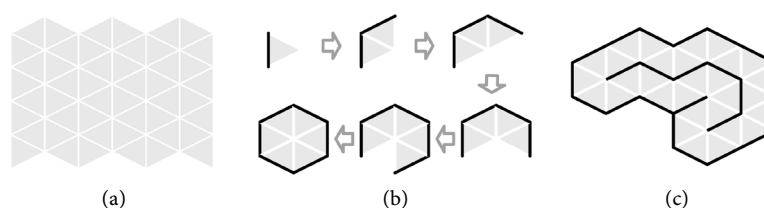

| (a) | (b) | (c) |

Figure 2. Loops of triangles. (a) Triangle tiling of $\mathbb{R}^2$. (b) Formation of a loop of length 6. Triangles are connected to two of the three adjacent triangles via the shared edge. Thick line segments indicate the "normal vector" edge. (c) The loop b_1 of length 30 given in **Figure 1**.

3.2. Fusion and Fission of Loops

"Fusion and fission" of loops are defined by identifying sets of loops with the same "shape".

Definition 3 (The shape $|b|$ of a loop b). Let $U \in FL(TT)$ and $b \in lp(U)$. The shape $|b| \subset TT$ of b is the union of the triangles swept by b.

Remark 1. Some areas are not obtainable as the shape of a loop.

Definition 4 (Equivalence relation "$\equiv$" on loops). Let $U_i \in FL(TT)$ and $b_i \in lp(U_i)$ ($i = 1, 2$). Then,

$$b_1 \equiv b_2 \quad \text{if and only if} \quad |b_1| = |b_2|.$$

Definition 5 (Fusion "+" of loops). Let $V \in FL(TT)$ and $\{c_1, c_2, \cdots, c_m\} \subset lp(V)$. The *fusion* $c_1 + c_2 + \cdots + c_k$ of c_i ($i = 1, 2, \cdots, m$) is defined if and only if there exists

$$U \in FL(TT) \quad \text{and} \quad b \in lp(U) \quad \text{such that} \quad |b| = |c_1| \bigcup |c_2| \bigcup \cdots \bigcup |c_m|.$$

Then, we write

$$b \equiv c_1 + c_2 + \cdots + c_m.$$

We often denote $c_1 + c_2 + \cdots + c_m$ by $\Sigma_i c_i$ (if there is no risk of confusion). Note that (1) $\Sigma_i c_i$ is not always defined, and (2) $\Sigma_i c_i$ is unique up to equivalence "$\equiv$" even if $\Sigma_i c_i$ is defined. c_i ($i = 1, 2, \cdots, m$) are said to *interact with each other* if $\Sigma_i c_i$ is defined. The inverse operation of fusion is called *fission*. Loops may have more than one fissions.

Example 1. Two fissions of c_1 are shown in **Figure 1(b)** and **Figure 1(c)**, *i.e.*,

$$\begin{cases} c_1 \equiv b_1 + b_2 + b_3 + b_4 + b_5 + b_6, \\ c_1 \equiv b_7 + b_8 + b_9 + b_{10} + b_{11} + b_{12}. \end{cases}$$

3.3. Computation of the Loops with a Given Shape

By stacking unit cubes diagonally in the three-dimensional Euclidean space $\mathbb{R}^3$, we can quickly compute a loop that fills a given two-dimensional "shape" on TT [1].

Example 2. The loop b_1 of **Figure 1(a)** is obtained by stacking unit cubes as shown in **Figure 3**. By stacking unit cubes so that the thick line segments on them form the contour of the shape, a loop with the contour is automatically obtained (**Figure 3(b)**). Each triangle is connected to two adjacent triangles via shared edge (thin line segment) (**Figure 3(c)**).

We can also compute a fusion and fission of loops instantly by placing/removing unit cubes on/from the stacked unit cubes.

Example 3. **Figure 4(a)** shows the fusion of two loops. Two loops of flow U_1 are fused into the loop of flow U_3 by placing a unit cube on U_1. Two loops of flow U_2 are also fused into the same loop of U_3 by removing a unit cube from U_2. U_3 has no loop decomposition other than U_1 and U_2.

As mentioned in the introduction, one of the features of our model is a simple mechanism of the allosteric regulation.
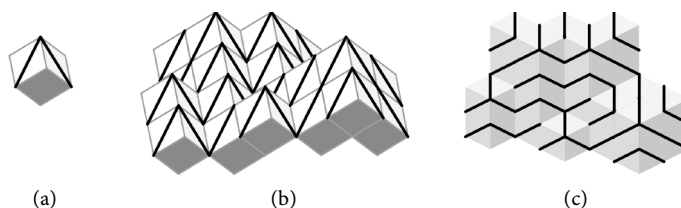
Figure 3. Stacking of unit cubes in $\mathbb{R}^3$. (a) A unit cube to be stacked. Each of the three top faces is divided into two triangles by a thick line segment. The thick edge of the triangles corresponds to the "normal vector" edge of the flow. (b) Stacked unit cubes. (c) The top view of the stacked cubes of (b). Triangles on the surfaces of the stacked cubes are connected via thin edges, forming the loop b_1 of **Figure 1**.
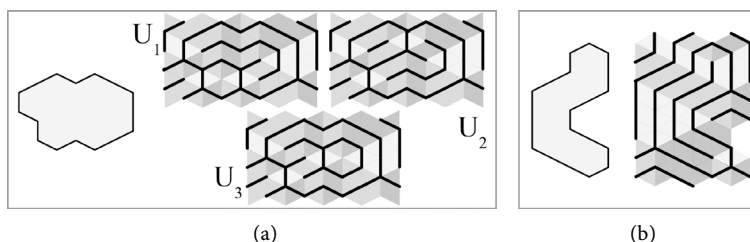

Figure 4. Computation of loops with a given shape. (a) A shape (left) and the corresponding flows (right). On the right are the top views of stacked unit cubes. Flows U_1 and U_2 consist of two loops and open trajectories. Flow U_3 consists of a loop and open trajectories. (The contained loop is considered to be part of the surrounding loop.) (b) Non-affine loop. Note that there is an "overhang" of stacked unit cubes in the lower right.

Definition 6 (Allosteric loop). Let $e \in lp(U)$ for some $U \in FL(TT)$. e is called *allosteric* if there exist $V \in FL(TT)$ and $e, s, c \in lp(V)$, such that

$$e + s + c \text{ is defined, but } e + s \text{ is not defined.}$$

Example 4. In **Figure 5**, e is allosteric. (That is, b_1 of **Figure 1** is allosteric.) Loops s and e don't fuse into a loop due to an "overhang" of stacked unit cubes.

4. Affine Flows *L* (Aka Rational Flows)

Now let's define the set *L* of flows mentioned in the introduction. The other set L^A is defined in the next section. We define not only the objects in a set, but also a binary relation between objects within the set. This is why we call our approach a "category theoretical" point view. Using the binary relation, we define the set $L^\wedge$ of $\{0,1\}$-valued functions on *L*. As for introduction to category theory, a number of standard textbooks are available, including [16] and [17].

4.1. The Set *L* of Affine Flows

Definition 7 (The set *L* of affine flows). An *affine flow* (*or rational flow*) is a flow that can be computed with a single set of stacked unit cubes. The set of all affine flows is denoted by *L*. In particular, $L \subset FL(TT)$. The loops of affine flows are called *affine loops* (*or rational loops*). Since we are concerned with the interaction between loops, we consider the contained loop to be part of the surrounding loop.
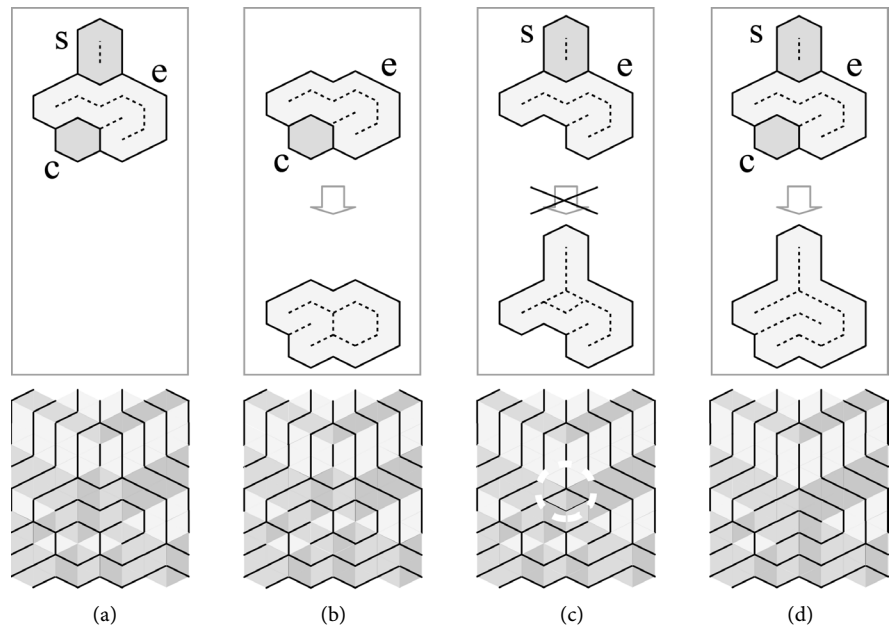
Figure 5. Allosteric regulation of loops. (a) Three loops. *e*, *s*, and *c* correspond to "Enzyme", "Substrate", and "aCtivator", respectively. Shown below is the top view of the corresponding stacked unit cubes. (b) Fusion of loop *e* and loop *c*. (c) Loops *e* and *s*. They don't fuse into a loop without *c* because of the "overhang" enclosed by the white dashed line. (d) Fusion of *e*, *s*, and *c*.

Remark 2. In order to show the similarity to the relation between the integers and the rational numbers, flows of *L* are called *rational* in the introduction.

Example 5 (Non-affine loop). Figure 4(b) gives an example of non-affine loops. We need two sets of stacked unit cubes to cover the whole shape: one for the upper half and the other for the lower half.

Definition 8 (Unit loop). Affine loops of length 6 are called *unit loops*. They always form a hexagonal shape (Figure 2(b)). Unit loops play a role similar to the role "1 (the unity)" plays in the set $\mathbb{Z}$ of integers.

Notation 1. In the following, we usually denote unit loops by a_i and x_i. Loops in general are denoted by b, b_i, c, c_i, y_i, z_i and others.

4.2. The Category *Cat(L)* of Affine Flows

Definition 9 (Binary relation "<" on *L*). Let $U, V \in L$. Then, $U < V$ if and only if there exist loops $b_1, b_2, \cdots, b_m \in lp(U)$ for any $c \in lp(V)$ such that

$$c \equiv b_1 + b_2 + \cdots + b_m.$$

In other words, *U* is a subdivision of *V* with respect to the fusion "+" operation. Note that $U < U$.

Definition 10 (*Cat(L)*). The category $Cat(L)$ of affine flows is a pair of
1) the set *L* of affine flows, and
2) the binary relation < on *L* defined above.

Definition 11 (Fusion "+" of flows). Let $V \in L$ and $\{U_1, U_2, \cdots, U_m\} \subset L$. Suppose that $lp(V) = \{c\}$ and $lp(U_i) = \{b_i\}$ ($i = 1, 2, \cdots, m$). Then, we write

$$V \equiv U_1 + U_2 + \cdots + U_m$$

if $c \equiv b_1 + b_2 + \cdots + b_m$.

Lemma 1.

1) Let $U, V \in L$. Then, $U > V$ if $lp(U) \subset lp(V)$.

2) Let $U_1, U_2 \in L$. Let $lp(U_1) = \{b_1\}$ and $lp(U_2) = \{b_2\}$. Suppose that $b_1 \neq b_2$. Then

$$U_1 < U_2 \text{ and } U_2 < U_1 \text{ if } b_1 \equiv b_2.$$

That is, "<" is not a partial order.

Definition 12 (Equivalence relation "$\equiv$" on L). Let $U, V \in L$. Then, we define an equivalent relation "$\equiv$" on L by

$$U \equiv V \text{ if and only if } U < V \text{ and } V < U.$$

Remark 3. Definition 11 and Definition 12 are consistent.

Remark 4. Let $U, V \in L$. Let $lp(U) = \{b_1, b_2, \cdots, b_m\}$ and $lp(V) = \{c_1, c_2, \cdots, c_k\}$. Then, $U \equiv V$ implies

$$m = k \text{ and } \{|b_1|, |b_2|, \cdots, |b_m|\} = \{|c_1|, |c_2|, \cdots, |c_k|\}.$$

That is, U and V are divided into the same set of shapes (*i.e.*, $|b_i| = |c_j|$ for some i and j), but the shapes may have different "folding patterns" of trajectories (*i.e.*, $b_i \neq c_j$).

Definition 13 (A terminal flow Z). Let $Z \in L$ such that $lp(Z) = \varnothing$. Then, $U < Z$ for any $U \in L$. Z is *called a terminal flow* of $Cat(L)$. Z is determined uniquely up to equivalence $\equiv$. In this paper, we consider a fixed terminal flow Z.

4.3. Basic Operations of *Cat(L)*

The least upper bound and the greatest lower bound for $S \subset L$ with respect to "<" is defined as follows [16] [17].

Definition 14 (*LB(S)* and *UB(S)*). Let $S \subset L$ be a finite subset. The set $LB(S)$ of *lower bounds* of S is defined by

$$LB(S) := \{V \in L | V < U \text{ for all } U \in S\}.$$

The set $UB(S)$ of *upper bounds* of S is defined by

$$UB(S) := \{V \in L | U < V \text{ for all } U \in S\}.$$

Definition 15 (*GLB(S)*). Let $S \subset L$ be a finite subset. A *greatest lower bound* of S is a maximal element of $LB(S)$. We denote the set of all the greatest lower bounds of S by $GLB(S)$, *i.e.*,

$$GLB(S) := \{V \in LB(S) | V \text{ is a maximal element of } LB(S)\}.$$

Note that $GLB(S)$ may contain more than one flow.

The author does not know the answer to the following question.

Question 1. Let $S \subset L$ be a finite subset. Does $U_1, U_2 \in GLB(S)$ imply $U_1 \equiv U_2$?

Definition 16 (Greatest lower bound mapping "∧" on *L*). Let $S \subset L$ be a finite subset. A *greatest lower bound mapping* "∧" is an assignment of an element of $GLB(S)$ to *S*, *i.e.*,

$$\wedge : P(L) \to L, \wedge(S) \in GLB(S),$$

where $P(L)$ is the set of all finite subsets of *L*. In this paper, we consider a fixed greatest lower bound mapping ∧.

Remark 6. Let $S \subset L$ be a finite subset. Then, $V < U$ for all $U \in S$ dose not imply $V < \wedge(S)$ because $GLB(S)$ may contain multiple flows.

Example 6. Shown in Figure 1(b) left is

$$U_1 \wedge U_2 \wedge U_3 \wedge U_4 \wedge U_5 \wedge U_6,$$

where $U_i \in L$ such that $lp(U_i) = \{b_i\}$ ($i = 1, 2, \cdots, 6$). Then,

$$lp(U_1 \wedge U_2 \wedge U_3 \wedge U_4 \wedge U_5 \wedge U_6) = \{b_1, b_2, b_3, b_4, b_5, b_6\}.$$

Definition 17 (Least upper bound mapping "∨" on *L*). A *least upper bound mapping* "∨" is defined on *L* similarly, *i.e*,

$$\vee : P(L) \to L, \vee(S) \in LUB(S).$$

where

$$LUB(S) := \{V \in UB(S) | V \text{ is a minimal element of } UB(S)\}.$$

Note that $LUB(S)$ may contain more than one flow. In this paper, we consider a fixed least upper bound mapping ∨ on *L*.

Notation 2. Let $S \subset L$. For simplicity, we often write $\vee(S)$ and $\wedge(S)$ instead of $\vee_{V \in S} V$ and $\wedge_{V \in S} V$, respectively.

Definition 18 (Consistency and disjointness). Let $U, V \in L$. *U* and *V* are called *consistent* if

$$LB(\{U, V\}) \neq \varnothing.$$

U and *V* are called *disjoint* if

$$U \vee V \equiv Z.$$

Notation 3. By abuse of notation, we often write $U \wedge V \neq \varnothing$ instead of $LB(\{U, V\}) \neq \varnothing$.

Finally, let's define a basic operation of the propositional calculus [18] [19].

Definition 19 (Exponential U^B). Let $U, V, B \in L$. An exponential U^B of *U* for *B* is defined by

$$U^B := \{W \in L | V < W \text{ if and only if } B \wedge V < U\}.$$

In particular, $B \wedge W < U$ if $W \in U^B$. Note that $U^B = \varnothing$ if $B \wedge U = \varnothing$.

Remark 7. $W \in U^B$ if and only if

$$Hom_L(B \wedge V, U) = Hom_L(V, W).$$

See the next subsection for the definition of $Hom_L(,)$.

In theoretical computer science, an "exponential object" U^B corresponds to a computer program that takes input *B* and produces output *U*. In the following,

we will specify affine loops (*i.e.*, proteins) as an exponential.

4.4. *Hom* Function of *Cat*(*L*)

Definition 20 (The category of {0,1}). Let $\{0,1\}$ be the set of two integers. A binary relation "<" is defined on $\{0,1\}$ by

$$0 < 0,\ 0 < 1 \text{ and } 1 < 1.$$

Multiplication "·" is defined on $\{0,1\}$ by

$$1 \cdot 1 = 1 \text{ and } 1 \cdot 0 = 0 \cdot 1 = 0 \cdot 0 = 0.$$

Definition 21 ($L^{\wedge}$). A $\{0,1\}$ -valued function F on L is called contravariant if $U_1 < U_2$ implies $F(U_1) > F(U_2)$ for $U_1, U_2 \in L$. We denote the set of all contravariant $\{0,1\}$ -valued functions on L by $L^{\wedge}$.

Let F and G be $\{0,1\}$ -valued functions on L. Then, multiplication $F \cdot G$ of F and G is defined by

$$(F \cdot G)(V) := F(V) \cdot G(V).$$

We sometimes write FG instead of $F \cdot G$. The multiplication of $F_1, F_2, \cdots, F_m$ is often denoted by $\Pi_i F_i$ (if there is no risk of confusion).

Definition 22 (*Hom* functions on *L*). *Let* $V \in L$. *The contravariant Hom function* $Hom_L(\ ,V) \in L^{\wedge}$ *is defined by*

$$Hom_L(U,V) = \begin{cases} 1 & \text{if } U < V \\ 0 & \text{otherwise} \end{cases}$$

Lemma 3. *Let* $U,V \in L$. *Then,* U *and* V *are consistent if and only if* $Hom_L(V \wedge U, U) = 1$.

Lemma 4. Let ι_L be the mapping from L into $L^{\wedge}$ defined by

$$\iota_L : L \to L^{\wedge},\ \iota_L(V) := Hom_L(\ ,V).$$

Then, ι_L is injective (up to equivalence $\equiv$). That is, we can regard L as a subset of $L^{\wedge}$.

Notation 4. Let $V \in L$. By identifying V with $Hom_L(\ ,V)$, we set

$$V(X) = Hom_L(X,V)$$

for $X \in L$.

Lemma 5.

1) $Hom_L(U, \wedge_i V_i) = \Pi_i Hom_L(U,V_i)$.
2) $Hom_L(\vee_i U_i, V) = \Pi_i Hom_L(U_i,V)$.
3) Suppose that $\Sigma_i V_i$ is defined. Then,

$$Hom_L(U, \Sigma_i V_i) = \Pi_i Hom_L(U,V_i) = 1 \quad \text{if } U < \wedge_i V_i.$$

Remark 8. Let $U,V \in L$. Then $V(U) = 1$ if and only if $U < V$.

Definition 23 ($L^{\wedge\wedge}$). We denote the set of all contravariant $\{0,1\}$ -valued functions on $L^{\wedge}$ by $L^{\wedge\wedge}$.

Lemma 6 (Spectral decomposition). *Let* ι_{L2} *be the mapping from L into* $L^{\wedge\wedge}$ *defined by*

$$\iota_{L2} : L \to L^{\wedge\wedge}, \iota_{L2}(U)(Y) := Y(U).$$

Then, ι_{L2} is injective (up to equivalence $\equiv$). In particular, we obtain a "*spectral decomposition*"

$$\left\{ Y \in L^{\wedge} \middle| Y(U) = 1 \right\}$$

of $U \in L$.

4.5. Loop Representation of Affine Flows

Now let's describe the loops of a flow of L using the binary relation "<" instead of the lp function.

Definition 24 (One-loop flow). Let $U \in L$. U is called a one-loop flow if

$$U < V \text{ implies } V \equiv U \text{ or } Z.$$

Definition 25 (Generators of a flow). Let $U \in L$. The set of generators of U are a set $\{V_1, V_2, \cdots, V_m\}$ of one-loop flows such that:

1) $U \equiv \wedge V_i \in L$, and
2) $\wedge_{i \in [1,m] \setminus \{k\}} V_i \nless V_k$ ($k = 1, 2, \cdots, m$).

Note that the set of generators consists of a finite number of flows and is uniquely determined by U.

Notation 5 (Direct sum "$\oplus$"). Let $U \in L$. Let $\{V_1, V_2, \cdots, V_m\}$ be a set of one-loop flows. We write

$$U \equiv \oplus_i V_i$$

if and only if

$$\begin{cases} U \equiv \wedge_i V_i, \\ V_i \vee V_j \equiv Z \text{ if } i \neq j, \left(i.e., V_i \text{ and } V_j \text{ are disjoint}\right). \end{cases}$$

Lemma 7. Let $U, V_1, V_2, \cdots, V_m \in L$ be a set of one-loop flows. Then,

$$U \equiv \Sigma_i V_i \text{ implies } U > \oplus_i V_i.$$

Example 7. Shown in **Figure 1(b)** left is

$$U_1 \oplus U_2 \oplus U_3 \oplus U_4 \oplus U_5 \oplus U_6 \ (< C_1),$$

where $U_i \in L$ such that $lp(U_i) = \{b_i\}$ ($i = 1, 2, \cdots, 6$).

Lemma 8 (Direct sum representation of a flow). Let $U \in L$. Let $S = \{V_1, V_2, \cdots, V_m\}$ be the generators of U. Let $P_i \subset S$ ($i = 1, 2, \cdots, k$) be a partition of S, *i.e.*,

$$S = P_1 \cup P_2 \cup \cdots \cup P_k \text{ and } P_i \cap P_j = \varnothing \text{ if } i \neq j.$$

Then,

$$U \equiv W_1 \oplus W_2 \oplus \cdots \oplus W_k,$$

where $W_i \equiv \wedge_{V \in P_i} V$. In particular,

$$U \equiv V_1 \oplus V_2 \oplus \cdots \oplus V_m,$$

Lemma 9 (Exponential U^B of U for B). Let $U, B \in L$. Let $U \equiv V_1 \oplus V_2 \oplus \cdots \oplus V_m$. Suppose that $B \equiv \oplus_{i \in J} V_i$ for some $J \subset [1, m]$. Then,

$$\oplus_{i \in I \setminus J} V_i \in U^B.$$

Definition 26 (Maximal-cover flow). Let $U, V \in L$ and $V < U$. V is called a maximal-cover flow of U if

$$V < W < U \quad \text{implies} \quad W \equiv V \text{ or } U.$$

Definition 27 (Co-generators of a flow). Let $U \in L$. The set of *co-generators* of U are a set $\{V_1, V_2, \cdots, V_m\}$ of maximal-cover flows of U such that:

1) $U \equiv \vee V_i \in L$, and
2) $V_i \not\prec V_j$ if $i \neq j$.

Note that the set of co-generators consists of a finite number of flows and is uniquely determined by U.

Definition 28 (Consistent component of co-generators). Let $U \in L$. Let $S = \{W_1, W_2, \cdots, W_m\}$ be the set of co-generators of U. Let $W_k \in S$. The subset defined by

$$c\text{-}comp(W_k) := \{W_i \in S \mid W_i \wedge W_k \neq \varnothing\}$$

is called the *component consistent with* W_k. Then, there exist $\{W_{j1}, W_{j2}, \cdots, W_{jk}\} \subset S$ such that

$$\begin{cases} S = c\text{-}comp(W_{j1}) \bigcup c\text{-}comp(W_{j2}) \bigcup \cdots \bigcup c\text{-}comp(W_{jk}), \\ c\text{-}comp(W_{js}) \neq c\text{-}comp(W_{jt}) \quad \text{if} \quad js \neq jt. \end{cases}$$

$c\text{-}comp(W_{jk})$'s are called the *consistent components* of U.

Remark 9. If $c\text{-}comp(W_{jk})$ is a consistent components of U, then

$$\begin{cases} \wedge \left(c\text{-}comp(W_{jk}) \right) \neq \varnothing, \\ c\text{-}comp(W_{jk}) < U. \end{cases}$$

Notation 6 (Direct product "$\otimes$"). Let $U \in L$. Let $\{V_1, V_2, \cdots, V_m\} \subset L$ be a set of maximal-cover flows of U. We write

$$U \equiv \otimes_i V_i$$

if and only if

$$\begin{cases} U \equiv \vee_i V_i, \\ V_i \wedge V_j = \varnothing \text{ if } i \neq j, \left(i.e., V_i \text{ and } V_j \text{ are not consistent} \right). \end{cases}$$

Lemma 10. Let $U, V_1, V_2, \cdots, V_m, W_i, W_2, \cdots, W_k \in L$ be a set of one-loop flows. Suppose that $V_i \wedge W_j = \varnothing$ for all i, j. Then,

$$\begin{cases} U \equiv \Sigma_i V_i, \\ U \equiv \Sigma_j W_j \end{cases} \quad \text{implies} \quad U > \left(\oplus_i V_i \right) \otimes \left(\oplus_j W_j \right).$$

Lemma 11 (Direct product representation of a flow). Let $U \in L$. Let

$$U \equiv W_1 \vee W_2 \vee \cdots \vee W_m$$

be a maximal-cover representation of U. Let

$$\left\{ c\text{-}comp(W_{j1}), c\text{-}comp(W_{j2}), \cdots, c\text{-}comp(W_{jk}) \right\}$$

be the consistent components of U. Then,

$$U \equiv \wedge\left(c\text{-}comp\left(W_{j1}\right)\right) \otimes \wedge\left(c\text{-}comp\left(W_{j1}\right)\right) \otimes \cdots \otimes \wedge\left(c\text{-}comp\left(W_{jk}\right)\right).$$

Example 8. In the case of **Figure 1**, we have

$$C_1 > \left(U_1 \oplus U_2 \oplus \cdots \oplus U_6\right) \otimes \left(U_7 \oplus U_8 \oplus \cdots \oplus U_{12}\right).$$

Example 9. In the case of **Figure 4(a)**, we have

$$C_1 \equiv \left(B_1 \oplus B_2\right) \otimes \left(B_3 \oplus B_4\right) \otimes \left(U_3\right),$$

where $lp(U_1) = \{b_1, b_2\}$, $lp(U_2) = \{b_3, b_4\}$, and $lp(B_i) = \{b_i\}$ ($i = 1, 2, 3, 4$). That is,

$$\begin{cases} C_1 \equiv \left(B_1 \wedge B_2\right) \vee \left(B_3 \wedge B_4\right) \vee U_3, \\ \left(B_1 \wedge B_2\right) \wedge \left(B_3 \wedge B_4\right) = \varnothing, \\ \left(B_1 \wedge B_2\right) \wedge U_3 = \varnothing, \\ \left(B_3 \wedge B_4\right) \wedge U_3 = \varnothing. \end{cases}$$

4.6. The Fusion-Structure Presheaf *F* on the Sup Topology *K*

In quantum mechanics, each physical system is associated with a collection of data obtained when measuring some physical property of the system, *i.e.*, the spectrum of an observable quantity associated with the physical property. In our model of proteins, each protein, *i.e.*, "loop", is associated with the "spectrum" of the measurable quantity, *i.e.*, "shape".

Remark 10. Discrete differential geometry of triangles [1] provides a mechanism by which the observed result comes about.

Definition 29 (Sup topology *K*). The *sup topology K* on *L* is defined by

$$K(U) := \{W \in L | W < U\} \quad (U \in L).$$

Note that

$$\begin{cases} U = \vee\left(K(U)\right), \\ K(U) \subset K(V) \text{ if } U < V. \end{cases}$$

Elements of $K(U)$ are called *states* of the "biological entity" *U*.

Definition 30 (Fusion-structure presheaf *F*). Fusion-structure presheaf *F* on *L* is defined by

$$F(W) := \{V \in L | W < V\} \quad (W \in L).$$

Note that

$$\begin{cases} W = \wedge\left(F(W)\right), \\ F(U) \supset F(V) \text{ if } U < V. \end{cases}$$

Recall that $F(W)$ is the spectral decomposition of W (See Lemma 6). That is, $F(W)$ is the spectoral decomposition of the state W of the "biological entity" *U*.

Remark 11. As for the sup topology and presheaf, a number of standard textbooks are available, including [17] and [19].

Example 10. In the case of **Figure 4(a)**, we have

$$\begin{cases} K(U_1) = \{U_1, \cdots\}, & F(U_1) = \{B_1, B_2, U_1\}, \\ K(U_2) = \{U_2, \cdots\}, & F(U_2) = \{B_3, B_4, U_2\}, \\ K(U_3) = \{U_1, U_2, U_3, \cdots\}, & F(U_3) = \{U_3\}, \end{cases}$$

where B_i ($i = 1, 2, 3, 4$) are defined in Example 9.

Definition 31 (*L*-spectrum). Let $U \in L$. The *L*-spectrum of *U* is defined by

$$L\text{-}spec(U) := \bigcup \{F(W) | W \in K(U)\} = \{V \in L | W < V \text{ and } W < U\}.$$

Note that $L\text{-}spec(U)$ captures all the observable properties of the "quantum system" *U* of *L*.

Example 11. In the case of **Figure 4(a)**, we have

$$\begin{cases} L\text{-}spec(U_1) = F(U_1) \cup \cdots, \\ L\text{-}spec(U_2) = F(U_2) \cup \cdots, \\ L\text{-}spec(U_3) = F(U_1) \cup F(U_2) \cup F(U_3) \cup \cdots. \end{cases}$$

5. Integral Flows L^A

Here we define the other set L^A of flows mentioned in the introduction. Using the binary relation defined on L^A, we define the set $\left(L^A\right)^\wedge$ of $\{0,1\}$-valued functions on L^A.

5.1. The Category *Cat*(L^A) of Integral Flows

Definition 32 (The set L^A of integral flows). Let $A \in L$ be the flow consisting of only unit loops (**Figure 6(a)**). A flow $U \in L$ is called integral if $A < U$. The set of all integral flows is denoted by L^A, *i.e.*,

$$L^A := \{U \in L | A < U\} \subset L.$$

A is called the *base flow* of L^A. Loops of integral flows are called *integral loops*.

Remark 12. "$A < U$" means that each loop of *U* is obtained by fusing a set of loops of *A*. That is, *A* spawns all flows of L^A. In other words, *A* is a kind of "quantum vacuum". On the other hand, the terminal object *Z* does not generate anything because *Z* contains no loop. That is, *Z* is a kind of "classical vacuum".

Binary relations "<" and "≡" as well as mappings "∨" and "∧" are already defined on L^A because $L^A \subset L$. For example,
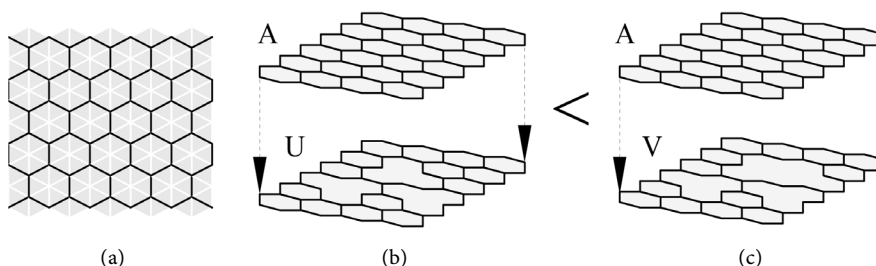

(a)   (b)   (c)

Figure 6. Integral flows. (a) The base flow *A*, consisting of only unit loops. (b) The binary relation $U < V$ on L^A.

$$U < V \quad \text{in } L^A \text{ if and only if } \quad U < V \quad \text{in } L$$

where $U, V \in L^A$ (**Figure 6(b)**).

Definition 33 ($Cat(L^A)$). The category $Cat(L^A)$ of integral flows is a pair of

1) the set L^A of integral flows, and

2) the binary relation < on L^A,

Lemma 12. Let $S \subset L^A$ be a finite subset. Then, $\vee(S), \wedge(S) \in L^A$.

Definition 34 (lp_A function). Let $U \in L^A$. The loop function lp_A is defined on L^A by

$$lp_A(U) := \{b \in lp(U) \mid b \text{ is not a unit loop}\}.$$

Loops of L^A are "measurable with unit loops" (*i.e.*, "computable") because of the proposition below.

Proposition 1. Let $U \in L^A$. Let $b \in lp_A(U)$. Then, there exists a unique finite subset $\{a_1, a_2, \cdots, a_m\}$ of unit loops such that

$$b \equiv a_1 + a_2 + \cdots + a_m.$$

That is, every integral loop is a fusion of unit loops.

Proof. It follows immediately from the definition of "<". □

Remark 13. Unit loops have the same shape, but behave differently with respect to "+". That is, they form different shapes depending on their relative placement. Because of the property, the algebra of integral flows is also called "hetero numbers", *i.e.*, numbers with "heterogeneous" unity elements.

The following proposition gives a characterization of allosteric loops (See Definition 6).

Proposition 2. Integral loops are not allosteric.

Proof. (Sketch of proof) Let $U \in L^A$. Let $b \in lp_A(U)$. Then, there are a set of unit loops $\{a_1, a_2, \cdots, a_m\}$ surrounding b. Suppose that there exists a_i and a_j such that $b + a_i + a_j$ is defined, but $b + a_j$ is not. As shown in **Figure 3(c)**, allosteric regulation is due to a "overhang" of a_j over a_i, *i.e.*, the difference of the "height" of a_i and a_j. On the other hand, a_i's are at the same "height" because they are surrounding an integral loop. In particular, (the absence of) a_i does not prevent a_j from interacting with b. □

5.2. *Hom* Function of $C(L^A)$

Definition 35 ($(L^A)^{\wedge}$). We denote the set of all contravariant $\{0,1\}$-valued functions on L^A by $(L^A)^{\wedge}$. Multiplication "·" is defined on $(L^A)^{\wedge}$ in the same way as $L^{\wedge}$.

Definition 36 (*Hom* functions on L^A). Let $V \in L^A$. The *contravariant Hom function* $Hom_A(\ , V) \in (L^A)^{\wedge}$ is defined by

$$Hom_A(U, V) = \begin{cases} 1 & \text{if } U < V \\ 0 & \text{otherwise} \end{cases}$$

Notation 7. Let $V \in L$. By identifying V with $Hom_A(\ , V)$, we set

$$V(X) = Hom_A(X, V)$$

for $X \in L^A$.

Definition 37 ($\left(L^A\right)^{\wedge\wedge}$). We denote the set of all contravariant $\{0,1\}$ -valued functions on $\left(L^A\right)^{\wedge}$ by $\left(L^A\right)^{\wedge\wedge}$.

Lemma 13 (Spectral decomposition).

1) Let ι_A be the mapping from L^A into $\left(L^A\right)^{\wedge}$ defined by

$$\iota_A : L^A \to \left(L^A\right)^{\wedge}, \iota_A(V)(X) \coloneqq V(X).$$

Then, ι_A is injective (up to equivalence " $\equiv$ "). That is, we can regard L^A as a subset of $\left(L^A\right)^{\wedge}$.

2) Let ι_{A2} be the mapping from L^A into $\left(L^A\right)^{\wedge\wedge}$ defined by

$$\iota_{A2} : L^A \to \left(L^A\right)^{\wedge\wedge}, \iota_{A2}(U)(Y) \coloneqq Y(U).$$

Then, ι_{A2} is injective (up to equivalence " $\equiv$ "). In particular, we obtain a *spectral decomposition*

$$\left\{ Y \in \left(L^A\right)^{\wedge} \middle| Y(U) = 1 \right\}$$

of $U \in L^A$.

Lemma 14.

1) $Hom_A(U, \wedge_i V_i) = \Pi_i Hom_A(U, V_i)$.

2) $Hom_A(\vee_i U_i, V) = \Pi_i Hom_A(U_i, V)$.

3) Suppose that $\Sigma_i V_i$ is defined. Then,

$$Hom_A(U, \Sigma_i V_i) = \Pi_i Hom_A(U, V_i) = 1 \quad \text{if } U < \wedge_i V_i.$$

Lemma 15 (Factorization of $U \in L^A$ in $L^{\wedge}$). *Let $U \in L^A$. Let $V_1, V_2, \cdots, V_m \in L$ such that $U \equiv V_1 + V_2 + \cdots + V_m$. Then,*

$$Hom_L(Y, U) = \Pi_i Hom_L(Y, V_i) = 1 \quad \text{if } Y < \wedge_i V_i,$$

where $Y \in L$.

Example 12. In the case of Figure 1, two factorizations of $C_1 \in L^A$ are considered *i.e.*,

$$\begin{cases} C_1(Y) = B_1 B_2 B_3 B_4 B_5 B_6(Y) = 1 & \text{if } Y < B_1 \wedge B_2 \wedge \cdots \wedge B_6, \\ C_1(Y) = B_7 B_8 B_9 B_{10} B_{11} B_{12}(Y) = 1 & \text{if } Y < B_7 \wedge B_8 \wedge \cdots \wedge B_{12}, \end{cases}$$

where $Y \in L$.

5.3. Closure of Affine Loops

First, we define an operation that replaces the "background" open trajectories of an integral flow with a set of unit loops.

Definition 38 (Localization/unlocalization). Let $U \in L^A$. A *localization* $loc(U)$ of U is the integral flow defined by

$$loc(U) \coloneqq U \vee A.$$

An *unlocalization* $unloc(U)$ of U is the integral flow defined by

$$unloc(U) \coloneqq \wedge \left(\left\{ B \in L^A \middle| U < B, B \text{ is a one-loop flow} \right. \right.$$

$$\left. \left. \text{such that } lp(B) = lp_A(B) \right\} \right),$$

where we set $\wedge(S) = Z$ if $S = \varnothing$.

Example 13. $loc(Z) = A$ and $unloc(A) = Z$.

Using the unlocalization operation, we define a projection of flows of L onto flows of L^A, *i.e.*, the "shadows" of affine flows on L^A.

Definition 39 (Closure *cl*(*B*) of $B \in L^A$). Let $B \in L$. Let $CL(B) \subset L^A$ be the set of all the integral flows that is consistent with B, *i.e.*,

$$CL(B) := \left\{ U \in L^A \big| B \wedge U \neq \varnothing \right\}.$$

Let $Loc\text{-}CL(B) \subset L^A$ be the set of all the minimal elements of $CL(B)$, *i.e.*,

$$Loc\text{-}CL(B) := \left\{ C \in L^A \big| C \text{ is a minimal element of } CL(B) \right\}.$$

A *closure mapping* $cl : L \to L^A$ is defined on L by

$$cl(B) = \begin{cases} B & \text{if } B \in L^A \\ unloc(E) \text{ for some } E \in Loc\text{-}CL(B) & \text{otherwise} \end{cases}$$

Note that $Loc\text{-}CL(B)$ may contain more than one flow. In this paper, we consider a fixed closure mapping cl on L. $cl(B)$ is called the *closure* (or *shadow*) of *B* on L^A.

Remark 14. In the introduction, it is implicitly assumed that $lp(cl(B))$ consists of one loop for any $B \in L$ (because of the definition of fusion).

The author does not know the answer to the following questions.

Question 2.

1) Let $B \in L$. Does $Loc\text{-}CL(B)$ consist of one flow?

2) Can one define the closure mapping cl in such a way that $lp(cl(B))$ consists of one loop for all $B \in L$?

Example 14. Figure 7(d) shows the closure $cl(B_1)$ of a flow B_1. In this case, $Loc\text{-}CL(B_1) = \{V\}$ and $cl(B_1) = unloc(V)$.
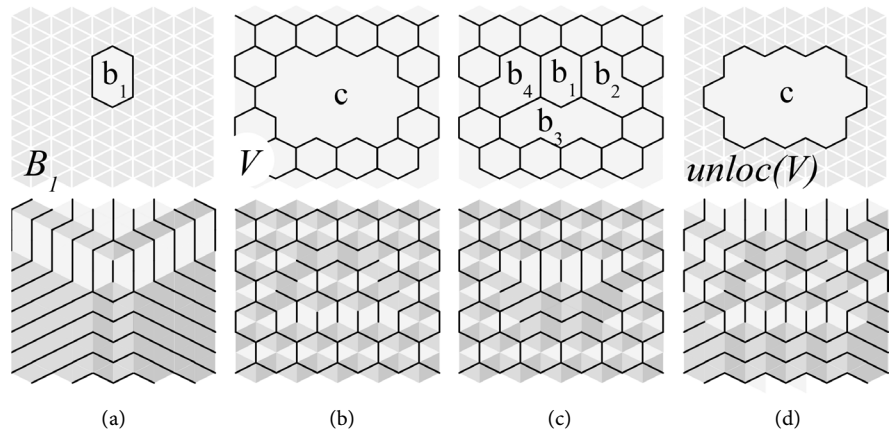

Figure 7. Closure of affine loops. (a) A flow $B_1 \in L$ consisting of one loop b_1 and open trajectories. Shown below is the top view of the corresponding stacked unit cubes. (b) The only flow V of $Loc\text{-}CL(B_1)$. (c) A fission of the loop c of V. Note that $B_2 \oplus B_3 \oplus B_4 \in V^{B_1}$, where B_i's are integral flows such that $lp_A(B_i) = \{b_i\}$. (d) The closure $cl(B_1) = unloc(V) \in L^A$.

Example 15. In the case of **Figure 1**,

$$\begin{cases} C_1 = cl(B_1), \\ C_i = cl(B_i) \ (i = 1, \cdots, 12). \end{cases}$$

Note that some pairs of flows have the same closure. For example,

$$cl(B_1) \equiv cl(B_7) \quad \text{and} \quad cl(B_2) \equiv cl(B_8).$$

Proposition 3. Let $U \in L$. Suppose that $lp(cl(U))$ consists of one loop. Then, there exists $W \in cl(U)^U$ such that

$$cl(U) \equiv U + W.$$

Moreover, there exist one-loop flows $V_1, V_2, \cdots, V_m \in L$ such that

$$cl(U) \equiv U + V_1 + V_2 + \cdots + V_m,$$

where

$$W \equiv V_1 \oplus V_2 \oplus \cdots \oplus V_m.$$

Proof. (Sketch of proof) By definition of $CL(B)$, there exists $V \in L$ such that

$$V < U \quad \text{and} \quad V < cl(U).$$

Suppose that $lp(U) = \{u\}$ for simplicity. Let $lp(cl(U)) = \{c\}$. By definition of "$<$", there exist loops $v_1, v_2, \cdots, v_k \in V$ and $b_1, b_2, \cdots, b_m \in V$ such that

$$\begin{cases} u \equiv v_1 + v_2 + \cdots + v_k, \\ c \equiv v_1 + v_2 + \cdots + v_k + b_1 + b_2 + \cdots + b_m. \end{cases}$$

Then,

$$c \equiv u + b_1 + b_2 + \cdots + b_m.$$

$\square$

Remark 15. Suppose that $lp(cl(U))$ consists of multiple loops, *i.e.*,

$$lp(cl(U)) = \{c_1, c_2, \cdots, c_l\}.$$

Let $V \in L$ such that

$$V < U \quad \text{and} \quad V < cl(U).$$

Then, there exist loops $v_1, v_2, \cdots, v_k \in lp(V)$ and $b_{i,1}, b_{i,2}, \cdots, b_{i,m_i} \in lp(V)$ ($i = 1, 2, \cdots, l$) such that

$$\begin{cases} u \equiv v_1 + v_2 + \cdots + v_k, \\ c_i \equiv \Sigma_{j \in P_i} v_j + b_{i,1} + b_{i,2} + \cdots + b_{i,m_i} \quad (i = 1, 2, \cdots, l), \end{cases}$$

where $P_1, P_2, \cdots, P_l$ is a partition of $[1, k]$. It follows that there exists $W \in L$ such that $V < W$ and

$$|cl(u)| = |c_1| \bigcup |c_2| \bigcup \cdots \bigcup |c_l| = |u| \bigcup |w_1| \bigcup |w_2| \bigcup \cdots \bigcup |w_n|,$$

where $lp(W) = \{u, w_1, w_2, \cdots, w_n\}$.

Remark 16. Recall that categories $Cat(L)$ and $Cat(L^A)$ consists of a set of "objects" and a binary relation "$<$" on them. However, the closure mapping does

not preserve the binary relation, *i.e.,*

$$U < V \quad \text{dose not imply} \quad cl(U) < cl(V).$$

Lemma 17. Let $U \in L$. Suppose that $Loc\text{-}CL(U)$ consists of only one flow, *i.e.,*

$$Loc\text{-}CL(U) = \{loc(cl(U))\}.$$

If $cl(U) \not\equiv cl(V)$, then

$$Hom_L(U,V) = 0 \quad \text{implies} \quad Hom_A(cl(U), cl(V)) = 0.$$

Remark 17. *Let $U \in L$. If $Loc\text{-}CL(U)$ consists of only one flow, then*

$$Y \wedge U = \varnothing \quad \text{if and only if} \quad Y \wedge cl(U) = \varnothing$$

for $Y \in L^A$.

6. Defining Equations of Affine Loops

Now let's consider the problems mentioned in the introduction step by step. For intuitive understanding, we use loops instead of flows in this section.

6.1. In the Case of Self-Computable Loops

We begin by considering the second half of Problem 2.

Problem 3. Find integral loops $y_1, y_2, \cdots, y_m$ such that

$$\begin{cases} y_1 \equiv x_1 + x_2 + \cdots + x_n, \\ y_2 \equiv f_2(x_1, x_2, \cdots, x_n), \\ y_3 \equiv f_3(x_1, x_2, \cdots, x_n), \\ \cdots, \\ y_m \equiv f_m(x_1, x_2, \cdots, x_n), \end{cases}$$

where $x_1, x_2, \cdots, x_n$ are unit loops, and $f_1, f_2, \cdots, f_m$ are fusions of a subset of $\{x_1, x_2, \cdots, x_n\}$.

Remark 18. For simplicity, the author uses "=" instead of "≡" in the introduction.

Definition 40 (Self-computable loops). Let $b_1, b_2, \cdots, b_m$ be integral loops. Let $a_1, a_2, \cdots, a_n$ be unit loops. Suppose that

$$\begin{cases} y_1 = b_1, \ y_2 = b_2, \ \cdots, \ y_m = b_m, \\ x_1 = a_1, \ x_2 = a_2, \ \cdots, \ x_n = a_n \end{cases}$$

is a solution to Problem 3. Then, we call b_1 a *self-computable* loop. b_1 is determined uniquely up to the rotational and mirror symmetries by m fusion interactions $f_1, f_2, \cdots, f_m$ between $a_1, a_2, \cdots, a_n$.

Example 16 (Loop triangles of size two). Let's consider an integral loop y_1 such that

$$y_1 \equiv x_1 + x_2 + x_3,$$

where x_1, x_2, x_3 are unit loops. That is, there is a fusion interaction between the three unit loops. (Note that there is no interaction between two unit loops.) **Figure 8(b)** gives a solution to the problem, *i.e.,*

$$\begin{cases} y_1 = b_1, \\ x_1 = a_1,\ x_2 = a_2,\ x_3 = a_3. \end{cases}$$

y_1 is called a *loop triangle of size two*.

Example 17 (Loop triangles of size three). Let's consider integral loops y_1, y_2, y_3, y_4 such that

$$\begin{cases} y_4 \equiv x_1 + x_2 + x_3 + x_4 + x_5 + x_6, \\ y_1 \equiv x_1 + x_2 + x_3, \\ y_2 \equiv x_2 + x_4 + x_5, \\ y_3 \equiv x_3 + x_5 + x_6, \end{cases}$$

where $x_1, x_2, \cdots, x_6$ are unit loops. That is, there are four interactions between six unit loops. **Figure 8(b)** and **Figure 8(c)** give a solution to the problem, *i.e.*,

$$\begin{cases} y_1 = b_1,\ y_2 = b_2,\ y_3 = b_3,\ y_4 = b_4, \\ x_1 = a_1,\ x_2 = a_2,\ x_3 = a_3,\ x_4 = a_4,\ x_5 = a_5,\ x_6 = a_6. \end{cases}$$

y_4 is called a *loop triangle of size three*.

For any positive integer n, loop triangles of size n are similarly defined.

Lemma 18. Loop triangles of size n are self-computable for any positive integer n.

6.2. In the Case of Integral Loops

Now let's consider the first half of Problem 2. Subsets of a self-computable loop are obtained as a solution to the following problem.

Problem 4. *Find integral loops* $z, y_1, y_2, \cdots, y_m$ *such that*

$$\begin{cases} z \equiv g(x_1, x_2, \cdots, x_n), \\ y_1 \equiv x_1 + x_2 + \cdots + x_n, \\ y_2 \equiv f_2(x_1, x_2, \cdots, x_n), \\ y_3 \equiv f_3(x_1, x_2, \cdots, x_n), \\ \cdots, \\ y_m \equiv f_m(x_1, x_2, \cdots, x_n), \end{cases}$$

where $x_1, x_2, \cdots, x_n$ are unit loops and $g, f_1, f_2, \cdots, f_m$ are fusions of a subset of $\{x_1, x_2, \cdots, x_n\}$.

By embedding a given integral loop in a self-computable loop, we obtain a set of defining equations of the integral loop. For example, by choosing a large positive integer n, one can embed the loop in a loop triangle of size n.

Lemma 19. Integral loops are obtained as a solution to Problem 4.

Example 18. By embedding the loops $c_1, c_2, \cdots, c_6$ of **Figure 1** in a loop triangle of size 7 (**Figure 9**) at once, we obtain

$$\begin{cases} c_1 \equiv a_4 + \cdots + a_{10} + a_{12} + \cdots + a_{15} + a_{19} + \cdots + a_{21} + a_{26} + a_{27}, \\ c_2 \equiv a_4 + a_5 + a_7 + a_8 + a_9 + a_{12} + a_{13}, \\ c_3 \equiv a_4 + \cdots + a_{10} + a_{12} + a_{13} + a_{14}, \\ c_4 \equiv a_5 + a_6 + a_8 + a_9 + a_{10} + a_{13} + a_{14}, \\ c_5 \equiv a_{14} + a_{15} + a_{19} + a_{20} + a_{21} + a_{26} + a_{27}, \\ c_6 \equiv c_5. \end{cases}$$
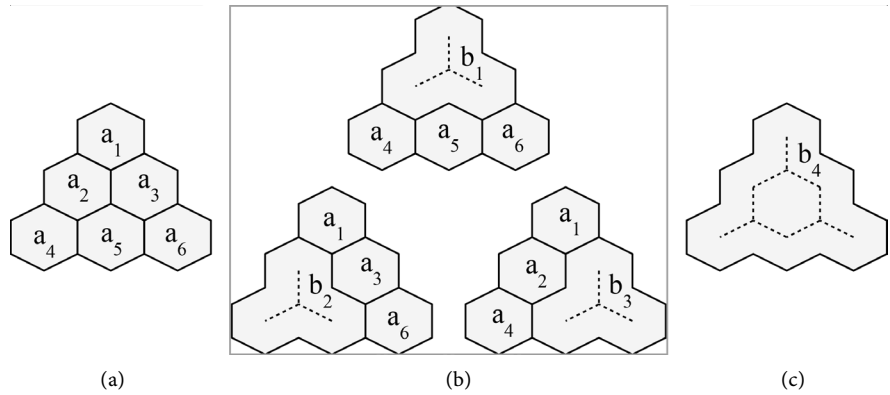
Figure 8. Loop triangles. (a) Six unit loops $a_1, a_2, \cdots, a_6$ of the base flow A. (b) Loop triangles of size two. $b_1 \equiv a_1 + a_2 + a_3$, $b_2 = a_2 + a_4 + a_5$, and $b_3 \equiv a_3 + a_5 + a_6$. (c) A loop triangle of size three $b_4 \equiv a_1 + a_2 + a_3 + a_4 + a_5 + a_6$. (Recall that the contained loop is considered to be part of the surrounding loop.)
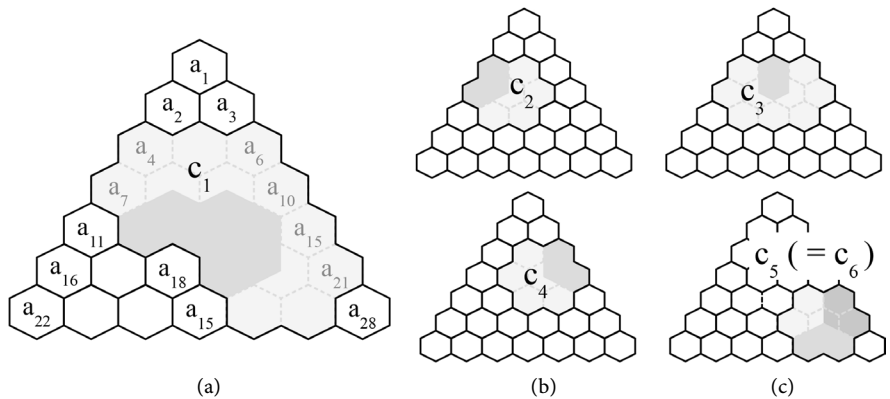

Figure 9. Subsets of loop triangles. (a) A loop triangle (consisting of $a_1, a_2, \cdots, a_{28}$) of size 7 and the integral loop c_1 of **Figure 1**, where the position of b_1 is colored dark grey. (b) Integral loops c_i of **Figure 1**, where the position of the corresponding b_i is colored dark grey.

That is,

$$\begin{cases} z_1 = c_1, \ z_2 = c_2, \ \cdots, \ z_6 = c_6, \\ x_1 = a_1, \ x_2 = a_2, \ \cdots, \ x_{28} = a_{28}. \end{cases}$$

gives a solution to Problem 2 of the introduction.

6.3. In the Case of Non-Integral Loops

Finally, let's consider Problem 1 of the introduction.

Problem 5. For a given set of flows $\{C_1, C_2, \cdots, C_m\} \subset L^A$, find a flow $W_1 \in L$ such that there exists a set of flows $\{W_2, W_3, \cdots, W_m\} \subset L$ that satisfies

$$\begin{cases} cl(W_1) \equiv W_1 + W_2 + \cdots + W_m \quad (\text{fusion}), \\ cl(W_i) \equiv C_i \quad (i = 1, 2, \cdots, m). \end{cases}$$

Remark 19. For simplicity, the author uses "=" instead of "$\equiv$" in the introduction. The dual version of Problem 5 provides a simpler description of the

problem (See below).

Remark 20. In Problem 5, it is implicitly assumed that $lp(cl(W_1))$ consists of one loop (because of the definition of fusion).

We denote the set of all the solutions to the problem by

$$C_1/C_2C_3C_4C_5\cdots C_m.$$

By analogy with quantum mechanics, C_1 is called the "output spectral flow", and $C_2, C_3, \cdots, C_m$ are called the "input spectral flows".

Example 19. Let $U_1 \in L$. Then,

$$U_1 \in cl(U_1)/cl(U_2)cl(U_3)\cdots cl(U_m),$$

if $cl(U_1) \equiv U_1 \oplus U_2 \oplus \cdots \oplus U_m$ and $lp(cl(U_1))$ consists of one loop.

Example 20. In the case of **Figure 1**,

$$\begin{cases} C_1/C_2C_3C_4C_5C_6 = \{B_1\}, \\ C_1/C_2C_3C_{10}C_{11}C_6 = \{B_7\}. \end{cases}$$

Moreover, since $cl(B_5) \equiv cl(B_6) \equiv cl(B_{12}) \equiv C_5$, we have

$$\{B_5, B_6, B_{12}\} \subset C_5/C_5C_5.$$

The dual version of Problem 5 mentioned above is given by

Problem 6. For a given set of flows $\{C_1, C_2, \cdots, C_m\} \subset L^A$, find a flow $W_1 \in L$ such that there exists a set of flows $\{W_2, W_3, \cdots, W_m\} \subset L$ that satisfies

$$\begin{cases} cl(W_1)(Y) > W_1W_2\cdots W_m(Y) \text{ for } Y \in L^A, \\ cl(W_i)(Y) = C_i(Y) \text{ for } Y \in L \quad (i = 1, 2, \cdots, m). \end{cases}$$

Remark 21. Let $U, W \in L$. If $cl(U) \not\equiv cl(W)$, then

$$W(U) = 0 \text{ implies } cl(W)(cl(U)) = 0.$$

7. Discussion

The author has shown that it is possible to describe protein-like shapes using algebraic equations if one takes a category theoretical approach. Along the way, also provided is the basis for describing the relation between various fissions of a shape, *i.e.*, *L-spectrum* of a flow $U \in L$, albeit briefly.

In the previous attempts [6] [7], the author tried to describe the shape of a loop by considering the relation between all the fissions of the loop, *i.e.*, the *L-spectrum* of the corresponding flow. Although transitions between states correspond to the addition or removal of unit cubes in the "stacked-unit-cube description", it is not yet clear how to describe the relation between various fissions of the given shape.

In this study, the author obtains a set of defining equations of a given shape by assuming the existence of integral flows. That is, a fusion-based approach rather than the previous fission-based approach. By referring to the construction of the rational numbers, he describes the shape of loops algebraically in a way that is compatible with the description style of quantum mechanics.

One of the features of the study is that it deals with both protein folding and protein shape simultaneously. This provides us with a simple model of allostery. Another feature is that it expresses both geometrical and algebraic features of protein interactions simultaneously. This allows us to describe the shape of proteins algebraically. The author is not aware of any similar multifaceted studies by other researchers on protein shape. However, they are results for a simple model and he considers this to be a starting point rather than an end point of the study.

In relation to experimental studies, the author hopes that this study will provide options on the underlying general principles that govern the engineering of self-assembling molecules such as proteins. For example, our model indicates that

allostery is a remnant of a giant protein molecule, i.e., interaction between a surface region and the core of the molecule.

It also proposes an approach to the question "How to obtain a well-defined shape with desired properties by folding a chain of subunits" [20]. That is,

first construct the shape with unity elements, then divide it into a set of folded chains.

(Of course, in order to do that, we need to find something that can be used as a unity element.)

However, so far there are no experimental results to support these proposals.

8. Conclusions

Protein design starts with a specification. Since the function of proteins is largely determined by their shape, the specification should include a description of the shape of the target protein. This paper proposes a novel category theoretic approach to describe protein's shape, *i.e.*, a description of their shape by a set of algebraic equations.

However, this paper considers the approach in a very simplified model. That is, in the two-dimensional case of the mathematical toy model proposed in [1]. Moreover, some notions were only defined and not specifically considered.

The author hopes that this paper will serve as one of the starting points for a variety of related researches. Shown below are some of the future research directions the author considers:

1) Base change.

The base flow A is obtained by partitioning $\mathbb{R}^2$ into hexagons. On the other hand, flows of triangles on a closed surface are considered in [20]. By changing the base flow A with another base flow (*i.e.*, another flow consisting of only unit loops), we can consider flows on various surfaces. For example, the surface flow on a rhombic dodecahedron corresponds to a base flow consisting of four unit loops.

2) Fission of unity elements (Higher dimensional cases).

In the case of flows of tetharadrons (*i.e.*, 3-simplices), unity elements are

rhombic dodecahedral loop of length 24. Unlike the case of triangles (*i.e.*, 2-simplices), they split into four loops of length 6 [1]. In general, unity elements of flows of *n*-simplices have fissions if $n > 2$. It is a challenge to describe fissions of unity elements in higher dimensional cases.

3) Characterization of allosteric proteins.

Allostery is at play in all processes in the living cell, and increasingly in drug discovery [2]. Our model indicates that allosteric proteins are produced as a result of the fission of larger protein molecules. That is, in our model, the design of allosteric proteins corresponds to the specification of fission of larger proteins. However, the author does not know how to specify the fission of proteins (*i.e.*, the fission of loops of *n*-simplices).

4) Symmetry of the shape of proteins.

Here the "symmetry" means the relation between the flows of *L-spectrum* defined in Subsection 4.6. The question is to what extent the symmetry controls the shape? In particular, for a given *L-spectrum*, find a flow *U* that has the same *L-spectrum*.

5) "Biologic" logic.

The relations between proteins (*i.e.*, loops) are defined using the fusion and fission of proteins, where allostery (*i.e.*, global interaction) occurs naturally [3]. What kind of logic can be made by building theories based on the fusion and fission of proteins?

Our model may be too simple to describe the ecology of actual proteins. But it is said that the simpler the model, the broader the range of applications. The author believes that this paper not only provides a new perspective on protein engineering, but also promotes further collaboration between biology and other disciplines.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Morikawa, N. (2017) Discrete Differential Geometry and the Structural Study of Protein Complexes. *Open Journal of Discrete Mathematics*, **7**, 148-164. https://doi.org/10.4236/ojdm.2017.73014

[2] Tsai, C.-J. and Nussinov, R. (2014) A Unified View of "How Allostery Works". *PLoS Computer Biology*, **10**, e1003394. https://doi.org/10.1371/journal.pcbi.1003394

[3] Morikawa, N. (2018) Global Geometrical Constraints on the Shape of Proteins and Their Influence on Allosteric Regulation. *Applied Mathematics*, **9**, 1116-1155. https://doi.org/10.4236/am.2018.910076

[4] Michalopoulos, I., Torrance, G.M., Gilbert, D.R. and Westhead, D.R. (2004) TOPS: An Enhanced Database of Protein Structural Topology. *Nucleic Acids Research*, **32**, D251-D254. https://doi.org/10.1093/nar/gkh060

[5] Taylor, W.R., Chelliah, V., Hollup, S.M., MacDonald, J.T. and Jonassen, I. (2009) Probing the "Dark Matter" of Protein Fold Space. *Structure*, **17**, 1244-1252. https://doi.org/10.1016/j.str.2009.07.012

[6] Morikawa, N. (2003) Research Project: Toward Galois Theory of Protein-Like Objects. http://www.genocript.com/papers/project_Galois.pdf

[7] Morikawa, N. (2004) Research Project: Protein as Numbers. http://www.genocript.com/papers/project_numbers.pdf

[8] Rosen, R. (1958) The Representation of Biological Systems from the Standpoint of the Theory of Categories. *Bulletin of Mathematical Biophysics*, **20**, 317-341. https://doi.org/10.1007/BF02477890

[9] Leteliera, J.-C., Soto-Andradeb, J., Abarzua, G.F., Cornish-Bowden, A. and Cárdenas, M. (2005) Organizational Invariance and Metabolic Closure: Analysis in Terms of (*M*, *R*) Systems. *Journal of Theoretical Biology*, **238**, 949-961. https://doi.org/10.1016/j.jtbi.2005.07.007

[10] Varenne, H. (2013) The Mathematical Theory of Categories in Biology and the Concept of Natural Equivalence in Robert Rosen. *Revue d'Histoire des Sciences*, **66**, 167-197. https://doi.org/10.3917/rhs.661.0167

[11] Rietman, E.A., Karp, R.L. and Tuszynski, J.A. (2011) Review and Application of Group Theory to Molecular Systems Biology. *Theoretical Biology and Medical Modelling*, **8**, Article No. 21. https://doi.org/10.1186/1742-4682-8-21

[12] Edelsbrunner, H. (1995) The Union of Balls and Its Dual Shape. *Discrete & Computational Geometry*, **13**, 415-440. https://doi.org/10.1007/BF02574053

[13] Li, J., Mach, P. and Koehl, P. (2013) Measuring the Shapes of Macromolecules—and Why It Matters. *Computational and Structural Biotechnology Journal*, **8**, e201309001.

[14] Vishveshwara, S., Brinda, K.V. and Kannan, N. (2002) Protein Structure: Insights from Graph Theory. *Journal of Theoretical and Computational Chemistry*, **1**, 187-211. https://doi.org/10.1142/S0219633602000117

[15] Penner, R.C., Knudsen, M., Wiuf, C. and Andersen, J.E. (2011) An Algebro-Topological Description of Protein Domain Structure. *PLoS ONE*, **6**, e19670. https://doi.org/10.1371/journal.pone.0019670

[16] MacLane, S. (1998) Categories for the Working Mathematician. 2nd Edition, Springer-Verlag, New York.

[17] Kashiwara, M. and Schapira, P. (2006) Categories and Sheaves. Springer-Verlag, Berlin. https://doi.org/10.1007/3-540-27950-4

[18] McLarty, C. (1992) Elementary Categories, Elementary Toposes (Oxford Logic Guides 21). Oxford University Press Inc., New York.

[19] MacLane, S. and Moerdijk, I. (1992) Sheaves in Geometry and Logic: A First Introduction to Topos Theory. Springer-Verlag, New York.

[20] Morikawa, N. (2019) Design of Self-Assembling Molecules and Boundary Value Problem for Flows on a Space of *n*-Simplices. *Applied Mathematics*, **10**, 907-946. https://doi.org/10.4236/am.2019.1011065