数理解析研究所講究録
第1704巻 2010年 172-186

Remarks to study about distribution of DNA-knot by use of Jones polynomial of topological invariant

Isamu Ohnishi[0] Takashi Yoshino[1]

Department of Mathematical and Life Sciences,
Graduate School of Science, Hiroshima University

Abstract

In this paper, a finer mechanism of transcription process of gene as a cost-saving and resource-saving transportation system of codewords information is investigated by studying distribution of the DNA knots by use of topological invariants to λ phage's structural operation as one of interesting examples. This work is just a finer mathematical analysis to the mechanism by use of topological invariants, but better understanding makes better using, and moreover, it is known that there are some cases in which some pieces of global topological information is very important for transcription process. To use the topological invariants has another important point, because it makes the simulation time be shorter and will be able to make simulation itself be precise even in the case of big crossing number. We use two types of topological invariants: Alexander polynomials and Jones polynomials. Generally speaking, Alexander polynomials are easy to use, but have lower separating ability. On the other hand, Jones polynomials are somewhat difficult to utilize, but have finer separating ability. We compare the results gotten by both ways of simulations to discuss about both advantages and disadvantage to suggest that they should be used according to the objectives. Moreover, these results are compared with the actual biochemical experimental results to make an interpretation to what happens in the cell, and we will state perspectives in the future.

1 Introduction

Due to the central dogma of molecular biology, the transcription-translation process based on informations on DNA is essential for decision making in a cell, but nowaday, it is well-known that there only are fewer genes than expected on a DNA, compared to the diversity of results of expression of genes. For example, in human-being, before the genome project, it was said that there must be more than 100,000 genes, although only 22,000 approximately genes have existed actually. Several devices for transport of informations on DNA make it possible and, as a consequence, it saves costs about the transcription-translation system. Analysing the mechanism is not only important for progress of molecular biology, but also it has a possibility for the transportation system of complex informations, at least in nano-scale, to be improved very well by use of the mechanism. For the purpose, as the first step, it is nescessary for the biological system to be analysed and to be understood in details experimentally as well as mathematically. When the transcription and translation occur, genomes are variously modified, and it is because of the diversity of expression. Typical operations are "conservative site specific recombination" and "transposition". These two operations are responsible for a lot of events in a cell.

"Conservative site specific recombination" occurs at defined sequence elements in a DNA. Recombinase proteins recognize these sequence elements and act to containing the recombination sites. Three types of rearrangements are common: DNA insertion, DNA deletion, and DNA inversion. These rearrangements

[0] isamu_o@math.sci.hiroshima-u.ac.jp
[1] m083450@hiroshima-u.ac.jp

have many functions, including insertion of a viral genome into that of the host cell during infection, ro-solving DNA multimers, and altering gene expression. There are two families of conservative site-specific recombinases. Both families cleave DNA using a protein-DNA covalent intermediate. Transposition is a class of recombination that moves mobile genetic elements, so called transposons, to new genomic sites. There are three major classes of transposons. Transposons existin the genomes of all organisms, where they can constitute a huge distribution of the total DNA sequence. They are a major cause of mutations and genome rearrangements. Transposition is often regurated to help ensure that transposons do not cause too much of a disruption to the genome of the host cell. Control of transposon copy number and regulation of the choice of new insertion sites are commonly observed. (See [7] for the details about the biological point of views.)

As in the above summary, by use of very clever spontaneous mechanisms in nano-level in a cell, the cost-saving effective transport system is established and realized. A kind of leaning of distribution for DNA is often produced by such kind of operations of the enzyme of nucleic acid. If this kind of leaning can be calculated experimantally and it can be interpreted theoretically, then it should be useful as an index about degree of the operations: How much was it operated by the enzyme ? If so, it also is an important index in order to understand the control and the reguration. We pick up topological type of DNA knot operated by macro-phage in E. coli (See in [2]). As in Procaryote, there is creature with ring type DNA. Distribution of topological type of DNA knot and its changing give us an index of the degree of the operations. In this paper, calculation method and its result are presented by use of topological invariants of knot: Alexander polynomials and the Jones polynomials. Both are well known as the topological invariants for the knots and each quantity has its advantages and its disadvantages. We calculate the Distribution of topological type of DNA knots in both ways, and consider about meaning of these. Moreover we compare them to the biochemical experimental results, and make a study about how much it can be operated to give a theoretical result.

Before stating our method and result, we revise the structure of DNA and some examples of nucleic acid enzyme changing DNA structure:

1.1 Topological Structures of a DNA

The structure of DNA is a double helix, much like a ladder that is twisted into a spiral shape. In 1953 the double-helix structure of a DNA was proposed by J. D. Watson and F. H. Crick with technique to build a molecular model. The double helix is the ideal model that J. D. Watson and F. H. Crick arrived at from many studies about a DNA and seven next important characteristics are emphasized in the structure.

1. The double helix is formed by two polynucleotide.
2. A purine and a pyrimidine ring orient it in the inside of the double helix.
3. The base in complementary relations is bound by hydrogen bond.
4. There is 10.4 base pair per spiral 1 round per minute.
5. As for two polynucleotide of the double helix, a direction is reverse (reverse parallel) each .
6. Double helix has two kinds of grooves, the major groove and the minor groove.
7. The double helix is clockwise twining (right handed).

When a double-stranded DNA of the double-helix structure is twisted more or it is untied, the double helix is more twisted as a whole. These are called the supercoiled structure (a positively supercoiled DNA is the former and a negatively supercoiled DNA is the latter, respectively).

The ringed supercoiled DNA closed by covalent bond has "the twist" and "the writhe". "The twist" is the spiral number of revolutions that one chain surrounds the chain of the other, that is the number of times that one chain completely coils the chain of the other. "The twist" in a right-handed spiral is defined as plus. "The writhe" is a number that the double strand coils oneself or the double strand is coiled cylindrically like the cord of the telephone. The total value of "the twist" and "the writhe" cannot be changed unless we cut off the DNA. We consider of the topoisomerase for an example as an enzyme

changing topological structure of the DNA.

1.2 Topoisomerase

We here make an introduce to Topoisomerase as an important example of enzyme of nucleic acid which is changing the DNA topology while it works. The topoisomerase is the enzyme which can change structures of the DNA by cutting the single strand or the double strand of the DNA temporarily.

The topoisomerase has two types. Type I topoisomerases cut the single strand of the DNA temporarily and connect it again after dipping the other single strand. Type II topoisomerases cut the double strand temporarily and close after dipping another double strand which is not cut.

Both a prokaryote and organism with a nucleus have type I and II topoisomerase which can remove the supercoiled structure from the DNA. Prokaryotes possess two structurally similar type II topoisomerases, Gyrase and Topo IV, which differ in their function. Gyrase generates negative supercoils, but Topo IV relaxes positive supercoils. In addition, it has reported that Gyrase is strongly bound with DNA, but Topo IV is weak [1].

The global structure of DNA strongly ties to the expression pattern of DNA by the property of Topo IV and Gyrase, since the different DNA structure tends to produce different DNA modification even under the same local condition. For example, when Gyrase was added to the positively supercoiled DNA, Gyrase unties DNA since Gyrase generates a negatively supercoils, but when Gyrase was added to the negatively supercoiled DNA, Gyrase twists DNA more. Thus, the global structure of the DNA is important for an enzyme.

It is known that the global topological structures of DNA tend to affect the expression of transcripts. DNA is affected by other nucleic acids and enzymes in the process of DNA transcription. The global structure of DNA strongly ties to the expression pattern of DNA, since the different DNA structure tends to produce different DNA modification even under the same local condition. Then it has been reported that the global topological structures of DNA influences expression of transcripts by performing various topological changes in genetic recombination [?]. In this paper we describe the topological structures of DNA in a phage.

1.3 Structures of a λ phage

A λ phage is a virus which infects bacteria. It is well known that it has tails and a regular icosahedral head(see Fig.1). The phage is the simplest system to study a basic process of life. The genome is small generally and the genome reproduce only after invading a host cell (a bacteria cell). Then a gene is expressed. A genome tends to be rearranged while infection.

Because the system is simple, the phage was used very well at the molecular biological dawn. Actually, the phage was indispensable for the development of this field. The system is also most suitable for a study of the DNA replication, the gene expression and basic structures of the recombination today. Furthermore, a phage is important as a vector of the DNA recombination, and it is used to examine the activity inducting variation of various compounds.

Phages usually pack genomes (generally, DNA) to a mantle consisting of subunits of the protein. A subunit has a part making head structure (packing genomes) and an another part making tail structure. The phage particle is glued to the outside of the bacteria host cell with a tail and injects the phage genome to the cell. Because each phage glues to a specific molecule (usually, protein) in the cell surface, only a cell having a coping acceptor is infected with the phage.

A phage has a double-stranded DNA controlling a genetic information in the head. The linear DNA packed in the head of phages forms various types of knots. As the part of knots, Fig.2 shows knots 0_1, 3_1, 4_1, 5_1, 5_2, 6_1, 6_2, 6_3, 7_1, 7_2, 7_3, 7_4, 7_5, 7_6 and 7_7.

It is the important step of the DNA studies to investigate the Distribution of various types of knots and clarify the reasons of the phenomena.

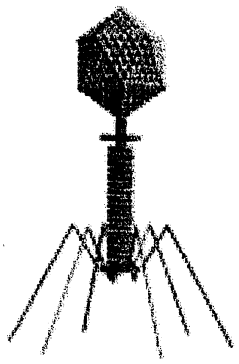

Fig. 1 A bacteriophage

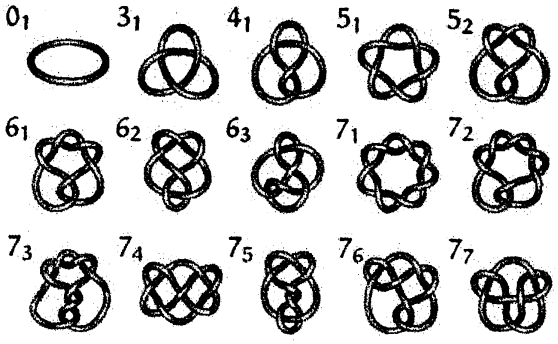

Fig. 2 The prime knots having below 7 crossing

1.4 Results of biological experiments

The DNA inside the phage was examined types of knots which is made inside the phage and the probability by J. Arsuaga et. al. [2]. First, purified DNA was analyzed by two-dimensional gel electrophoresis. By gel electrophoresis, we can distinguish some knot types with the number of the same crossing number. To the direction of I of Fig. 1 by [2], DNA samples were run at $0.8V/cm$ for $40h$ at room temperature. After a 90° rotation of the gel, the direction of II of Fig. 1 by [2] was run in the same electrophoresis buffer at $3.4V/cm$ for $4h$ at room temperature. This gel electrophoresis segregated the linear DNA molecules and segregated knots with six and more crossings to two groups. Knot populations of low crossing number are showed in Fig. 1 by [2]. The number of the right in Fig. 1 by [2] is indicated the crossing number.

Next, the projection obtained by this experiment was compared with the projection of known knot types to examine the knots corresponding DNA knots separated by electrophoresis (see [2]). The gel velocity at low voltage of individual knot populations resolved by two dimensional electrophoresis (Right of Fig. 2 by [2]) is compared with the gel velocity at low voltage of twist knots (3_1, 4_1, 5_2, 6_1 and 7_2) of a $10 - kb$ nicked plasmid (Center of Fig. 2 by [2]) and with known relative migration distances of some knot types (Left of Fig. 2 by [2]). By this comparison, some knot types of each population are distinguished. In addition to the unambiguous knots 3_1 and 4_1, the knot population of five crossings matched the migration of the knot 5_1. The knot 5_2 appeared to be negligible or absent. The knot population of seven crossings matched the migration of the torus knot 7_1 rather than the twist knot 7_2. Yet, we cannot identify this gel band as the knot 7_1, because other possible knot types of seven crossings cannot be excluded.

Then, several indicators led us to believe that the second arch of the gel ($6'\sim9'$ of Fig. 1 by [2]) consists of mainly composite knots (see Def.2.6 in Chapter 2). First, the arch starts at knot populations containing six crossings, and no composite knots of fewer than six crossings exist. Second, the population of six crossings matched the migration at low voltage of the granny knot (composite of a 3_1 plus a 3_1, indicated as $3_1\#3_1$) , although the square knot (the other possible composite of six crossings, $3_1\# - 3_1$) cannot be excluded. Third, consistent with the low amount of 4_1 knots, the size of the seven-crossing subpopulation is also reduced; thus, any composite seven-crossing knot is $3_1\#4_1$ (or $-3_1\#4_1$). The increased gel velocity at high voltage of composite knots relative to prime knots (see Def.2.5 in Chapter 2) of the same crossing number likely reflects distinct flexibility properties of the composites during electrophoresis.

Densitometer readings confirmed the apparent scarcity of the knot of four crossings (knot 4_1) relative to the other knot populations in the main arch of the gel. Fig.3 shows the results of the Distribution of knots 3_1, 4_1, 5_1 and 5_2 by biological experiments. However, the scarcity of the knot 4_1 relative to the

knot of three crossings (knot 3_1) and to other knot populations is enhanced if we make the correction for DNA molecules plausibly knotted outside the viral capsid. In such a case, one can predict that 38% of the total number of observed 3_1 knots and 75% of the observed 4_1 knots are formed by random knotting in free solution.

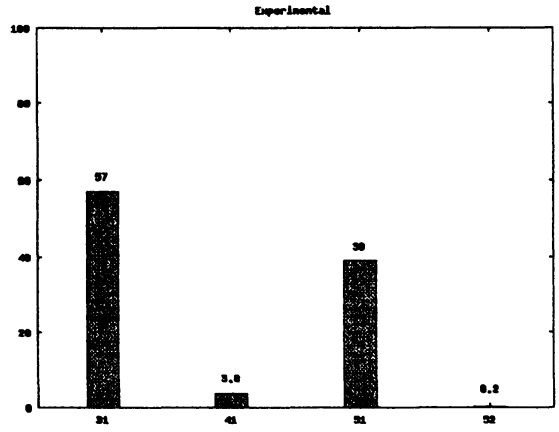

Fig. 3 The results of the Distribution of knots 3_1, 4_1, 5_1 and 5_2 by biological experiments [2]

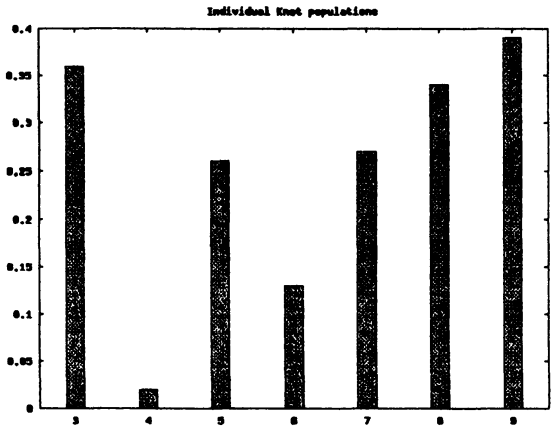

Fig. 4 Quantification of the individual knot populations of three to nine crossings

1.5 Outline of this study

To study the geometry of the DNA inside a λ phage, we simulate Distribution of the various types of knots and compared to those of biological experiments.

First, we simulate under the assumption that DNA knots inside the phage are generated at random. With Mersenne Twister we randomly make the knot topology having n crossings ($n = 3, 4, \cdots, 8$). This process is repeated $1,000,000$ times to obtain $1,000,000$ knots. Then we classified these knots with the invariant for the knot and estimate the Distribution of the knots. These trials are repeated 100 times and we consider the mean value of all trials as the conclusive Distribution of the above-mentioned types of knots.

Next, we examined "the writhe value" Wr at each crossing. After eliminating matrices having the small value of the total amount $\sum Wr$, we performed the same simulation on the remaining matrices. The Distribution of the knots 3_1, 4_1, 5_1 and 5_2 after considering "the writhe value" were similar to the biological results.

2 Preparations from Topology

Def. 2.1 (Alexander matrix) Let G be a group and let $\alpha : G \to G/[G,G]$ be the Abel map. The image $A(G,\alpha)$ of Jacobian by the Abel map α is called **Alexander matrix** of G.

Def. 2.2 (Alexander polynomial) The g. c. d. of determinant of $(n-1) \times (n-1)$ submatrix of $n \times n$ Alexander matrix is called **Alexander polynomial**.

Def. 2.3 (Trivial knot) Let $K \subset S^3$ be a knot. If there exists $K \subset S^2$ for a two-dimensional sphere $S^2 \subset S^3$, then K is called **the trivial knot**.

Def. 2.4 (Connected sum of knot) Let $K_1 \subset S^3$, $K_2 \subset S^3$ be knots and let $B_1 = S^3 - intN(x_1)$, $B_2 = S^3 - intN(x_2)$ for points $x_1 \in K_1$, $x_2 \in K_2$. If we set $S_i = \partial B_i$ $(i = 1, 2)$, then S_i is a two-dimensional sphere and $K_i \cap S_i$ is a one-dimensional sphere in S_i.

Then we identify S_1 with S_2 and identify $K_1 \cap S_1$ with $K_2 \cap S_2$ by a homeomorphic mapping $(S_1, K_1 \cap S_1) \to (S_2, K_2 \cap S_2)$, and we connect $(B_1, K_1 \cap B_1)$ and $(B_2, K_2 \cap B_2)$.

The obtained knot $K \subset S^3$ is called **the connected sum** of K_1 and K_2, we use the notation $K_1 \sharp K_2$. Conversely, it is called that K is decomposed K_1 and K_2.

Def. 2.5 (Prime knot) For the arbitrary decomposition $K = K_1 \sharp K_2$ of the knot K, if at least one of K_1 or K_2 is the trivial knot, then K is called **the prime knot**. However, we define that the trivial knot is not the prime knot.

Def. 2.6 (Composite knot) If there exists a decomposition $K = K_1 \sharp K_2$ of the knot K such that K_1 and K_2 is the non-trivial knot which is not prime, then K is called **the composite knot**.

Def. 2.7 (Skein relation) Let t be the nonzero complex number and let L_+, L_- and L_0 be three knots that only one part crosses like a Fig.5 and the other part is totally same.

Then

$$tV_{L_+}(t) - t^{-1}V_{L_-}(t) = -(t^{1/2} - t^{-1/2})V_{L_0}(t)$$

is called **Skein relation**.

Fig. 5 Three knots L_+, L_- and L_0

Def. 2.8 (Jones polynomial) The polynomial $V_L(t)$ such that Skein relation hold is called **Jones polynomial**. However, Jones polynomial of the trivial knot of one component defines 1.

Def. 2.9 (Bracket polynomial) The polynomial $< D >$ of one variable A obtained by applying next three rules to the link diagram D is called **Bracket polynomial**:

(i) $< U >= 1$;

(ii) $< DU >= -(A^2 + A^{-2}) < D >$;

(iii) $< \times >= A < \asymp > + A^{-1} <)(>$;

$< \asymp >= A <)(> + A^{-1} < \times >$;

where U is a trivial knot, DU is an union with a trivial knot and we apply (iii) locally at each crossing of the link diagram.

Then for the link diagram D of an arbitrary knot L with the direction, Jones polynomial $V_L(t)$ has a relation with Bracket polynomial $< D >$ by

$$V_L(A^{-4}) = (-A^3)^{-\omega(D)} < D >,$$

where $\omega(D)$ is the sum of "the writhe value" at each crossing of the link diagram D (see Section 3.1.1).

It is known that Alexander polynomial and Jones polynomial, which are the invariant for knots, is effectively used for the classification of knots based on the topology. We can classify all prime knots having below 8 crossing by a combination of the value of $\Delta(-2)$ and $\Delta(-3)$ when we use Alexander polynomial $\Delta(t)$. Then we can distinguish the mirror images of knots and knots having more crossings when we use Jones polynomial.

3 Simulations

3.1 Alexander polynomial

3.1.1 Elements of Alexander matrix

We shall project the knot on a place along an arbitrarily chosen axis, while drawing breaks at the crossing points in the part of the curve that lies below. Now the projection of the knot amounts to the set of segments of curves, which are called the generators. Let us fix arbitrarily the direction of passage of the generators and number them, having selected arbitrarily the first generator. The crossing that separates the kth and $(k+1)$th generators will be called the kth crossing. The crossings are of two types (see Fig.6). The writhe value Wr is $Wr = -1$ at type I crossing and $Wr = 1$ at type II crossing. The elements of Alexander matrix $A(t)$ are defined as follows [3].

1. When $i = k$ or $i = k+1$, independently of the type of crossing: $a_{kk} = -1$, $a_{kk+1} = 1$.
2. When $i \neq k$, $i \neq k+1$, for a type I crossing: $a_{kk} = 1$, $a_{kk+1} = -t$, $a_{ki} = t-1$.
   for a type II crossing: $a_{kk} = -t$, $a_{kk+1} = 1$, $a_{ki} = t-1$.
3. All the elements except a_{kk}, a_{kk+1} and a_{ki} are zero.

Here k is the number of crossing and i is the number of the overpassing generator. Then these relationships hold for all $k = 1, \cdots, n$ under the condition of the substitution $n+1 \to 1$.

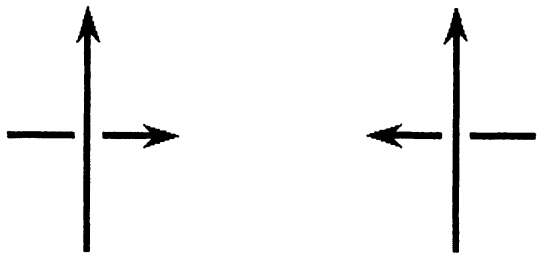

Fig. 6 The type I crossing (left) and the type II crossing (right)

3.1.2 Generation of Alexander matrix

First, we fix the dimension n of Alexander matrix from 3~8 with Mersenne Twister at random. Second, we fix the element a_{kk} $(k = 1, 2, \cdots, n)$ of $n \times n$ matrix $A(-2), A(-3)$ $(n = 3, 4, \cdots, 8)$ substituted $t = -2, t = -3$ with Mersenne Twister as to satisfy above list. When $i \neq k$ and $i \neq k+1$, we take i with Mersenne Twister at random (i must not take k and $k+1$). This work is repeated from $k = 1$ to $k = n$ and we make the $n \times n$ matrix corresponding to the random knot.

3.1.3 Reidemeister moves

By Reidemeister moves, we untie the part which can untie of the knot. Reidemeister moves I, II (Fig.7) are the next operation [3].

I. If the kth row contains only two nonzero elements $a_{kk} = -1$ and $a_{kk+1} = 1$, then we should add the column k to the column $k+1$. Then we delete the kth column and the kth row and renumber the rows and column afresh.

II. If in two adjacent rows of the matrix having the numbers k and $k+1$ elements having the value $t-1$ lie in a single column and elements equal to $t-1$ are lacking in the $(k+1)$th

column, then we should add the kth column to the $(k+2)$th, and then delete the kth and $(k+1)$th rows and columns and renumber all the rows and columns afresh.

The above list I and II correspond Reidemeister moves I and II, respectively.

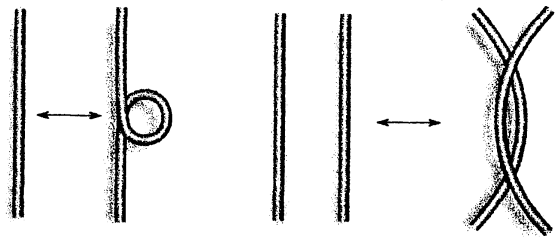

Fig. 7 Reidemeister moves I (left), II (right)

3.1.4 Classification of knots

We estimate the value $\Delta(-2)$, $\Delta(-3)$ of the determinant of the $(n-1) \times (n-1)$ submatrix and classify knots by the combination of the value of the determinant. Here, we removed all rows and columns from the first row and column to the nth row and column one by one. Then we treated as a knot when the value $\Delta(-2)$, $\Delta(-3)$ of all determinants was equal or different $\pm 2^m$ times, $\pm 3^m$ times (m is an integer).

3.1.5 Reliability of the simulation

We gave 1,000,000 matrices at random and counted the number which generated each knot and estimated the Distribution. These trials were repeated 100 times and we considered the mean value of all trials as the conclusive Distribution of the types of knots.

3.1.6 Addition of "the writhe"

After the simulation of the random matrices, we employed "the writhe value" in the simulation. we examined "the writhe value Wr" at each crossing. After eliminating matrices having the small value of the total amount $\sum Wr$, we performed the same simulation on the remaining matrices.

3.2 Jones polynomial

3.2.1 Recurrent algorithm of Bracket polynomial

Because we can find Jones polynomial from Bracket polynomial if we find "the writhe value" of the knot, we constitute algorithm by Bracket polynomial [4].

For an arbitrary link diagram D, we paint with two colors of the black and white so that adjacent domains do not become the same color. Here, infinite domain is the white. Next, we choose one point in each black domain. If two black domains are adjacent, we link the point that we chose of each domain and we establish the sign of the edge based on Fig.8. Thus, the plane graph S with the sign is obtained.

We define that $S\backslash e$ is the graph which deleted the edge e from the graph S and S/e is the graph which contracted the edge e from S (that is the edge e is condensed and the top of both ends correspond one point). Then we define that the loop is the edge where both ends are the same tops and the bridge is the edge where the connected component of the graph increases one when the edge is deleted.

It is known that the contraction / deletion formula of the edge holds for graph S with the sign. For example, when the edge linking two adjacent black domains is positive, the operation which tie two black domains and the operation which divide two black domains is equivalent to the contraction and the deletion, respectively.

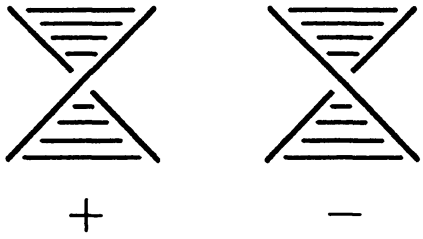

Fig. 8 The sign in two adjacent black domains

$$< S >= A < S/e > +A^{-1} < S\backslash e >$$

Then when the edge e is the positive bridge, $S\backslash e$ is disconnectedness (that is the number of the component of the knot corresponding to $S\backslash e$ increases one), but $S\backslash e$ and S/e is the isomorphism, and the following relations hold between them.

$$< S\backslash e >= (-A^2 - A^{-2}) < S/e >$$

Therefore, they hold the following relations when the edge e is the positive bridge.

$$\begin{aligned} < S > &= A < S/e > +A^{-1} < S\backslash e > \\ &= -A^{-3} < S\backslash e > \end{aligned}$$

It is similar in the case of the negative bridge or the loop too, therefore they hold the following recurrence relations.

$$< S >= \begin{cases} -A^{-3} < S/e > & \text{if } e \text{ is the positive bridge} \\ -A^{3} < S/e > & \text{if } e \text{ is the negative bridge} \\ -A^{3} < S\backslash e > & \text{if } e \text{ is the positive loop} \\ -A^{-3} < S\backslash e > & \text{if } e \text{ is the negative loop} \\ A < S/e > +A^{-1} < S\backslash e > & \text{if } e \text{ is positive (otherwise)} \\ A < S\backslash e > +A^{-1} < S/e > & \text{if } e \text{ is negative (otherwise)} \end{cases}$$

Then when S consists of one edge (the positive bridge C^+, the negative bridge C^-, the positive loop L^+, the negative loop L^-), Bracket polynomials hold the following relations.

$$< C^+ >= -A^{-3},\ < C^- >= -A^3,\ < L^+ >= -A^3,\ < L^- >= -A^{-3}$$

Furthermore, when the edge set of S is the empty, Bracket polynomial has 1 from 2.9 (i).

3.2.2 Generation of the random plane graph with the sign

First, we fix the number of the top from 2~9 with Mersenne Twister at random and also fix the number of the edge from 3~9 at random. Next, we choose tops linked each edges at random and give the sign for each edges. For the sign of each edges, we tried three kinds; at random, only plus and only minus. Thus, we generate the random plane graph with the sign.

3.2.3 Classification of knots

We evaluate the recurrence algorithm of Bracket polynomial for the given graph. Then we classify knots by the polynomial which is finally obtained. Here, because we considered the difference of the value of "the writhe value" of the knot, we also treat as the knot when the polynomial which is finally obtained is different $(-A)^{3m}$ times (m is an integer).

3.2.4 Reliability of the simulation

We give 100,000 matrices at random and counted the number which generate each knot and estimate the Distribution. These trials are repeated 100 times and we consider the mean value of all trials as the conclusive Distribution of the types of knots.

3.2.5 Decrease of the BDD size

Since Bracket polynomial is the same for 2-isomorphic graphs, we may represent 2-isomorphic minors among them by one of these members. However, testing the 2-isomorphism of arbitrary graphs is hard in general, and finding all 2-isomorphic minors may be difficult.

The 2-isomorphism between two graphs whose edges have an identity map can be determined in linear time [5]. For this reason, we may restrict ourselves just to find 2-isomorphic minors whose corresponding edges have the same order in the original graph S. By modification, the expansion tree becomes an acyclic graph (edges are directed from a parent to a child). This acyclic graph has a single source (the original graph S) and the m-th level may be regarded as a single sink. We call this acyclic graph the BDD of all spanning trees of a graph S with respect to the edge ordering $e_1, e_2, \cdots, e_m$. This is because this acyclic graph has strong connection with the BDD (Binary Decision Diagram) representing all the spanning tres of a graph. The size of this BDD is defined to be the number of its nodes. The width of this BDD is defined to be the maximum among the numbers of nodes at each level.

In graph S with the sign, any signs of remaining edges don't change even if we contract or delete the edge e. Therefore, the isomorphic minor which is also the same order of edges is the isomorphism including the sign of each edge. That is, we can unify all isomorphic minor which is also the same order of edges and calculate it.

We can reduce the number of minor if we take the order of calculating edges and tops well. First we decide the first top p_1. The top p_1 is chosen as follows.

1. We choose the top which connects to a lot of same edges.
2. (If there are several tops such as 1,) we choose the top with a little number of connected tops.

Next we decide the top p_2. We choose the top p_2 which is worked above 1., 2. for tops connected to the top p_1. Furthermore we choose also the top p_3 which is worked above 1., 2. for tops connected to the top p_1. If the top p_3 isn't connected to the top p_1, we choose the top p_3 which is worked above 1., 2. for tops connected to the top p_2. We choose after the top p_4 equally. Thus we decide the order of tops and calculate the graph from the edge which is connected to small number tops. In this way, we were able to calculate Jones polynomial up to the maximum 20 crossings. However, in this algorithm, the calculation is finished only Jones polynomial up to the maximum 16 crossings because the calculation is taken quite a long time. Before decreasing BDD size, for n crossings, the number of minor was need 2^n theoretically, but by our simulation, the number of minor was decreased 44, 52, 100, 80, 109, 169 for 11, 12, 13, 14, 15, 16 crossings respectively. However, we may make better algorithm if we see to ref.[4].

4 Results, Discussions and Concluding remarks

4.1 Comparison of the results by Alexander polynomial and those by Jones polynomial

When knots are generated at random, Fig.9 shows the results by Alexander polynomial and Fig.11 shows those by Jones polynomial (*-suffix in Fig.11 represents the mirror images of the knots). In both results, the distribution of the knot 3_1 is the highest, the distribution of 4_1 is the second highest, 5_2 is the third and 5_1 is the fourth. In the group of the knots having 6 crossing, the distribution of the knot 6_2 is the highest, the distribution of 6_3 is the second and 6_1 is the third. In this way, the distribution of each knot obtains the almost same tendency, even if any topological invariant is used. Furthermore, by the results by Jones polynomial, the population of each knot possesses the approximately same as the one of the mirror images at each knot group having the same number of the crossing.

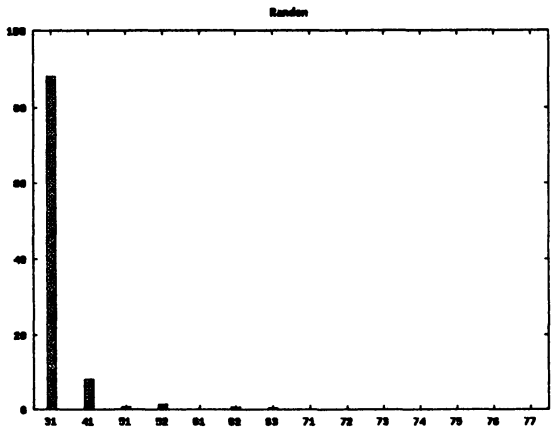

Fig. 9 Distribution distinguished knots generated at random by Alexander polynomial

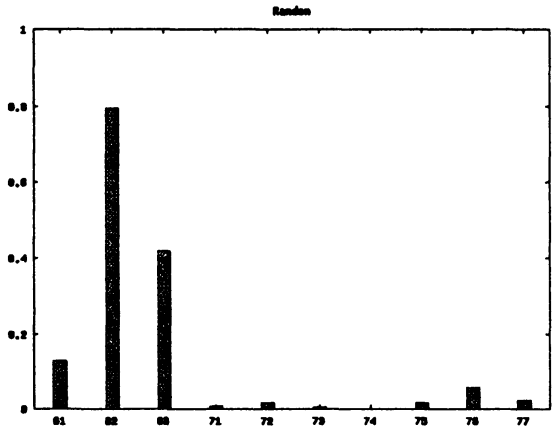

Fig. 10 Enlarged picture of the part of 6~7 crossings of Fig.9

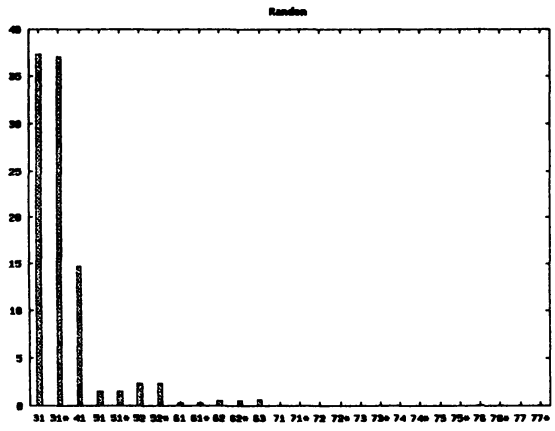

Fig. 11 Distribution distinguished knots generated at random by Jones polynomial

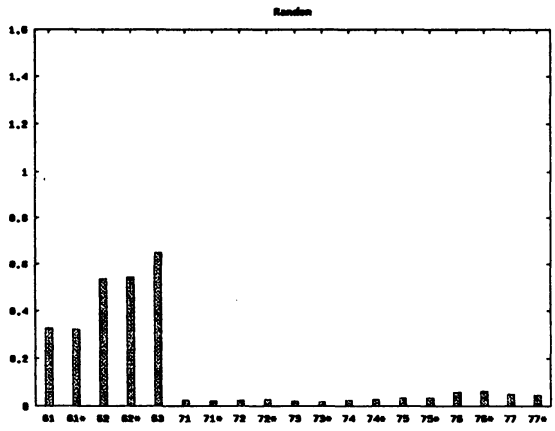

Fig. 12 Enlarged picture of the part of 6~7 crossings of Fig.11

4.2 Comparison of the results of the simulation and those of biological experiments

About knots 3_1, 4_1, 5_1 and 5_2, we compare the results of the simulation by Alexander polynomial with the results of biochemical experiments. The distribution of knot types that emerges in biochemical experiments done by Professors J. Arsuaga et. al. [2] is shown in Fig.13. In the biochemical experiments the distribution of the knot 3_1 is the highest and the Distribution of 5_1 is the second highest, while 4_1 and 5_2 are low (Fig.13). In contrast, the distribution of knots that are computationally generated at random (Fig.14) is very different from the results of biochemical experiments. In Fig.14, the abundance of 4_1 is higher than 5_1 and 5_2.

However, by employing "the writhe value" in the simulation, the result is changed. The distribution of the knots 3_1, 4_1, 5_1 and 5_2 after considering "the writhe value" tends to be rather similar to the biochemical result (Fig.16).

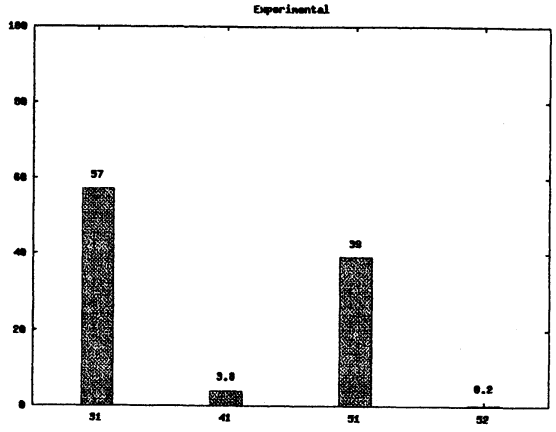

Fig. 13 The results of biological experiments by J. Arsuaga et. al. [2]

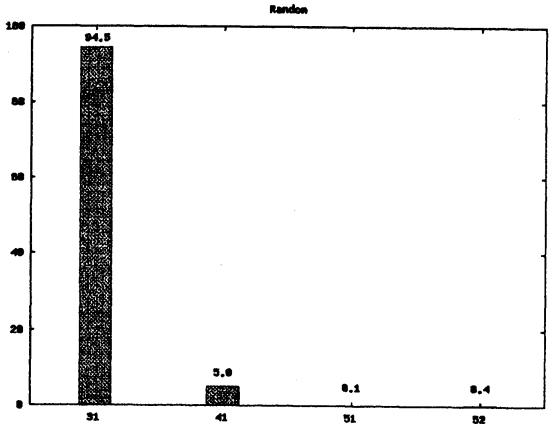

Fig. 14 Distribution of the knots 3_1, 4_1, 5_1 and 5_2 generated at random

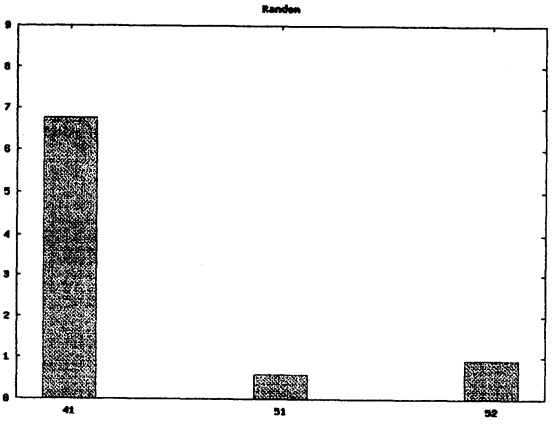

Fig. 15 Distribution of the knots 4_1, 5_1 and 5_2 generated at random

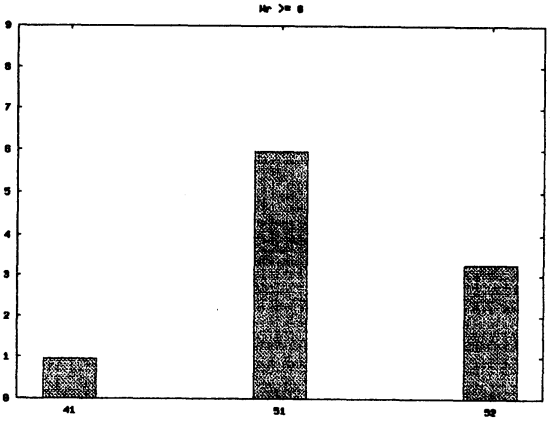

Fig. 16 Distribution of the knots 4_1, 5_1 and 5_2 after examining "the writhe value"

4.3 The results in case of giving lean of the total sign on all the edges by Jones polynomials

When we give only the plus sign to all the edges, the population of the mirror images is higher in the most knots group having the same crossing number (Fig.17). On the other hand, when we give only the minus sign, the inverse result happens (Fig.18).

Furthermore, let us see how the distribution changing, according to varying the total sign of the edges. We denote the absolute value of the total sign of all the edges as $|L|$. For $|L| \geq 0$, the distribution of the knots are decreasing with the change of the crossing number(Fig.19). However, for $|L| \geq 8$, the population of the knots with 4 crossings are rapidly decreased (Fig.20) and makes a hollow in the bar graph. This also has the same tendency as in the biochemical experiment done by Professors J. Arsuaga et. al. [2].

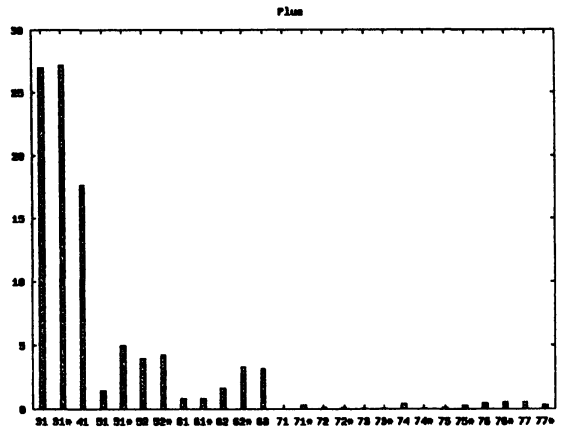

Fig. 17 Distribution with only plus for the sign of all edges

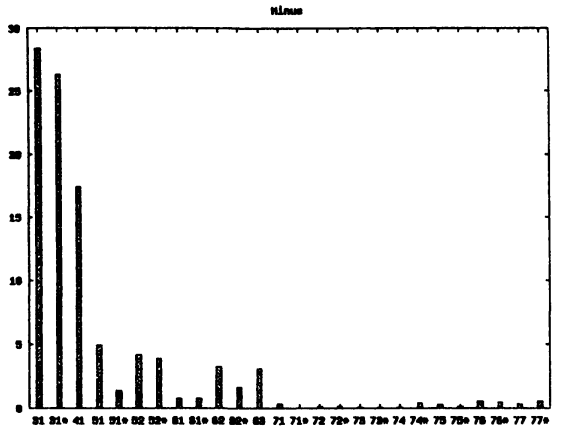

Fig. 18 Distribution with only minus for the sign of all edges

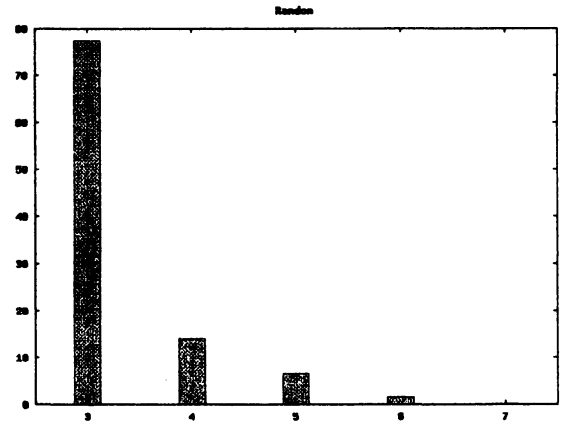

Fig. 19 Distribution of the knots for $|L| \geq 0$

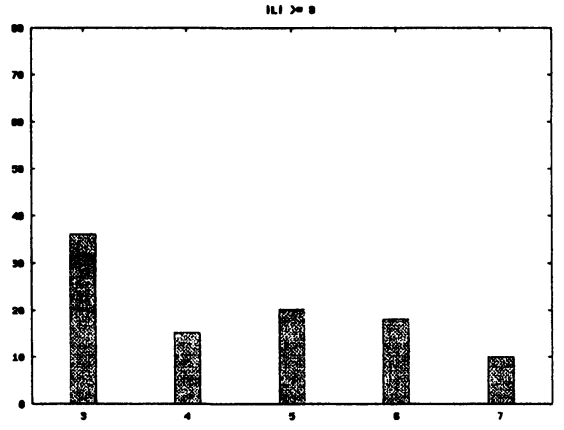

Fig. 20 Distribution of the knots for $|L| \geq 8$

4.4 Discussions

We conclude that DNA knots inside the λ phage are not generated randomly and they would have some bias related to "the writhe value". This means that the enzyme of nucleic acids have an operation which makes DNA knots having such a kind of leaning in the cell. This is a piece of important information about the total operation of the enzymes, and moreover, can be one of the keys by which we will solve more precisely a piece of the mechanism producing the cost-saving and resource-saving transportation system of the transcription-translation process in the cell. It is important partially because such kind of leaning may cause some obstacles for transcriptions, if the leaning is too heavy. Interestingly and marvelously, it is very important that the topological invariants extract global information about DNA knots mathematically. We hope that, if this direction of research is proceeded a little bit more and more, then it is possible that such purely mathematical notions are more and more useful for this kind of analysis and may get a more advanced application, because several examples are known there in which the global topological informations are remarkably important. Further, Jones polynomials can separate a kind of enantiomers (a kind of mirror image of knots) from the other, as we have done here. If a

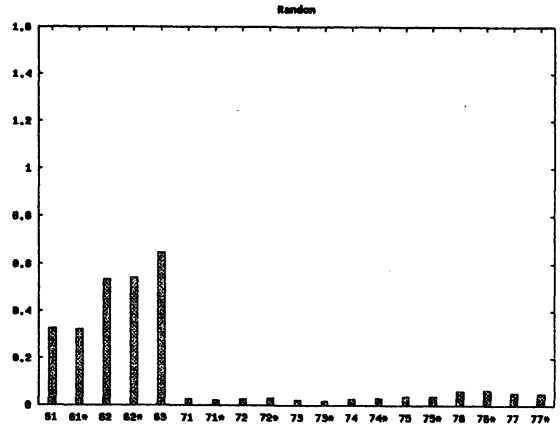

Fig. 21 Enlarged picture of the part of 6~7 crossings at the maximum 9 crossings

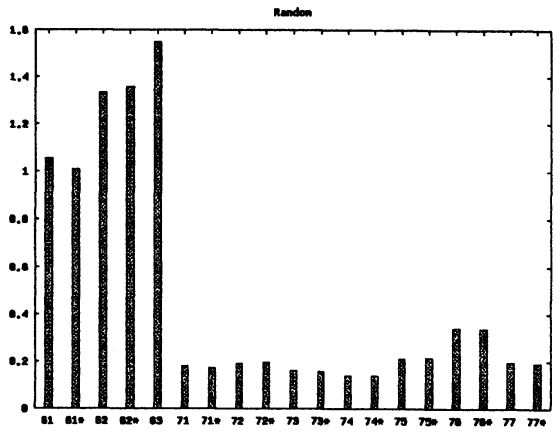

Fig. 22 Enlarged picture of the part of 6~7 crossings at the maximum 16 crossings

distribution separating such a kind of enantimers is obtained biochemically experimentally, this type of simulations may be compared with it to get something important about the enzyme's operation. For instance, our simulations indicates that both the one and the enantimer has the almost same ratio in the distribution (this is because the initial state is randomly given.) If there is a remarkable leaning in biochemical experiments under some conditions, then we will conjecture that the enzyme has this kind of operation, which is maybe interesting dependingly upon cases.

While the crossing number is very low, these simulations have been done well as a tendency by use of both topological invariants. Especially, we can give the leaning of "the writhe value" explicitly, when using Alexander polynomials, so that the result is clearer than in the simulation of Jones polynomials in this point of view. This is an advantage of Alexander polynomials, although Alexander polynomials has much lower resolution for knots than the Jones polynomials (See also [9]). For instance, Alexander polynomials cannot separate some nontrivial knots from the trivial one. But, as we choose the testing population randomly a hundred times to test it and take the average distribution among a hundred attempts, we can expect that the no-well separating objects are randomly scattered a kind of good way. Therefore, Alexander polynomials have a piece of reason to make a simulation with these, if the crossing number is less. Especially, it is used easily and takes a shorter time to spend than Jones polynomials.

On the other hands, we should not use Alexander polynomials in case of separating much more types of knots. It is well-known that there exists a nontirvial knot which is not separated from the trivial knot, and moreover it is known that there are several examples indicating low separating ability even in the case of knots with few numbers of crossing. In the actual biochemical experiments, the types of knots separated explicitly are from 3 to 7 (See also in Fig.13 of Professors J. Arsuaga et.al. [2]), but there seems to be much more types of knots and much more quantity of knots, which cannot be separated by this experiment and which are gotten into "the other" types. If we use Jones polynomials, instead, then theoretically we can separate them exactly in the case of the crossing number less than or equal to nine (This is easily ensured by computer calculations, and we have also done it already).

In Jones polynomials case, there is one difficult point how to give "the writhe value" explicitly in simulations. Here, although we cannot give the leaning of "the writhe value" explicitly, instead, we give the bias of the total sign of all the edges in the scheme by which Jones polynomials can be calculated. By this device, the knots are implicitly given the leaning, but we cannot measure how much it is explicitly. The resulting bar graph also has the same tendency as Alexander polynomials, but the hollow is smaller than in the case of Alexander polynomials. It seems that it is very difficult to give the leaning of "the writhe value" explicitly in the definition of the Jones polynomials so far. But in spite of such a disadvantage, the Jones polynomials are a very powerful tool, which is much higher resolution of knots

than the Alexander's ones. This can separate even the mirror images from the knots. If biochemical experiments get much stronger separation ability, then the Jones polynomials method is more and more useful.

This kind of simulations' results can be a kind of index of degree of operation of the enzyme of nucleic acid. This is because the operation is repeated a lot of times, and the leaning is more and more. But we must point out that the Jones polynomials cannot always separate two arbitrarily given knots. For example, it is known that 10_{129} has the exactly same polynomial as 8_8 (See [8]). Furthermore, it is known that there are infinitely many distinguished knots obtaining the same polynomial constructed concretely (See [6]). But the distribution itself can be affected very little, because the initial states are given randomly, so that it is expected that such kind of unseparable knots emergies randomly.

Now, we mension how the simulation results agree with the biochemical experiments' distribution of knot types. We make a glance at Fig. 13 to notice that 5_1 is much bigger comparably than 4_1 and 5_2. This suggest that it has a kind of leaning of "the Writhe value" of the knots. Our simulation ensured this property theoretically. Namely, in both ways of simulations, if "the writhe value" is let biased more and more, then the shape of distribution of the topological knot type is similar to the one of the experiment more and more (See Fig.s 14, 15, and 16).

When moving eyes to the finer part, we find some little differences of distributions between the biochemical experiments and our simulation results. For example, in the case of actual biochemical experiments, 4_1 is greater than 5_2 in Fig.13 , and on the other hand, the inverse is holds in the both cases of simulation done by us in Fig.14, although 5_1 greater than the others is agree with the actual biochemical experiments. In another point, 6_*, and 7_* are too low in the whole distribution. One reason may be that the maximum crossing number of initial state is too small. The initail crossing number is only 11 maximally so far in our simulation. If the initial crossing number is let much bigger, then the distribution is changing and expecting to be more similar shape of the bar graph. In fact, if the maximum values of initial crossing number is changed from 9 to 16, then ratios of distribution of 6_*, and 7_* are increasing (See Fig.21 and Fig.22). On the other hand, for the first disagreement, it is possible that there may be some unknown biological mechanism. In the actual biochemical experiment, the 4-crossing knot population is very low, although 3_1 is rather big. This is caused of some kind of biological actions. In the future, if we make the simulation methods be better to calculate knots having more crossings in the shorter time, we will make a finer interpretation and a precise prediction.

References

[1] G. Charvin, T. R. Strick, D. Bensimon and V. Croquette, *Topoisomerase IV Bends and Overtwists DNA upon Binding,* Biophysical Journal, **Vol.89,** (2005), 384–392.

[2] J. Arsuaga, M. Vazquez, P. McGuirk, S. Trigueros, D. W. Sumners and J. Roca, *DNA knots reveal a chiral organization of DNA in phage capsids,* PNAS, **Vol.102,** (2005), 9165–9169.

[3] M. D. Frank-Kamenetskii and A. V. Vologodskii, *Topological aspects of the physics of polymers: The theory and its biophysical applications,* Sov.Phys.-Usp.(Engl.Transl.), **Vol.24,** (1981), 679–696.

[4] K. Sekine, H. Imai and K. Imai, *Computation of the Jones Polynomial,* Transactions of JSIAM, **Vol.8,** (1998), 341–354.

[5] K. Sekine, H. Imai and S. Tani, *Computing the Tutte Polynomial of a Graph of Moderate Size,* Proceedings of the 6th International Symposium on Algorithms and Computation (ISAAC' 95), Lecture Notes in Computer Science, **Vol.1004,** (1995), 224–233.

[6] T. Kanenobu, *Infinitely many knots with the same polynomial,* Proc. Amer. Math. Soc. **Vol.97,** (1986), 158–161.

[7] J. D. Watson, T. A. Baker, S. P. Bell, A. Gann, M. Levine, and R. Losick, "Molecular biology of the gene" (the sixth edition), Pearson, Benjamin Cummings.

[8] W. B. R. Lickorish, "An Introduction to Knot Theory", Springer.

[9] T. Yoshino, and I. Ohnishi, *Applied analysis to DNA knot by topological invariants,* Research report of RIMS in Kyoto University, **1616,** (2008), 181 – 194.