Scientific Research Publishing

Open Journal of Discrete Mathematics, 2023, 13, 63-85
https://www.scirp.org/journal/ojdm
ISSN Online: 2161-7643
ISSN Print: 2161-7635

A Novel Design Method for Protein-Like Molecules from the Perspective of Sheaf Theory

Naoto Morikawa

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

How to cite this paper: Morikawa, N. (2023) A Novel Design Method for Protein-Like Molecules from the Perspective of Sheaf Theory. *Open Journal of Discrete Mathematics*, **13**, 63-85.
https://doi.org/10.4236/ojdm.2023.133007

Received: February 23, 2023
Accepted: June 10, 2023
Published: June 13, 2023

Abstract

Proteins perform a variety of functions in living organisms and their functions are largely determined by their shape. In this paper, we propose a novel mathematical method for designing protein-like molecules of a given shape. In the mathematical model, molecules are represented as loops of n-simplices (2-simplices are triangles and 3-simplices are tetrahedra). We design a new molecule of a given shape by patching together a set of smaller molecules that cover the shape. The covering set of small molecules is defined using a binary relation between sets of molecules. A new molecule is then obtained as a sum of the smaller molecules, where addition of molecules is defined using transformations acting on a set of $(n+1)$-dimensional cones. Due to page limitations, only the two-dimensional case (*i.e.*, loops of triangles) is considered. No prior knowledge of Sheaf Theory, Category Theory, or Protein Science is required. The author hopes that this paper will encourage further collaboration between Mathematics and Protein Science.

Keywords

Discrete Differential Geometry, Protein Design, Sheaf Theory, Protein Structure

1. Introduction

Proteins are folded sequences of amino acids, which perform variety of functions in cells. They perform their functions by interacting with other proteins as well as small molecule ligands (in enzyme-substrate interactions).

In protein-protein interactions, proteins interact each other by forming temporary complexes of proteins called "reaction intermediates". Stability of reaction intermediates then depends on shape complementarity at the protein-protein interfaces (*i.e.*, contact area on surface).

In protein-ligand interactions, proteins bind to one or more small molecule ligands at pockets (or grooves) on their surfaces. Specificity and affinity of the interactions then depend on shape complementarity at the ligand-binding pockets.

In both cases, the functions of proteins are largely determined by their shape. Since structural data for thousands of protein-protein interfaces and ligand-binding pockets are available in the PDB database [1], it is conceivable that artificial proteins could be created by combining these known structures. On the other hand, Mathematics has Sheaf Theory as a framework for patching local data together to obtain global data.

In this paper, we propose a novel design method for artificial protein-like molecules (*i.e.*, folded sequences of basic units) with a given shape. In the method, a new molecule is obtained from a given set of known molecules using the framework of Sheaf Theory. The design of protein-like molecules is carried out in two steps:

1) Specify the shape of a new molecule.

2) Find a folded sequence of basic units that forms the specified shape.

Note that it is not trivial to combine proteins with known structures to form a new protein (*i.e.*, a folded sequence of amino acids). For example, since a local surface structure is often formed by multiple amino acid fragments which are distant in the amino acid sequence, the local surface structure may be unfolded in the new molecule if the corresponding fragments are not arranged adequately in the new amino acid sequence (in other words, proteins are neither "rigid" like holomorphic functions nor "flexible" like continuous functions).

In this paper, protein-like molecules are represented as a closed trajectory in a flow of n-simplices. Due to page limitations, only the two-dimensional case (*i.e.*, flows of 2-simplices) is considered. We then propose a novel design method, called the "incremental design method", which uses the framework of Sheaf Theory to compute a closed trajectory (*i.e.*, a new molecule) from a given set of shorter closed trajectories (*i.e.*, smaller known molecules). We believe this method is essential, especially when designing hybrids of known proteins.

In the past, mathematical studies of protein structure have been concerned mostly with the classification and characterization of their structure [2]-[7]. The author is unaware of any other mathematical studies on the design of protein-like molecules by other researchers. For an overview of protein-like molecules, see [8]. No prior knowledge of Protein Science, Sheaf Theory [9], nor Category Theory [10] is required.

A quick review of Sheaf Theory is given: Let U be a subset of a 2D Euclidean space R^2. Let $A=\{V_1,\cdots,V_n\}$ be a covering of U, *i.e.*, a set of subsets in R^2 such that $U=\bigcup_{i=1}^{n} V_i$. Suppose that each subset V of R^2 is associated with a set $F(V)$ of mathematical data. Let σ be a function $\sigma(V_i)\in F(V_i)$ defined "consistently" on A. In Sheaf Theory, we can compute a value of $F(U)$ by patching together the values $\{\sigma(V_1),\cdots,\sigma(V_n)\}$ on A. For example, in the case of the sheaf of continuous functions on R^2, $F(V)$ is the set of continuous functions defined on an

open set V in R^2. We then obtain a "global" continuous function on an open set U by patching together "local" continuous functions $\sigma(V_i)$ on V_i.

Figure 1 illustrates the design method proposed in this paper. In our case, $F(V)$ is a set of closed trajectories on V. Figure 1(a) is an example of our design method. Given a subset U of R^2 (left end) and a covering $\{V_1, V_2\}$ of U (second from left). Suppose that closed trajectories $\psi_1 \in F(V_1)$ and $\psi_2 \in F(V_2)$ are given (third from left). We then obtain a closed trajectory $\phi \in F(U)$ (right end) by patching together the two closed trajectories $\psi_1 \in F(V_1)$ and $\psi_2 \in F(V_2)$ (enclosed closed trajectories are considered part of the enclosing closed trajectory).

Here's where the problem comes up. In the case of sheaves, we can compute the global data on U by patching together the local data on "any" covering of U (if they are "consistent"). On the other hand, computation fails for some covering A in our case. (Note that $\sigma(V_i)$ can be the empty set because the restriction of an element of $F(U)$ on V_i may not be contained in $F(V_i)$.) Figure 1(b) is an example of unsuccessful computation. Given a subset U of R^2 (left end) and a covering $\{V_1', V_2'\}$ of U (second from left). Suppose that $\psi_1' \in F(V_1)$ and $\psi_2' \in F(V_2)$ are given (third from left), Then, patching together $\psi_1' \in F(V_1)$ and $\psi_2' \in F(V_2)$, we obtain two closed trajectories. In Section 5, we consider sufficient conditions for "local" flows on a covering to produce a single closed trajectory.

This paper is organized as follows. Section 2 explains the loop model of protein-like molecules. Section 3 defines a differential geometric structure on a triangular mesh B. Section 4 formulates the protein design problem from the perspective of Sheaf Theory, where the design problem is rephrased into the "incremental design problem". Section 5 studies the incremental design problem. Due to page limitations, we only consider the case where a covering consists of two smaller molecules. Finally, Section 6 presents discussion and future directions.

2. The Loop Model of Protein-Like Molecules

Shapes of molecules are given as a region on a hexagonal mesh H. Molecules then correspond to a closed trajectory on a triangular mesh B, which is a subdivision of H. New molecules are designed using a differential structure defined on B.

2.1. Regions on a Hexagonal Mesh H

Figure 2 is explained in this subsection. Shown in Figure 2(a) is the honeycomb mesh obtained by dividing a 2D Euclidean plane R^2 into a set of regular hexagons. H denotes the set of all hexagons of the mesh. A subset S of H is called *connected* if each $h_a \in S$ shares a side with another $h_b \in S$ (*i.e.*, for each $h_a \in S$, there exists another $h_b \in S$ such that h_a and h_b share a side).

Shown in Figure 2(b) is a connected subset $S = \{h_1, h_2, \cdots, h_7\}$ of H. Since hexagons of H do not overlap each other, we write

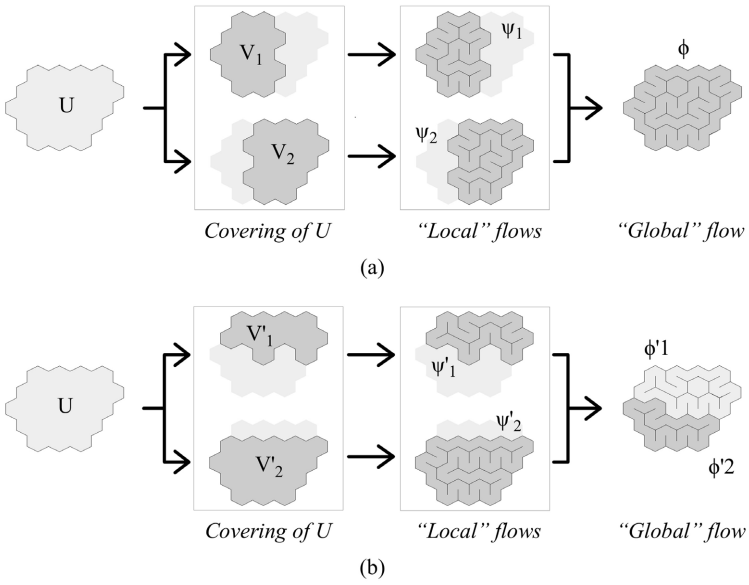

Figure 1. The design method for protein-like molecules proposed in this paper.

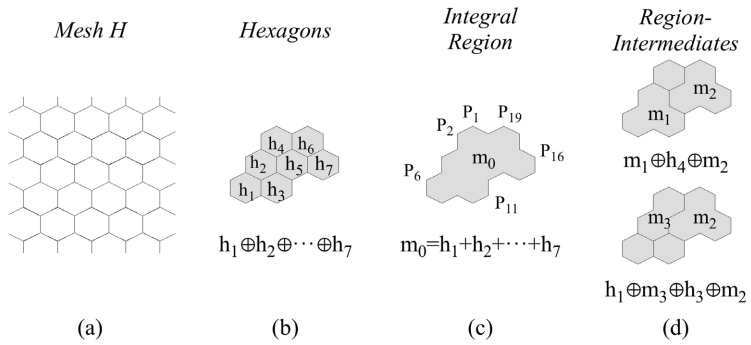

Figure 2. The mathematical model of the shape of protein-like molecules.

$$S = h_1 \oplus h_2 \oplus \cdots \oplus h_7 . \tag{1}$$

If S consists of only one hexagon h_1, we write either $S = h_1$ or $S = \{h_1\}$.

Shown in **Figure 2(c)** is an integral region on H, defined as follows. (Addition "+" of hexagons will be defined later in this subsection.)

Definition 2.1. (Integral Region) An *integral region* m_0 on H is a hole-free subset of R^2 covered by a connected finite subset $S = \{h_1, h_2, \cdots, h_n\}$ of H. The *hexagonal base* of m_0 is then defined by

$$(m_0)_H := h_1 \oplus h_2 \oplus \cdots \oplus h_n . \tag{2}$$

For example, the hexagonal base of m_0 of **Figure 2(c)** is shown in **Figure 2(b)**. I_H denotes the set of all integral regions on H. Note that $H \subset I_H$, *i.e.*, hexagons of H are integral regions.

The *set difference* between two integral regions $m_1, m_2 \in I_H$ is defined by

$$m_1 \setminus m_2 := \bigcup \{h \in H \mid h \subset m_1 \text{ and } h \not\subset m_2\}, \tag{3}$$

where $\bigcup \{h_1, h_2, \cdots, h_n\} := h_1 \cup h_2 \cup \cdots \cup h_n$. The *hexagonal base* $(m_1 \setminus m_2)_H$ of

$m_1 \setminus m_2$ is defined in the same way as for integral regions (it may have holes).

Lemma 2.2. Let $m_1, m_2 \in I_H$. Then, $m_1 \setminus m_2 \in I_H$ if $m_2 \not\subset m_1$.

Shown in Figure 2(d) are region-intermediates on *H*, defined as follows. Let $M_0 = \{m_1, m_2, \cdots\} \subset I_H$. M_0 is called *connected* if each $m_i \in M_0$ shares a side with another $m_j \in M_0$. M_0 is called *disjoint* if m_i's do not overlap each other. If M_0 is disjoint, we write

$$M_0 = m_1 \oplus m_2 \oplus \cdots. \tag{4}$$

If M_0 consists of only one integral region m_1, we write either $M_0 = m_1$ or $M_0 = \{m_1\}$.

Definition 2.3. (Region-Intermediate) A *region-intermediate* M_0 on *H* is a connected finite disjoint subset $S = \{m_1, m_2, \cdots, m_n\}$ of I_H. Since *S* is disjoint, we write

$$M_0 = m_1 \oplus m_2 \oplus \cdots \oplus m_n. \tag{5}$$

The *hexagonal base* of M_0 is then defined by

$$(M_0)_H := (m_1)_H \oplus (m_2)_H \oplus \cdots \oplus (m_n)_H. \tag{6}$$

RI denotes the set of all region-intermediate on *H*.

Finally, addition of integral regions is defined using addition of directed polygonal chains as shown below.

Definition 2.4. (Directed Polygonal Chain) Let $P_1, P_2, \cdots, P_n, P_{n+1}$ be points in R^2. A *directed polygonal chain* $P_1P_2\cdots P_{n+1}$ in R^2 is a set of directed line segments defined by

$$P_1P_2\cdots P_{n+1} := \{P_1P_2, P_2P_3, \cdots, P_nP_{n+1}\}, \tag{7}$$

where P_iP_j denotes the directed line segment from P_i to P_j. If $P_{n+1} = P_1$, we obtain a *closed directed polygonal chain* in R^2. $|P_1P_2\cdots P_nP_1|$ denotes the area of R^2 bounded by $P_1P_2\cdots P_nP_1$ in R^2.

Let $C_0 = \{c_1, c_2, \cdots\}$ be a set of closed directed polygonal chains in R^2. C_0 is called *disjoint* if $|c_i|$'s do not overlap each other. If C_0 is disjoint, we write

$$C_0 = c_1 \oplus c_2 \oplus \cdots. \tag{8}$$

If C_0 is disjoint, the area $|C_0|$ of R^2 bounded by C_0 is defined by

$$|C_0| := |c_1| \oplus |c_2| \oplus \cdots. \tag{9}$$

Let $m_0 \in I_H$. Since m_0 has no hole, we have

$$m_0 = |P_1P_2\cdots P_nP_1| \tag{10}$$

for some $P_1, P_2, \cdots, P_n \in R^2$, where the vertices are labeled counter-clockwise. In the case of Figure 2(c),

$$m_0 = |P_1P_2\cdots P_{20}P_1|. \tag{11}$$

Definition 2.5. (The Boundary Operator ∂ on *RI*) Let $m_0 = |P_1P_2\cdots P_nP_1| \in I_H$, where the vertices are labeled counter-clockwise. The *boundary* ∂m_0 of m_0 is defined by

$$\partial m_0 := P_1P_2\cdots P_nP_1. \tag{12}$$

In this paper, the boundary of an integral region is always given the counter-clockwise orientation. Let $M_0 = m_1 \oplus m_2 \oplus \cdots \oplus m_n \in RI$. Since $\{\partial m_1, \partial m_2, \cdots, \partial m_n\}$ is disjoint, the *boundary* ∂M_0 of M_0 is defined by

$$\partial M_0 := \partial m_1 \oplus \partial m_2 \oplus \cdots \oplus \partial m_n \tag{13}$$

(See Equation (8)) Note that $|\partial M_0| = M_0$.

Addition of integral regions is defined as follows.

Definition 2.6. ($m_1 + m_2$) Let $m_1 = |P_1P_2 \cdots P_nP_1|, m_2 = |Q_1Q_2 \cdots Q_nQ_1| \in I_H$ such that they do not overlap. *Addition* of ∂m_1 and ∂m_2 is defined by

$$\partial m_1 + \partial m_2 := \{P_iP_j \in \partial m_1 \mid P_jP_i \notin \partial m_2\} \cup \{Q_iQ_j \in \partial m_2 \mid Q_jQ_i \notin \partial m_1\}. \tag{14}$$

In other words, the same line segments in opposite directions (*i.e.*, P_iP_j and P_jP_i) are cancelled when added. *Addition* of m_1 and m_2 is then defined by

$$m_1 + m_2 := |\partial m_1 + \partial m_2|. \tag{15}$$

In the case of Figure 2(c), we have

$$m_0 = h_1 + h_2 + \cdots + h_7. \tag{16}$$

Let's denote the set of all natural numbers $\{1,2,3,\cdots\}$ by $\boldsymbol{N}$.

Lemma 2.7.

The set of all integral regions on H is given by

$$I_H = \{h_1 + \cdots + h_n \mid n \in \boldsymbol{N}, \{h_i\} \subset H \text{ is connected and hole-free}\}. \tag{17}$$

The set of all region-intermediate on H is given by

$$RI = \{m_1 \oplus \cdots \oplus m_n \mid n \in \boldsymbol{N}, \{m_i\} \subset I_H \text{ is connected and disjoint}\}. \tag{18}$$

Proof. They follow immediately from the definitions.■

Remark 2.8. Integral domains play the role that "integers" do for rational numbers. That is, "rational" regions are obtained by dividing integral regions into loops of triangles [11].

2.2. Loops on a Triangular Mesh *B*

Figure 3 is explained in this subsection. Shown in Figure 3(a) is the triangular mesh obtained by dividing every hexagon of H into 6 equilateral triangles. B denotes the set of all triangles of the mesh.

Definition 2.9. (Trajectories of Triangles) A *trajectory of triangles* on B is a sequence of triangles of B connected by a common edge. (No direction is assigned to a trajectory.) The edges not used to connect adjacent triangles are called the *normal edges* (of the trajectory) at the triangle (*i.e.*, the "normal vector" of the trajectory). In figures, normal edges are indicated by thick line segments.

Definition 2.10. (Loops of Triangles) A *loop* on B is a closed trajectory of triangles on B. In this paper, protein-like molecules are represented as a loop of triangles of B. $|lp|$ denotes the area in R^2 swept by a loop lp, where the area enclosed by lp is also included in $|lp|$ (for example, ϕ of Figure 1(a)).

A loop lp_0 of length 6 is called a *hexagonal loop*. A hexagonal loop lp_0 is denoted by h_0 if $|lp_0| = h_0 \in H$. In other words, $h_0 \in H$ denotes both a loop of length 6 and a hexagon of H, *i.e.*, $|h_0| = h_0$.

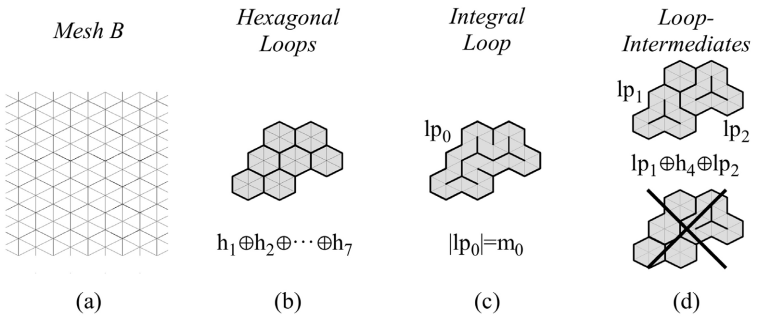

Figure 3. The mathematical model of protein-like molecules.

Remark 2.11. The *hexagonal base* $(M_0)_H$ of $M_0 \in RI$ defined above is a region-intermediate consisting of hexagons as well as a loop-intermediate consisting of loops of length 6.

Shown in **Figure 3(b)** is a set $L=\{h_1,h_2,\cdots,h_7\}$ of seven hexagonal loops on H. Since h_i's do not overlap each other, we write

$$L=h_1\oplus h_2\oplus\cdots\oplus h_7. \tag{19}$$

If L consists of only one hexagon h_1, we write either $L=h_1$ or $L=\{h_1\}$.

Shown in **Figure 3(c)** is an integral loop on B, defined as follows.

Definition 2.12. (Integral Loop) A loop lp_0 on B is called *integral* if $|lp_0|$ is an integral region on H. I_B denotes the set of all integral loops, *i.e.*,

$$I_B=\{lp_0\,|\,lp_0 \text{ is a loop on } B \text{ such that } |lp_0|\in I_H\}. \tag{20}$$

$lp_0\in I_B$ is called an *implementation* of $m_0\in I_H$ if $|lp_0|=m_0$. For example, lp_0 of **Figure 3(c)** is an implementation of m_0 of **Figure 2(c)**.

Let $L=\{lp_1,lp_2,\cdots\}\subset I_B$. The set $|L|$ of integral regions associated with L is defined by

$$|L|:=\{|lp_1|,|lp_2|,\cdots\}\subset I_H. \tag{21}$$

If $|L|$ is disjoint, we write

$$L=lp_1\oplus lp_2\oplus\cdots. \tag{22}$$

If L consists of only one integral loop lp_1, we write either $L=lp_1$ or $L=\{lp_1\}$.

Shown in **Figure 3(d)** top is a loop-intermediate on B, defined as follows.

Definition 2.13. (Loop-Intermediate) A *loop-intermediate* on B is a finite subset $L_0=\{lp_1,lp_2,\cdots,lp_n\}$ of I_B such that $|L_0|$ is a region-intermediate on H. Since $|L_0|$ is disjoint, we write

$$L_0=lp_1\oplus lp_2\oplus\cdots\oplus lp_n. \tag{23}$$

LI denotes the set of all loop-intermediate on B, *i.e.*,

$$LI:=\{lp_1\oplus\cdots\oplus lp_n\,|\,n\in\boldsymbol{N},lp_i\in I_B(i=1,2,\cdots,n),|lp_1\oplus\cdots\oplus lp_n|\in IH\}. \tag{24}$$

Let $M_0\in RI$. $L_0\in LI$ is called an *implementation* of M_0 if $|L_0|=M_0$. Note that some region-intermediates have no implementation. For example, m_3 of **Figure 2(d)** has no implementation (**Figure 3(d)** bottom).

Now, fusion and fission of integral loops are defined using addition of the corresponding integral regions (Addition of integral loops is considered in Section 4 below).

Definition 2.14. (Fusion and Fission of Integral Loops) Let $lp_0 \in I_B$. Let $L_0 = lp_1 \oplus lp_2 \oplus \cdots \oplus lp_n \in LI$. lp_0 is called the *fusion* of L_0 if

$$|lp_0| = |lp_1| + |lp_2| + \cdots + |lp_n|. \tag{25}$$

L_0 is then called a *fission* of lp_0. In **Figure 3**, both $h_1 \oplus h_2 \oplus \cdots \oplus h_7$ and $lp_1 \oplus h_4 \oplus p_2$ are fissions of lp_0.

Finally, let's define flows of triangles on *B*.

Definition 2.15. (Flows of Triangles on *B*) A *flow ψ of triangles* on *B* is an assignment of normal edges to triangles of *B*, *i.e.*,

$$\psi(t) := \text{the set of normal edges of } t \ (t \in B). \tag{26}$$

ψ is called *regular* at *t* if $\psi(t)$ consists of one edge (*i.e.*, *t* is connected to exactly two adjacent triangles). ψ is called *regular* if ψ is regular at all triangles of *B*. FLW_R denotes the set of all regular flows on *B*.

A triangle *t* of *B* is called *singular* if it is not regular. Singular triangles are called *branch*, *terminal*, or *isolated* triangles when they have no normal edges (*i.e.*, connected to all the adjacent triangles), two normal edges (*i.e.*, connected to only one adjacent triangle), or three normal edges (*i.e.*, connected to no adjacent triangles), respectively.

Remark 2.16. We often consider trajectories of triangles without explicit reference to the corresponding flow ψ. A triangle *t* is then called regular if ψ is regular at *t*.

Definition 2.17. (Disjoint Unions $L(\psi)$ and $M(\psi)$) Let $\psi \in FLW_R$. $L(\psi)$ denotes the set of all the loops of ψ. Since trajectories of ψ do not overlap, we have

$$L(\psi) := lp_1 \oplus lp_2 \oplus \cdots, \tag{27}$$

where lp_i's are the loops of ψ. $L(\psi)$ is called the *loops associated with* ψ. $M(\psi)$ denotes the associated regions, *i.e.*,

$$M(\psi) := |lp_1| \oplus |lp_2| \oplus \cdots, \tag{28}$$

$M(\psi)$ is called the *regions associated with* ψ.

Lemma 2.18. Let $\psi \in FLW_R$.

1) $L(\psi) \in LI$ if $L(\psi)$ is finite, connected, and hole-free.
2) $M(\psi) \in RI$ if $M(\psi)$ is finite and connected.

2.3. Design Problem for Protein-Like Molecules

In the loop model of protein-like molecules, the shape of a new molecule is an integral region m_0. A new molecule of the shape m_0 then is an implementation lp_0 of m_0. The problem we consider in this paper is now defined as follows.

Problem 2.19. (Design of Protein-like Molecules) Given $M_0 = \{m_0\} \in RI$,

find $L_0 = \{lp_0\} \in LI$ such that $|lp_0| = m_0$.

In the next section, the problem is rephrased using a differential geometric structure on B.

3. Differential Geometric Structure on B

A differential geometric structure on B is naturally obtained by embedding the honeycomb mesh H in a 3D Euclidean space R as shown in this section. We denote the set of all real numbers by R.

3.1. Embedding of H in R^3

Shown in Figure 4(a) is a unit cube in R^3 and its orthogonal projection on the plane H_0 in R^3 defined by

$$H_0 := \{(x,y,z) \in R^3 \mid x+y+z=0\}. \tag{29}$$

H is embedded in H_0 using unit cubes in R^3, as explained below.

Definition 3.1. (Unit Cubes in R^3) Let $(a,b,c) \in R^3$. $[a,b,c]$ denotes the unit cube at (a,b,c), *i.e.*,

$$[a,b,c] := [a,a+1] \times [b,b+1] \times [c,c+1] \subset R^3, \tag{30}$$

where $[x,y]$ is the closed interval in R between x and y. If $P=(a,b,c) \in R^3$, then $[a,b,c]$ is also written as $[P]$. UC denotes the set of all unit cubes at the integer lattice Z^3, *i.e.*,

$$UC := \{[a,b,c] \mid (a,b,c) \in Z^3\}. \tag{31}$$

The *height* $ht_{UC}([a,b,c])$ of $[a,b,c] \in UC$ is defined by

$$ht_{UC}([a,b,c]) := a+b+c. \tag{32}$$

Remark 3.2. $[a,b,c]$ is given by

$$\begin{aligned}[a,b,c] = \{&(a,b,c)+(u,0,0)+(0,v,0)+(0,0,w) \in R^3 \mid (u,v,w) \in R^3 \\ &\text{such that } u+v+w=1 \text{ and } u,v,w \geq 0\}.\end{aligned} \tag{33}$$

Shown in Figure 4(a) top is a unit cube $[a,b,c] \in UC$ with vertices $O=(a,b,c)$, $P=(a+1,b,c)$, $Q=(a,b+1,c)$, $R=(a,b,c+1)$, $U=(a+1,b+1,c)$, $V=(a,b+1,c+1)$, $W=(a+1,b,c+1)$, and $X=(a+1,b+1,c+1)$. The vertical diagonals OU, OV, and OW are drawn as thick line segments.

Definition 3.3. (Projection π of R^3 onto H_0) π is the orthogonal projection of R^3 onto H_0 defined by

$$\pi(x,y,z) := ((2x-y-z)/3, (-x+2y-z)/3, (-x-y+2z)/3). \tag{34}$$

Shown in Figure 4(a) bottom is the orthogonal projection of OU, OV, and OW onto H_0, forming part of a hexagonal mesh on H_0.

Definition 3.4. ("Bumpy" Mesh H_{bump}) H_{bump} is the "bumpy" honeycomb mesh defined on the top surfaces of

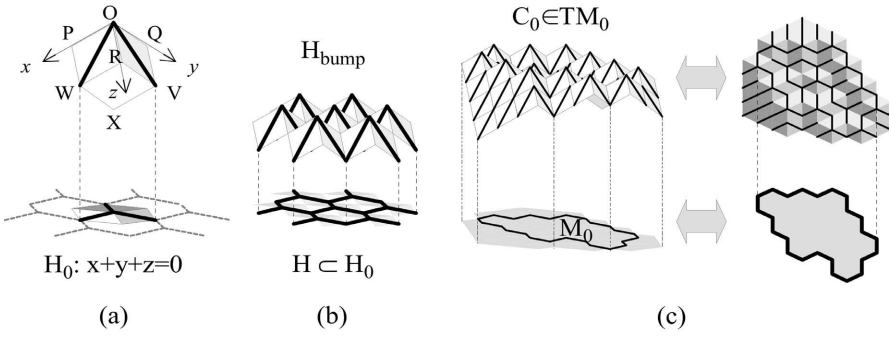

Figure 4. Differential geometric structure on the trianglular mesh *B*.

$$UC_0 := \left\{ [x,y,x] \in UC \mid ht_{UC}([x,y,z]) = 0 \right\}. \tag{35}$$

The edges of the mesh are the vertical diagonals. H_{bump} denotes the set of "bumpy" hexagons drawn on UC_0 (Figure 4(b) top).

Shown in Figure 4(b) bottom is the projection of H_{bump} onto H_0 by π. In the following, we identify H with $\pi(H_{bump}) \subset H_0$. An embedding of B in H_0 is then obtained by dividing every hexagon of $\pi(H_{bump})$ into 6 equilateral triangles.

3.2. Tangent Cones

Shown in Figure 4(c) top-left is a tangent cone C_0 to a region-intermediate M_0, defined as follows. Roughly speaking, a 3D cone with multiple tops is obtained by stacking unit cubes diagonally (from (∞,∞,∞) to $(-\infty,-\infty,-\infty)$).

Definition 3.5. (Tangent Cone *Cone A*) Let $A \subset Z^3$. The *tangent cone Cone A* generated by A is defined by

$$Cone\ A := \left\{ (x,y,z) \in R^3 \mid \max_{(a,b,c)\in A} \left\{ \min\{x-a, y-b, z-c\} \right\} \geq 0 \right\}. \tag{36}$$

TCONE denotes the set of all tangent cones, *i.e.*,

$$TCONE := \left\{ Cone\ A \mid A \subset Z^3 \right\}. \tag{37}$$

$P(TCONE)$ denotes the set of all subsets of *TCONE*, *i.e.*, the power set of *TCONE*.

Definition 3.6. (Tops and Bottoms of *Cone A*) Let $C_0 \in TCONE$. The *top vertices* of C_0 are the peaks of the cone. $top(C_0)$ denotes the set of all top vertices of C_0. Note that $C_0 = Cone\ top(C_0)$. The *bottom vertices* of C_0 are the dents of the cone which are peaks if we look up the cone from (∞,∞,∞). $bottom(C_0)$ denotes the set of all bottom vertices of C_0.

We define flows of triangles on B using tangent cones in R^3.

Definition 3.7. (The Flow ψ_{C_0} on *B*) Let $C_0 \in TCONE$. Note that the surfaces of C_0 consist of the top faces of unit cubes of *UC*. Taking their vertical diagonals as normal edge, we obtain a regular flow of triangles on the surface of C_0 (Figure 4(c) top-right). Projecting the regular flow onto H_0 by π, we obtain a regular flow on B. The regular flow is called the *flow on B induced by* C_0

and denoted by ψ_{C_0}.

Definition 3.8. FLW_{TCONE} denotes the set of all regular flows on B induced by tangent cones, *i.e.*,

$$FLW_{TCONE} := \{\psi_C \mid C \in TCONE\} \subset FLW_R. \tag{38}$$

RI_{TCONE} and LI_{TCONE} denote the corresponding set of region-intermediates and loop-intermediates, respectively, *i.e.*,

$$RI_{TCONE} := \{M \in RI \mid \exists C \in TCONE \text{ such that } M = M(\psi_C)\}. \tag{39}$$

$$LI_{TCONE} := \{L \in LI \mid \exists C \in TCONE \text{ such that } L = L(\psi_C)\}. \tag{40}$$

The author has no proof of the following claim.

Claim 3.9. $FLW_{TCONE} = FLW_R$.

In this paper, only flows of FLW_{TCONE} are considered.

Remark 3.10. $RI_{TCONE} \neq RI$. For example, m_3 of **Figure 2(d)** is not contained in RI_{TCONE}.

3.3. Tangent Cones to M_0

Here we define a tangent "space" to $M_0 \in RI$.

Definition 3.11. (The Boundary Cone $BC(\partial M_0)$) Let

$$M_0 = m_1 \oplus m_2 \oplus \cdots \oplus m_n \in RI_{TCONE}, \tag{41}$$

where

$$m_i = \left|P_{i1}P_{i2}\cdots P_{ik_i}P_{i1}\right| \ (i = 1,2,\cdots,n). \tag{42}$$

The *boundary cone* $BC(\partial M_0)$ *to* ∂M_0 is defined by

$$BC(\partial M_0) := Cone\{Q_{i1}, Q_{i2}, \cdots, Q_{ik_i} \mid i = 1,2,\cdots,n\}, \tag{43}$$

where Q_{ij}'s are points on the top surfaces of UC_0 such that

$$\pi(Q_{ij}) = P_{ij}. \tag{44}$$

Remark 3.12. Since π is a one-to-one mapping between the top surfaces of UC_0 and H_0, the boundary cone $BC(\partial M_0)$ exists for all $M_0 \in RI_{TCONE}$.

Definition 3.13. (The Set TM_0 of Tangent Cones) Let $M_0 \in RI_{TCONE}$. The set TM_0 of tangent cones to M_0 is defined by

$$TM_0 := \{C \in TCONE \mid top(BC(\partial M_0)) \cup bottom(BC(\partial M_0)) \subset sur(C)\}, \tag{45}$$

where $sur(C)$ is the surfaces of C, *i.e.*,

$$sur(C) := \{(x,y,z) \in R^3 \mid \max_{(a,b,c)\in top(C)} \{\min\{x-a, y-b, z-c\}\} = 0\}. \tag{46}$$

Note that $BC(\partial M_0) \in TM_0$.

Remark 3.14. $C \in TM_0$ dose not imply $top(BC(\partial M_0)) \subset top(C)$.

Lemma 3.15. (The Base Tangent Cone $C_{base}(M_0)$) Let $M_0 \in RI_{TCONE}$. There exists $C_{base}(M_0) \in TM_0$ such that

$$L(\psi_{C_{base}(M_0)}) = (M_0)_H, \tag{47}$$

$$ht_{UC}\left(\left[a,b,c\right]\right)=0 \quad \text{for} \quad (a,b,c)\in top\left(C_{base}\left(M_0\right)\right). \tag{48}$$

$C_{base}\left(M_0\right)$ is called the *base tangent cone* associated with M_0.

Proof. Since π gives a one-to-one mapping between the top surfaces of UC_0 and H_0, the result follows immediately.■

Definition 3.16. (Mapping *T*) Assigning TM_0 to each $M_0\in RI_{TCONE}$, we obtain a mapping T from RI_{TCONE} to $P(TCONE)$. Let $S\subset RI_{TCONE}$. A *section* σ of T on S is a mapping from S to $P(TCONE)$ such that $M\left(\psi_{\sigma(M_i)}\right)=M_i$ for all $M_i\in S$. $\Gamma_T\left(S\right)$ denotes the set of all sections of T on S.

The design problem is now rephrased as follows.

Problem 3.17. (Design of Protein-like Molecules) Given $M_0=\left\{m_0\right\}\in RI_{TCONE}$, find $C_0\in TM_0$ such that $M\left(\psi_{C_0}\right)=M_0$.

4. Loop Design Problem from the Perspective of Sheaf Theory

To mimic Sheaf Theory, "subsets" of a region-intermediate M_0 are defined using a binary relation over RI_{TCONE}. A "covering" $S=\left\{M_1,M_2,\cdots,M_n\right\}$ of M_0 is then defined as a set of region-intermediates such that M_0 is the least upper bound of S with respect to the binary relation. An implementation of M_0 is obtained as the sum of implementations of $M_i\left(i=1,2,\cdots,n\right)$, where addition is defined using transformations on TM_0 as shown below.

4.1. Binary Relation over RI_{TCONE} and LI_{TCONE}

Shown in Figure 5(a) is a binary relation over RI_{TCONE}, defined as follows.

Definition 4.1. (Binary Relation ≤ over RI_{TCONE}) Let $M_a,M_b\in RI_{TCONE}$. Then, $M_a\le M_b$ if and only if, for any $m_b\in M_b$,

$$\text{there exist } m_1,\cdots,m_n\in M_a \text{ such that } m_b=m_1+\cdots+m_n. \tag{49}$$

In figures, we often use the arrow $M_a\to M_b$ to indicate $M_a\le M_b$.

Shown in Figure 5(b) is the binary relation over LI_{TCONE} induced by the binary relation ≤ over RI_{TCONE}. That is,

Definition 4.2. (Binary Relation ≤ over LI_{TCONE}) Let $L_a,L_b\in LI_{TCONE}$. Then, $L_a\le L_b$ if and only if, for any $lp_b\in L_b$,

$$\text{there exists } lp_1,\cdots,lp_n\in L_a \text{ such that } \left|lp_b\right|=\left|lp_1\right|+\cdots+\left|lp_n\right|. \tag{50}$$

In figures, we often use the arrow $L_a\to L_b$ to indicate $L_a\le L_b$.

Remark 4.3. Notations such as $\left(RI_{TCONE},\le\right)$ and $\left(LI_{TCONE},\le\right)$ are used to explicitly indicate the binary relation equipped with a set.

Lemma 4.4. Let $M_1,M_2\in RI_{TCONE}$. Then,

$$T\left(M_1\right)\subset T\left(M_2\right) \text{ if } M_1\le M_2. \tag{51}$$

That is, T is a "covariant" mapping from $\left(RI_{TCONE},\le\right)$ to $\left(P\left(TCONE\right),\subset\right)$.

Shown in Figure 5(c) is examples of the greatest lower bound of loop-intermediates, defined as follows.

Definition 4.5. ($\wedge S$ and $\vee S$) Let $S\subset RI_{TCONE}$. The *greatest lower bound* $\wedge S$ of S is the greatest element of RI_{TCONE} that is less than or equal to each element

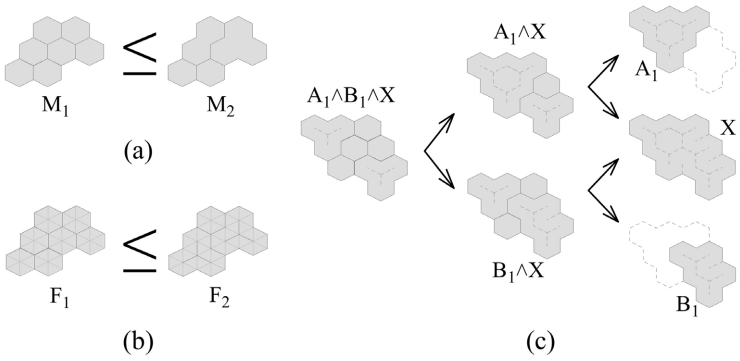

Figure 5. Binary relation ≤ over RI_{TCONE} and LI_{TCONE}.

of S. The *least upper bound* $\bigvee S$ of S is the least element of RI that is greater than or equal to each element of S. $\bigwedge S$ and $\bigvee S$ for $S \subset LI_{TCONE}$ are also defined similarly.

Remark 4.6. In general, there are multiple candidates for $\bigwedge S$ and $\bigvee S$. In such cases, select one of them arbitrarily. Because of this uncertainty, "$M_0 \le M_i$ for all $M_i \in S$" does not imply $M_0 \le \bigwedge S$.

Remark 4.7. $H \le M_0 \le \varnothing$ for any $M_0 \in RI_{TCONE}$, where $\varnothing$ denotes the empty set.

We use the following lemma to find "subsets" of a region-intermediate.

Lemma 4.8. Let $M_0 \in RI_{TCONE}$ and $C \in TCONE$. Then,

$$M(\psi_C) \le M_0 \text{ If and only if } C \in TM_0. \tag{52}$$

Proof. $M(\psi_C) \le M_0$ if and only if $\partial M_0 \subset \partial M(\psi_C)$. $\partial M_0 \subset \partial M(\psi_C)$ if and only if $top(BC(\partial M_0)) \cup bottom(BC(\partial M_0)) \subset sur(C)$. The result follows immediately. ■

4.2. Coverings of a Region-Intermediate

Two types of coverings are defined as follows.

Definition 4.9. (Coverings of a Region-Intermediate) Let $M_0 \in RI_{TCONE}$. Let $S \subset RI_{TCONE}$. S is called a *covering* of M_0 if $\bigvee S = M_0$.

Definition 4.10. (Topological Coverings of an Integral Region) Let $m_0 \in I_H$. Let $V = \{c_1, c_2, \cdots, c_n\} \subset I_H$. V is called a *topological covering* of m_0 if 1) $m_0 = \bigcup_{i=1}^{n} c_i$, and 2) for each $c_i \in V$, there exists another $c_j \in V$ such that $c_i \bigcap c_j \neq \varnothing$.

Remark 4.11. Since some integral regions have no implementation (*i.e.*, there exists $c_i \in I_H$ such that $c_i \neq |lp|$ for any $lp \in I_B$), topological coverings may have no sections on them.

Lemma 4.12. Let $m_0 \in I_H$. Let $\{c_1, c_2, \cdots, c_n\} \subset I_H$ be a topological covering of m_0. A covering of m_0 is then obtained by

$$\{m_0 \wedge c_1, m_0 \wedge c_2, \cdots, m_0 \wedge c_n\}. \tag{53}$$

Example 4.13. In Figure 5(c), $\{A_1, B_1\}$ is a topological covering of X. $\{A_1 \wedge X, B_1 \wedge X\}$ is a covering of X.

The proposed design method uses a specific type of covering (in **Problem 4.38**).

Definition 4.14. (Hexagonal Covering S_V of an Integral Region) Let $V=\{c_1,c_2,\cdots,c_n\}\subset I_H$ be a topological covering of $m_0\in I_H$. The *hexagonal covering* S_V of m_0 associated with V is defined by

$$S_V=\left\{c_1\oplus\left(m_0\setminus c_1\right)_H,c_2\oplus\left(m_0\setminus c_2\right)_H,\cdots,c_n\oplus\left(m_0\setminus c_n\right)_H\right\}. \tag{54}$$

Lemma 4.15. S_V is a covering of m_0.

The design problem is now rephrased as follows.

Problem4.16. (Incremental Design of Protein-like Molecules) Given 1) a target shape m_0: $M_0=\{m_0\}\in RI_{TCONE}$, 2) a topological covering V of m_0: $V=\{c_1,c_2,\cdots,c_n\}$, 3) a section σ of T on V: $\sigma\in\Gamma_T(V)$. Then, compute $C_0\in TM_0$ such that $M\left(\psi_{C_0}\right)=M_0$ by patching "local" loop-intermediates $L\left(\psi_{\sigma(c_1)}\right)$, $L\left(\psi_{\sigma(c_2)}\right)$, $\cdots$, and $L\left(\psi_{\sigma(c_n)}\right)$ together.

4.3. Transformations on LI_{TCONE} Induced by UC

To patch loop-intermediates together, we define addition of loop-intermediates using transformations on *TCONE*, defined as follows.

Definition 4.17. (*TRANS* (*TCONE*)) A *transformation* on *TCONE* is a mapping from *TCONE* to *TCONE*. *TRANS* (*TCONE*) denotes the set of all transformations on *TCONE*.

Let $A\in TRANS(TCONE)$ and $C\in TCONE$. We use the symbol "$\cdot$" to denote the transformation of C by A, *i.e.*, $A\cdot C$. $A\cdot C$ is also called the *action* of A on C. Let $A_1,A_2,\cdots,A_n\in TRANS(TCONE)$. We use the symbol "$\circ$" to denote the composition of transformations, *i.e.*,

$$A_1\circ A_2\circ\cdots\circ A_n\cdot C:=A_1\cdot\left(A_2\cdot\left(\cdots\left(A_n\cdot C\right)\right)\cdots\right). \tag{55}$$

Example 4.18. Unit cubes induce transformations on *TCONE* as follows. Let $C\in TCONE$, where $top(C)=\{P_1,P_2,\cdots,P_n\}$, *i.e.*,

$$C=Cone\{P_1,P_2,\cdots,P_n\}. \tag{56}$$

Taking the unit cube $[a,b,c]$ at $P_1=(a,b,c)$ from C, we obtain another tangent cone

$$C'=Cone\{P_1',P_2',P_3',P_2,\cdots,P_n\}, \tag{57}$$

where $P_1'=(a+1,b,c)$, $P_2'=(a,b+1,c)$, and $P_3'=(a,b,c+1)$. Conversely, putting the unit cube $[a,b,c]$ on C', we obtain the original cone C.

Definition 4.19. (The minimal *L*-cone C_L) Let $L\in LI_{TCONE}$ and $C\in T|L|$. C is called a L-cone if $L(\psi_C)=L$. The tangent cone C_L is the minimal L-cone with respect to set inclusion, *i.e.*, $C_L\subset C$ for any L-cone C. Since $L\in LI_{TCONE}$, $C_L\in T|L|$ always exists and uniquely determined by L.

Transformations on *TCONE* induce transformations on LI_{TCONE} as follows.

Definition 4.20. (Transformations on LI_{TCONE}) Let $A\in TRANS(TCONE)$ and $L\in LI_{TCONE}$. The *transformation of L by A* is defined by

$$A \cdot L := L\left(\psi_{A \cdot C_L}\right). \tag{58}$$

$A \cdot L$ is called the *action* of A on L.

Definition 4.21. (Transformations $P[a,b,c]$, $T[a,b,c]$, and $PT[a,b,c]$) Let $[a,b,c] \in UC$. Two transformations $P[a,b,c]$ and $T[a,b,c]$ on *TCONE* induced by $[a,b,c]$ is defined by

$$P[a,b,c] \cdot C := C \cup Cone\{(a,b,c)\}, \tag{59}$$

$$T[a,b,c] \cdot C := \left\{[x,y,z] \in C \mid [a,b,c] \notin Cone\{(x,y,z)\}\right\}, \tag{60}$$

where $C \in TCONE$. $P[a,b,c] \cdot C$ is called the *put & fill-action* by $[a,b,c]$ on C. $T[a,b,c] \cdot C$ is called the *take & clear-action* by $[a,b,c]$ on C. We denote the composition of P after T by PT, *i.e.*,

$$PT[a,b,c] := P[a,b,c] \circ T[a,b,c] \ \left([a,b,c] \in UC\right). \tag{61}$$

$PT[a,b,c] \cdot C$ is called the *take & put-action* by $[a,b,c]$ on C.

Remark 4.22. After the action of $PT[a,b,c]$ on $C \in TCONE$, the cube $[a,b,c]$ is always visible from $(-\infty,-\infty,-\infty)$. On the other hand, after the action of $P[a,b,c]$ on C, $[a,b,c]$ may not be visible from $(-\infty,-\infty,-\infty)$.

Lemma 4.23. Let $[a,b,c] \in UC$ and $C \in TCONE$. Then,

$$P[a,b,c] \cdot C = C \quad \text{if} \quad [a,b,c] \in C, \tag{62}$$

$$T[a,b,c] \cdot C = C \quad \text{if} \quad [a,b,c] \notin C. \tag{63}$$

Definition 4.24. ($TRANS_{\langle T,PT\rangle}(TCONE)$) $TRANS_{\langle T,PT\rangle}(TCONE)$ denotes the set of all the transformations on *TCONE* generated by finite compositions of $T[a,b,c]$ and $PT[a,b,c]$ ($[a,b,c] \in UC$), *i.e.*,

$$TRANS_{\langle T,PT\rangle}(TCONE) := \left\{A_1 \circ \cdots \circ A_n \mid n \in \boldsymbol{Z}, A_i = Pu \text{ or } PTu \text{ for some } u \in UC\right\}. \tag{64}$$

In general, $G_1 \circ G_2 \cdot C \neq G_2 \circ G_1 \cdot C$ for $G_1, G_2 \in TRANS_{\langle T,PT\rangle}(TCONE)$ and $C \in TCONE$

Example 4.25. Let $u_1 = [a,b,c], u_2 = [a',b',c'] \in UC$ such that $(a,b,c) \in Cone\{a',b',c'\}$. Then,

$$PTu_1 \circ PTu_2 \cdot Cone\{(a',b',c')\} \neq PTu_2 \circ PTu_1 \cdot Cone\{(a',b',c')\}, \tag{65}$$

$$Tu_1 \circ PTu_2 \cdot Cone\{(a',b',c')\} \neq PTu_2 \circ Tu_1 \cdot Cone\{(a',b',c')\}. \tag{66}$$

Definition 4.26 (Well-defined Transformations) Let $G = A_1 \circ A_2 \circ \cdots \circ A_n \in TRANS_{\langle T,PT\rangle}(TCONE)$. G is called *well-defined* if the action of G on *TCONE* does not depend on the order of A_i's, *i.e.*,

$$G \cdot C = A_{\rho(1)} \circ A_{\rho(2)} \circ \cdots \circ A_{\rho(n)} \cdot C \quad \text{for all} \quad C \in TCONE. \tag{67}$$

for any permutation ρ of $\{1,2,\cdots,n\}$.

Remark 4.27. If G is well-defined, removed unit cubes are removed forever and placed unit cubes are placed forever.

4.4. Addition on LI_{TCONE}

Addition of loop-intermediates is now defined using transformations on *TCONE*.

Definition 4.28. {Transformations on TM_0} Let $M_0 \in RI$. The set $TRANS(TM_0)$ of transformations on TM_0 is defined by

$$TRANS(TM_0) := \left\{ G \in TRANS_{\langle T,PT \rangle}(TCONE) \mid G \cdot C \in TM_0 \text{ for all } C \in TM_0 \right\}. \quad (68)$$

Lemma 4.29. Let $M_0 \in RI_{TCONE}$ and $G_1, G_2 \in TRANS(TM_0)$. Then,

$$G_1 \circ G_2 \in TRANS(TM_0). \quad (69)$$

Lemma 4.30. Let $L_0, L \in LI_{TCONE}$ and $G \in TRANS(T|L_0|)$. Then,

$$\text{If } L \le L_0, \text{ then } G \cdot L \le L_0. \quad (70)$$

Definition 4.31. (The Relative Transformation $G_H(L)$) Let $L \in LI_{TCONE}$. The *relative transformation* $G_H(L)$ of L with respect to H is defined by

$$G_H(L) := PT[P_1] \circ \cdots \circ PT[P_n] \circ T[Q_1 - (1,1,1)] \circ \cdots \circ T[Q_k - (1,1,1)], \quad (71)$$

where

$$\{P_1, \cdots, P_n\} = \{P \in top(C_L) \mid ht_{UC}([P]) \neq 0\}, \quad (72)$$

$$\{Q_1, \cdots, Q_k\} = \{Q \in bottom(C_L) \mid ht_{UC}([Q]) \neq 2\}. \quad (73)$$

$[Q-(1,1,1)]$ is defined by $[Q-(1,1,1)] := (a-1, b-1, c-1)$ for $Q = (a,b,c)$.

Lemma 4.32. Let $L \in LI_{TCONE}$ and $M \in RI_{TCONE}$ such that $|L| \le M$. Then, $G_H(L)$ is well-defined and

$$L = G_H(L) \cdot L_H \le G_H(L) \cdot M_H. \quad (74)$$

Remark 4.33. The hexagonal base M_H is a loop-intermediate consisting of loops of length 6 as well as a region-intermediate consisting of hexagons.

Lemma 4.34. Let $L \in LI_{TCONE}$ and $M \in RI_{TCONE}$ such that $|L| \le M$. Then,

$$G_H(L) \in TRANS(TM). \quad (75)$$

Addition of loop-intermediates is now defined as follows.

Definition 4.35. (Addition of Loop-Intermediates) Let $L_1, L_2, \cdots, L_n \in LI_{TCONE}$ and $M \in RI_{TCONE}$ such that $|L_1|, |L_2|, \cdots, |L_n| \le M$. Then,

$$L_1 + L_2 + \cdots + L_n := G_H(L_1) \circ G_H(L_2) \circ \cdots \circ G_H(L_n) \cdot M_H. \quad (76)$$

Remark 4.36. Addition $L_1 + L_2 + \cdots + L_n$ is defined with respect to M_H, which is not explicitly indicated in the formula.

Definition 4.37. (Section σ_{S_V} on S_V) Let $V = \{c_1, c_2, \cdots, c_n\} \subset I_H$ be a topological covering of $m_0 \in I_H$. Let $\sigma \in \Gamma_T(V)$. The section σ_{S_V} of T on the hexagonal covering S_V is defined by

$$\sigma_{S_V}(c_i \oplus (m_0 \setminus c_i)_H) := \sigma(c_i) \cup C_{base}((m_0 \setminus c_i)_H). \quad (77)$$

Note that

$$L\left(\psi_{\sigma_{S_V}(c_i \oplus (m_0 \setminus c_i)_H)}\right) = L\left(\psi_{\sigma(c_i)}\right) \oplus (m_0 \setminus c_i)_H \ (i = 1, \cdots, n). \quad (78)$$

Since $c_i \oplus (m_0 \setminus c_i)_H \le m_0 (i=1,\cdots,n)$, we can define addition

$$\sum_{i=1}^{n} L\left(\psi_{\sigma_{S_V}(c_i \oplus (m_0 \setminus c_i)_H)}\right) \tag{79}$$

by **Definition 4.35.**

Using addition of loop-intermediates, the design problem is now rephrased as follows.

Problem 4.38. (Incremental Design of Protein-like Molecules) Given 1) a target shape m_0: $M_0 = \{m_0\} \in RI_{TCONE}$; 2) a topological covering V of m_0: $V = \{c_1, c_2, \cdots, c_n\}$; 3) a section σ of T on V: $\sigma \in \Gamma_T(V)$. Then, we obtain $L_0 \in LI_{TCONE}$ such that $|L_0| \le M_0$ by

$$\begin{aligned} L_0 &:= \sum_{i=1}^{n} L\left(\psi_{\sigma_{S_V}(c_i \oplus (m_0 \setminus c_i)_H)}\right) \\ &= G_H(L_1) \circ G_H(L_2) \circ \cdots \circ G_H(L_n) \cdot (M_0)_H, \end{aligned} \tag{80}$$

where $L_i := L\left(\psi_{\sigma_{S_V}(c_i \oplus (m_0 \setminus c_i)_H)}\right)$. The question here is "when dose L_0 consist of a single loop?"

5. Incremental Design of Protein-Like Molecules ($N = 2$)

In general, the sum of loops is not a loop. In this section, we consider sufficient conditions for the L_0 of **Problem 4.38** to be a loop. Due to page limitations, we only consider a topological covering consisting of two integral regions. The incremental design problem is then given as follows

Problem 5.1. (Incremental Design of Protein-like Molecules ($n = 2$)) Given 1) a target shape m_0: $M_0 = \{m_0\} \in RI_{TCONE}$; 2) a topological covering V of m_0: $V = \{c_1, c_2\}$; 3) a section σ of T on V: $\sigma \in \Gamma_T(V)$. Then, we obtain $L_0 \in LI_{TCONE}$ such that $|L_0| \le M_0$ by

$$\begin{aligned} L_0 &:= L\left(\psi_{\sigma_{S_V}(c_1 \oplus (m_0 \setminus c_1)_H)}\right) + L\left(\psi_{\sigma_{S_V}(c_2 \oplus (m_0 \setminus c_2)_H)}\right) \\ &= G_H(L_1) \circ G_H(L_2) \cdot (M_0)_H, \end{aligned} \tag{81}$$

where $L_i := L\left(\psi_{\sigma_{S_V}(c_i \oplus (m_0 \setminus c_i)_H)}\right)$. Find sufficient conditions for L_0 to be a loop.

5.1. Closer Look at the Action of $TRANS_{\langle T,PT \rangle}(TCONE)$

Figure 6 shows the effect of the action of $TRANS(TM_0)$ on $L\left(\psi_{C_{base}(M_0)}\right)$ using the height of normal edges, defined as follows.

Definition 5.2. (Height of Normal Edges) Let P_1P_2 be a vertical diagonal of a top face of $[a,b,c] \in UC$. The *height* $ht_E(P_1P_2)$ of P_1P_2 is defined by

$$ht_E(P_1P_2) := ht_{UC}([a,b,c]). \tag{82}$$

Let $C \in TCONE$. Let Q_1Q_2 be a normal edge of the induced flow ψ_C such that $\pi(P_1P_2) = Q_1Q_2$, where P_1P_2 is the corresponding vertical diagonal on the surfaces of C. The *height* $ht_E(Q_1Q_2)$ of Q_1Q_2 (with respect to C) is defined by

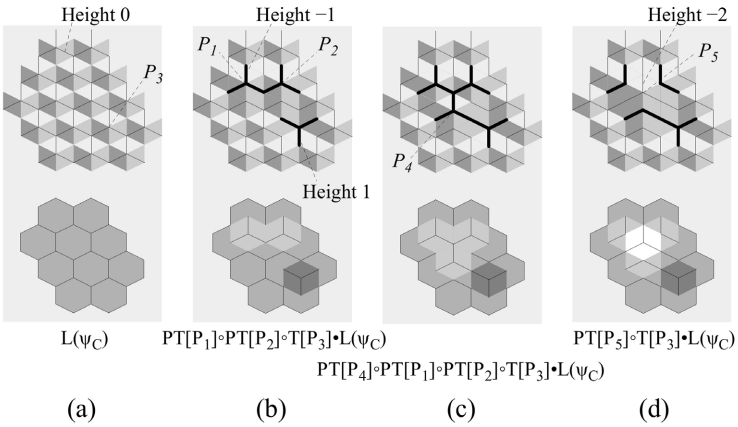

Figure 6. The action of $TRANS(TM_0)$ on $L\left(\psi_{C_{base}(M_0)}\right)$.

$$ht_E(Q_1Q_2) := ht_E(P_1P_2). \tag{83}$$

Shown in **Figure 6(a)** top is the top view of the base tangent cone

$$C_{base}(M_0) \in TM_0 \tag{84}$$

of some $M_0 \in RI$. Shown in **Figure 6(a)** bottom is the loop-intermediate

$$L\left(\psi_{C_{base}(M_0)}\right) \tag{85}$$

on B. By definition, the heights of all normal edges of $\psi_{C_{base}(M_0)}$ are 0.

Shown in **Figure 6(b)** top is the tangent cone

$$PT[P_1]\circ PT[P_2]\circ T[P_3]\cdot C_{base}(M_0). \tag{86}$$

In the upper part, two Y-shaped sets of normal edges are replaced by two inverted Y-shaped sets of normal edges (thick line segments) by putting the two unit cubes $[P_1]$ and $[P_2]$. In the lower part, a Y-shaped set of normal edges is replaced byan inverted Y-shaped set of normal edges (thick line segments) by taking the unit cube $[P_3]$.

Shown in **Figure 6(b)** bottom is the loop-intermediate

$$PT[P_1]\circ PT[P_2]\circ T[P_3]\cdot L\left(\psi_{C_{base}}(M_0)\right). \tag{87}$$

The light grey area indicates the projection image of $[P_1]$ and $[P_2]$ by π. The dark grey area indicates the projection image of the removed $[P_3]$ by π. Note that normal edges of different heights are not directly connected.

Lemma 5.3. If two normal edges n_1 and n_2 of a flow of triangles are connected, the difference of their heights is even, *i.e.*,

$$ht_E(n_1) \equiv ht_E(n_2) \bmod 2. \tag{88}$$

In **Figure 6(c)** top, the normal edges of height −1 are connected to the normal edges of height 1 by putting the unit cube $[P_4]$ of height −1 on the tangent cone of **Figure 6(b)**. In **Figure 6(c)** bottom, the light grey area (height −1) and the dark grey area (height 1) are now in contact.

In **Figure 6(d)** top, the normal edges of height −2 are connected to the nor-

mal edges of height 0 by putting the unit cube $\left[P_5\right]$ of height −2 on the tangent cone of Figure 6(c). In Figure 6(d) bottom, the white area indicates the projection image of $\left[P_5\right]$. Note that the grey area (height 0) and the white area (height −2) are in contact.

5.2. Computation of Loops from the Hexagonal Base

In the loop model, it is easier to design an integral loop from scratch than to design a "hybrid" of known integral loops, since the area enclosed by a loop lp is included in $\left|lp\right|$. (In protein science, it is a formidable task to design a novel artificial protein from scratch.)

Let $M_0=\left\{m_0\right\}\in RI_{TCONE}$. Disconnecting normal edges of $L\left(\psi_{C_{base}(M_0)}\right)$ along the boundary, we obtain $lp\in I_B$ such that $\left|lp_0\right|=m_0$ as explained below.

Shown in Figure 7(a) top is the top view of the base tangent cone

$$C_{base}\left(M_0\right)\in TM_0 \tag{89}$$

of some $M_0\in RI_{TCONE}$. Shown in Figure 7(a) bottom is the loop-intermediate

$$L\left(\psi_{C_{base}(M_0)}\right) \tag{90}$$

on B. By definition, the heights of all normal edges of $\psi_{C_{base}(M_0)}$ are 0.

In Figure 7(b), normal edges of height 0 is disconnected by putting unit cubes of height −1 along the boundary on $C_{base}\left(M_0\right)$. The light grey area in Figure 7(b) bottom indicates the projection image of the added unit cubes.

In Figure 7(c), normal edges of height 0 is disconnected by taking unit cubes of height 0 along the boundary from $C_{base}\left(M_0\right)$. The dark grey area in Figure 7(c) bottom indicates the projection image of the removed unit cubes.

In Figure 7(d) left, a loop-intermediate consisting of three integral loops is obtained by putting 8 unit cubes of height −1 and taking a unit cube of height 0 along the boundary. Then, taking another unit cube of height 0 at the meeting point of the boundaries of the three loops, we obtain the integral loop shown in Figure 7(d) right.

5.3. Sufficient Conditions for L_0 to be a Loop

Sufficient conditions for L_0 of **Problem 5.1** to be a loop are given using the two concepts defined below.

Definition 5.4. (The set $N_{ht=0}(L)$ of normal edges) Let $L\in LI_{TCONE}$. $N_{ht=0}\left(L\right)$ denotes the set of all the normal edges of height 0 contained in L.

Definition 5.5. (Rifts of a Loop-Intermediate) A *crack* of $L\in LI_{TCONE}$ is a polygonal chain of normal edges of L connected to the boundary of L. A crack is called a *rift* if it consists of more than one normal edge.

Lemma 5.6. Let $M_0\in RI$. Let $L_1,L_2\in LI_{TCONE}$ such that $\left|L_1\right|,\left|L_2\right|\le M_0$. Let $\sigma\in\Gamma_T\left(\left\{\left|L_1\right|,\left|L_2\right|\right\}\right)$. If $G_H\left(L_1\right)\circ G_H\left(L_2\right)$ is well-defined, then

$$N_{ht=0}\left(L\left(\psi_{\sigma(|L_1|)}\right)+L\left(\psi_{\sigma(|L_2|)}\right)\right)=N_{ht=0}\left(L\left(\psi_{\sigma(|L_1|)}\right)\right)\cap N_{ht=0}\left(L\left(\psi_{\sigma(|L_2|)}\right)\right). \tag{91}$$

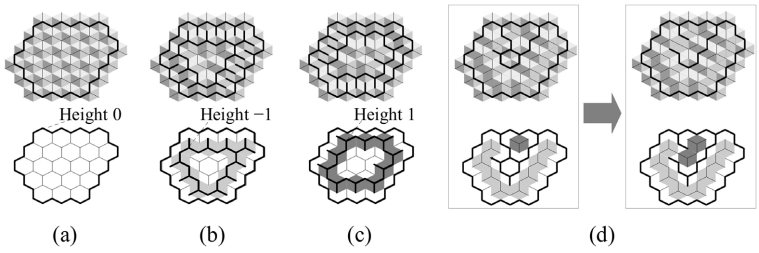

Figure 7. Computation of loops from the hexagonal base.

In general, the set $N_{ht=0}\left(\sum_i L\left(\psi_{\sigma(|L_i|)}\right)\right)$ shrinks monotonically as more loop-intermediates are added.

Proof. Because of **Remark 4.27**, removed normal edges of height 0 are removed forever. ■

Proposition 5.7. (Sufficient Conditions to be a Loop) Settings are the same as for **Problem 5.1.** Let $lp_i = L\left(\psi_{\sigma(c_i)}\right) \in I_B\left(i=1,2\right)$, *i.e.*,

$$L\left(\psi_{\sigma_{S_{lp}}\left(c_i \oplus (m_0 \setminus c_i)_H\right)}\right) = lp_i \oplus \left(m_0 \setminus c_i\right)_H \quad (i=1,2). \tag{92}$$

Then, L_0 consists of a single loop if $G_H\left(L_1\right) \circ G_H\left(L_2\right)$ is well-defined and one of the following three conditions are satisfied:

1) No cracks of lp_1 and lp_2 connect to $\partial M_0 \bigcap \partial\left(c_1 \bigcap c_2\right)$. There is at most only one rift of lp_1 that penetrates into $c_1 \bigcap c_2$ through $c_1 \setminus c_2$. No rift of lp_2 penetrates into $c_1 \bigcap c_2$ through $c_2 \setminus c_1$ (**Figure 8(a)** and **Figure 8(b)**).

2) No cracks of lp_1 and lp_2 connect to $\partial M_0 \bigcap \partial\left(c_1 \bigcap c_2\right)$. There is at most only one rift of lp_2 that penetrates into $c_1 \bigcap c_2$ through $c_2 \setminus c_1$. No rift of lp_1 penetrates into $c_1 \bigcap c_2$ through $c_1 \setminus c_2$.

3) Both lp_1 and lp_2 have a crack connected to $\partial M_0 \bigcap \partial\left(c_1 \bigcap c_2\right)$. No rift of lp_1 penetrates into $c_1 \bigcap c_2$ through $c_1 \setminus c_2$. No rift of lp_2 penetrates into $c_1 \bigcap c_2$ through $c_2 \setminus c_1$ (**Figure 8(c)**).

Proof. Since $G_H\left(L_1\right) \circ G_H\left(L_2\right)$ is well-defined, normal edges of height 0 are not added as a result of addition by **Lemma 5.6**. That is, cracks are extended only by normal edges of height *n*, where $n \in \boldsymbol{N}$ such that $n \neq 0$ and *n* is even. "No cracks of lp_1 and lp_2 connect to $\partial M_0 \bigcap \partial\left(c_1 \bigcap c_2\right)$" implies M_0 is not separated by a polygonal chain contained in $c_1 \bigcap c_2$.

Since lp_1 and lp_2 bring no normal edges of height n ($n \neq 0$) into $c_2 \setminus c_1$ and $c_1 \setminus c_2$, respectively, the result follows. ■

Shown in **Figure 9** are examples where the sufficient conditions are not satisfied.

In **Figure 9(a)**, both lp_1 and lp_2 have a rift that penetrates into $c_1 \bigcap c_2$ through $c_1 \setminus c_2$ and $c_2 \setminus c_1$, respectively. In **Figure 9(b)**, lp_1 has two rifts that penetrate into $c_1 \bigcap c_2$ through $c_1 \setminus c_2$. In **Figure 9(c)**, both lp_1 and lp_2 have a crack connected to $\partial M_0 \bigcap \partial\left(c_1 \bigcap c_2\right)$, and lp_1 has a rift that penetrates into $c_1 \bigcap c_2$ through $c_1 \setminus c_2$.

Figure 10(a) is the case given in **Figure 1(a)**. Shown in **Figure 10(a)** bottom are all normal edges of height 0, where both lp_1 and lp_2 have a rift consisting

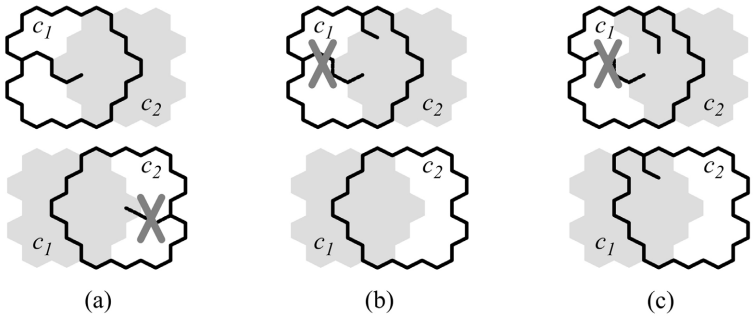

Figure 8. Sufficient conditions for L_0 of Problem 5.1 to be a loop.

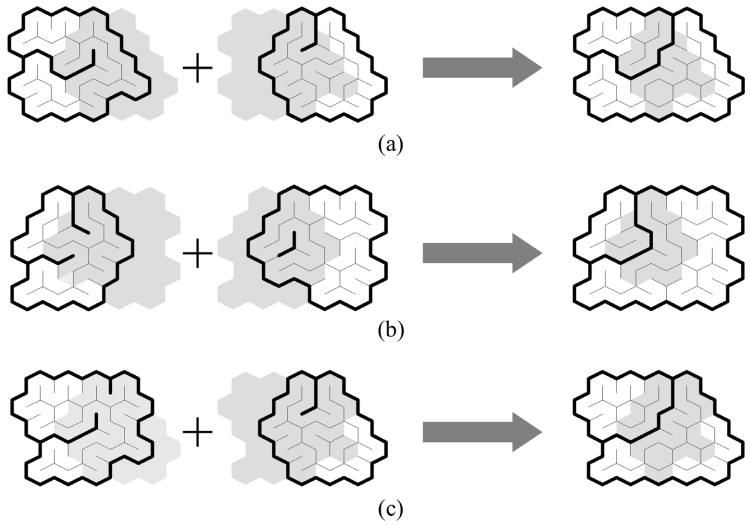

Figure 9. Examples where the sufficient conditions are not satisfied.

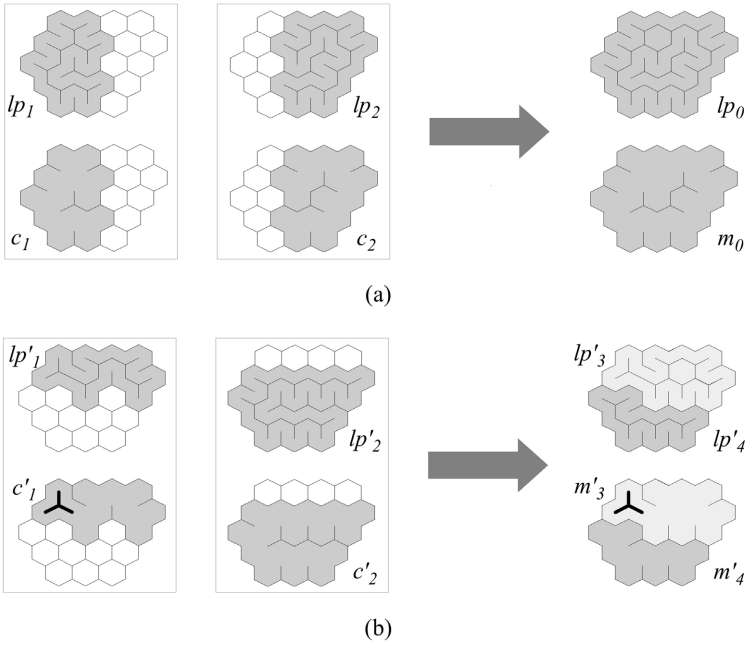

Figure 10. Incremental design of protein-like molecules (Examples given in **Figure 1**).

normal edges of height 0. All other normal edges are height 1. As a result of addition, some of the normal edges of height 0 are removed and we obtain the loop lp_0.

Figure 10(b) is the case given in **Figure 1(b)**. Shown in **Figure 10(b)** bottom are all normal edges of height 0 and height −1 (thick line segments). All other normal edges are height 1. In this case, $G(L_1, M_0) \circ G(L_2, M_0)$ is not well-defined and we cannot use **Lemma 5.6**. As a result of addition, the rift of lp_2' is extended by normal edges of height 0 and we obtain two loops lp_3' and lp_4'.

6. Discussion

A novel design method for protein-like molecules is proposed from the perspective of Sheaf Theory. In this method, a new molecule of a given shape is obtained as the sum of smaller molecules. Since the sum of loops is not a loop in general, sufficient conditions for a sum to be a loop are also considered. We believe this method is essential, especially when designing hybrids of known proteins.

Previous mathematical studies of protein structure have focused primarily on characterization and classification of structures, and the author is aware of no other mathematical research on protein design. As such, there is much room for improvement in this study, which is still in its infancy. The author hopes that this paper will inspire more mathematicians to become interested in the mathematical research on protein design.

As directions for future research, there are two directions. One is the study of three-dimensional case, in which protein-like molecules are represented as a loop of tetrahedra [12]. The other is the study of loops on various hexagonal meshes other than the "flat" mesh H considered in this paper [13]. Examples include hexagonal meshes on the surface of 3D molecules (*i.e.*, loops of tetrahedra). Note that a 2D triangular flow is induced on the surface of a complex of loops of tetrahedra.

In the three-dimensional case, two difficulties arise. First, the shape of a molecule is given on a mesh of dodecahedron, where a dodecahedron can be divided into four loops of tetrahedra (A hexagon cannot be divided into more than one loop of triangles). Second, the height of normal edges of tetrahedra is classified into three congruence classes of *modulo* 3, not two congruence classes of *modulo* 2.

Conflicts of Interest

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

References

[1] Berman, H.M., Westbrook, J., Feng, G., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Research*, **28**, 235-242. https://doi.org/10.1093/nar/28.1.235

[2] Taylor, W.R., May, A.C.W., Brown, N.P. and Aszódi, A. (2001) Protein Structure: Geometry, Topology and Classification. *Reports on Progress in Physics,* **64**, Article No. 517. https://doi.org/10.1088/0034-4885/64/4/203

[3] Albou, L.-P., Schwarz, B., Poch, O., Wurtz, J.M. and Moras, D. (2009) Defining and Characterizing Protein Surface Using Alpha Shapes. *Proteins*, **76**, 1-12. https://doi.org/10.1002/prot.22301

[4] Gromov, M. (2011) Crystals, Proteins, Stability and Isoperimetry. *Bulletin of the American Mathematical Society*, **48**, 229-257. https://doi.org/10.1090/S0273-0979-2010-01319-7

[5] Xia, K. and Wei, G.-W. (2014) Persistent Homology Analysis of Protein Structure, Flexibility and Folding. *International Journal for Numerical Methods in Biomedical Engineering*, **30**, 814-844. https://doi.org/10.1002/cnm.2655

[6] Penner, R.C. (2016) Moduli Spaces and Macromolecules. *Bulletin of the American Mathematical Society*, **53**, 217-268. https://doi.org/10.1090/bull/1524

[7] Zhao, R., Wang, M., Chen, J., Tong, Y. and Wei, G.-W. (2021) The de Rham-Hodge Analysis and Modeling of Biomolecules. *Bulletin of Mathematical Biology*, **82**, Article No. 108. https://doi.org/10.1007/s11538-020-00783-2

[8] 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

[9] Hartshorne, R. (1977) Algebraic Geometry. Springer-Verlag, New York. https://doi.org/10.1007/978-1-4757-3849-0

[10] Milewski, B. (2016) Category Theory for Programmers 1.1: Motivation and Philosophy. http://youtube.com/watch?v=I8LbkfSSR58

[11] Morikawa, N. (2020) On the Defining Equations of Protein's Shape from a Category Theoretical Point of View. *Applied Mathematics*, **11**, 890-916. https://doi.org/10.4236/am.2020.119058

[12] 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

[13] Morikawa, N. (2022) Discrete Exterior Calculus of Proteins and Their Cohomology. *Open Journal of Discrete Mathematics*, **12**, 47-63. https://doi.org/10.4236/ojdm.2022.123004