

RESEARCH PROJECT: PROTEINS AS NUMBERS

NAOTO MORIKAWA

1. OVERVIEW

1.1. About my research. I'm studying molecular biology as a frontier of mathematics. And my primary interest is construction of new mathematics rather than application of known mathematics. For example, I'm now constructing "hetero number theory" to study static structure of proteins and protein-protein interaction networks. And I'm proposing "arithmetic implementation" of cellular pathways, where proteins are represented by prime numbers. I'm also interested in denoting semantics of protein description languages by hetero numbers or natural numbers.

Recall that we use weights in scales for weighing objects and a ruler to measure their length. The set \mathbb{HN}^N of N -dimensional *hetero numbers* is a system of units for measuring shapes of objects such as proteins. It is characterized by correspondences between

- (1) protein-protein interaction and addition,
- (2) genetic code and the second derivative,

(Fig.1(a)). And \mathbb{HN}^N gives an example of commutative higher dimensional extension of the set \mathbb{N} of all natural numbers. In particular, we obtain "genetic codes" of natural numbers via a natural embedding of \mathbb{N} into \mathbb{PHN}^2 .

Arithmetic implementation of a process is a description of the process as an orbit of an action of integers on a subset of \mathbb{N} . Assigning prime numbers to proteins and their complexes, we obtain arithmetic implementations of biological activities conducted by proteins.

1.2. The research object.

Hetero number theory:

- Construction of hetero number theory.
- Protein structure prediction.
- Classification of the structure of proteins and protein-protein interaction networks.

Arithmetic implementation:

- Analysis of the dynamical system induced by the action.
- Analysis of evolution of cellular pathways.
- Classification of cellular pathways.

Protein description language:

- Description of semantics of protein description languages.

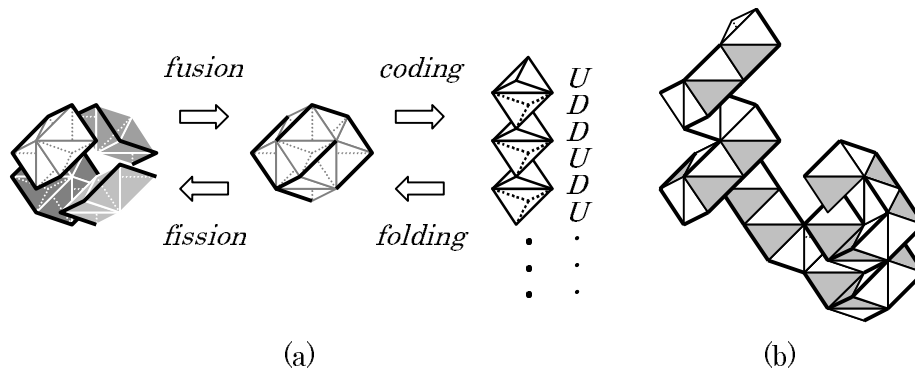


FIGURE 1. Four-dim. hetero numbers. (a): Features. (b): Simplex approximation of a protein (chain A of 2HIU).

2. HETERO NUMBER THEORY AND PROTEINS

2.1. Basic idea. Consider a unit cube in the three-dimensional Euclidean space \mathbb{R}^3 whose vertices are given by $v_1 = (0, 0, 0)$, $v_x = (1, 0, 0)$, $v_y = (0, 1, 0)$, $v_z = (0, 0, 1)$, \dots , and $v_{xyz} = (1, 1, 1)$. And draw lines $\overline{v_1v_{xy}}$, $\overline{v_1v_{yz}}$ and $\overline{v_1v_{xz}}$ (Fig.2(a)). Then, each of three upper faces is divided into two triangle tiles. For example, triangles $v_1v_xv_{xy}$ and $v_1v_yv_{xy}$ for the face $v_1v_xv_{xy}v_y$.

By piling up these cubes in the direction from v_{xyz} to v_1 , we obtain a drawing made up of triangle tiles (Fig.2(b)). *Three-dimensional hetero number* is the set of all closed trajectories of tiles defined by a drawing. For example, the left drawing of Fig.2(b) consists of one trajectory and the right consists of three.

Note that drawings are uniquely determined by their peaks. For example, the left of Fig.2(b) is determined by 4 peaks, center $(1, 1, 1)$, top $(2, 2, 0)$, left $(0, 2, 2)$, and right $(2, 0, 2)$. And we write $Cone^*\{xyz, x^2y^2, y^2z^2, x^2z^2\}$ for the drawing, where term $x^l y^m z^n$ corresponds to peak $(l, m, n) \in \mathbb{Z}^3$.

2.2. Simplex approximation of proteins. $Cone^*\{xyz, x^2y^2, y^2z^2, x^2z^2\}$ is also specified uniquely by ups (U) and downs (D) of the tiles along its closed trajectory:

$$D - U - U - D - D - U - D - U - \dots - D - D - U,$$

moving clockwise from the top (Fig.2(b) left). If we consider four-dimensional hetero numbers, we can describe proteins as trajectories of tetrahedrons (three-dimensional simplices), where one amino-acid corresponds to three tetrahedrons (Fig.1(c)). And the structure of a protein is also encoded into an U/D sequence.

2.3. Protein-protein interactions. Removing a cube at the peak $(1, 1, 1)$ from the left of Fig.2(b), we obtain the right one. And we denote the fusion of three closed trajectories as addition:

$$Cone^*\{xyz, x^2y^2, y^2z^2, x^2z^2\} \equiv Cone^*A_1 + Cone^*A_2 + Cone^*A_3,$$

where $A_1 = \{x^2y^2, xy^2z, x^2yz\}$, $A_2 = \{xy^2z, y^2z^2, xyz^2\}$, $A_3 = \{x^2yz, xyz^2, x^2z^2\}$.

Assigning hetero numbers to proteins and their complexes, we describe formations of protein complexes as addition of hetero numbers. For example, the above example denotes a formation of a protein complex of three proteins. For detailed discussion, see [1].

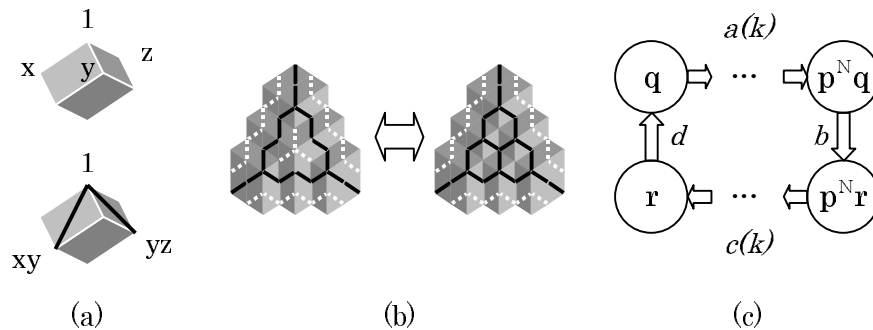


FIGURE 2. Three-dim. hetero numbers. (a): Unit cube in \mathbb{R}^3 . (b): Drawings of tiles. (c): Transition diagram of a process.

3. ARITHMETIC IMPLEMENTATION OF PROCESSES

3.1. Basic idea. Let $n = \prod_{1 \leq i \leq N} p_i^{l_i}$, $m = \prod_{1 \leq i \leq M} q_i^{k_i}$ be prime factorizations of $n, m \in \mathbb{N}$. And set $a = m - n \in \mathbb{Z}$. Then, addition $m = a + n$ induces an action $*$ of a on finite sets of primes:

$$a * \{p_1, \dots, p_1, \dots, p_N\} := \{q_1, \dots, q_1, \dots, q_M\}.$$

For example, $1 + 12 = 13$ defines an action of 1 on $\{2, 2, 3\}$: $1 * \{2, 2, 3\} = \{13\}$. That is, two proteins $2s$ and one protein 3 form a protein complex 13 by “interaction” 1.

For a finite set $\{p_1, \dots, p_k\}$ of primes, the action $*$ of \mathbb{Z} on $P := \{p_1^{l_1}, \dots, p_k^{l_k} \mid 0 \leq l_i \in \mathbb{Z}\} \cup \{0\} \subset \mathbb{N}$ is defined by

$$* : \mathbb{Z} \times P \rightarrow P, \quad a * s := \begin{cases} a + s & \text{if } a + s \in P, \\ 0 & \text{else.} \end{cases}$$

And $A * s := \sum_{a \in A} a * s$ if $\sum_{a \in A} a * s \in P$ for a finite subset A of \mathbb{Z} .

3.2. Arithmetic Implementation. Let $P = \{p^l q^m r^n \mid 0 \leq l, m, n \in \mathbb{Z}\}$. The process specified by the transition diagram of Fig.2(c) corresponds to the orbit defined by

$$s_{n+1} = A * s_n \quad (n > 0), \quad s_0 = q,$$

$$\text{and } A = \{a(k), b, c(k+1), d \mid k = 0, 1, \dots, N-1\} \subset \mathbb{Z},$$

where

$$a(k) := p^{k+1}q - p^kq, \quad b := p^N r - p^N q, \quad c(k) := p^{k-1}r - p^k r, \quad \text{and } d := q - r.$$

Then, for example, the process is implemented as an orbit of the action of

$$A = \{34, 102, -36, -78, -26, 4\} \text{ on } P = \{3^l 17^m 13^n \mid 0 \leq l, m, n \in \mathbb{Z}\}$$

if $N = 2$. See [2] for more examples.

REFERENCES

1. N.Morikawa, *Towards $\text{Sub}(\mathbb{Z}^N)$ implementation of protein-protein interactions*, 2004 (manuscript).
2. N.Morikawa, *$\text{Sub}(\mathbb{Z})$ implementation of biological activities conducted by proteins*, 2004 (manuscript).

E-mail address: mmorika@f3.dion.ne.jp