KYUNGPOOK Math. J. 53(2013), 385-395 http://dx.doi.org/10.5666/KMJ.2013.53.3.385

## DNA and the SU(3) Invariant of Knots and Links

Myeong-Ju Jeong $^*$ 

Department of Mathematics and Computer Science, Korea Science Academy of KAIST, Busan 614-822, Republic of Korea e-mail: mjjeong@kaist.ac.kr

DAE GY HONG Department of Obstetrics and Gynecology, Kyungpook National University School of Medicine, Daegu, Korea e-mail: chssa0220@hanmail.net

ABSTRACT. To analyze the enzyme reaction on DNA knots and links, we study tangle embedding and the number of reaction. By using the quantum SU(3) invariant of knots and links we get a necessary condition for a tangle to be embedded in a knot or link. Moreover we give a relationship between the number of reactions and the changes of the value of quantum SU(3) invariant for the corresponding knots and links in a processive recombination.

### 1. Introduction

DNA is a nucleic acid containing the genetic instructions used in the growth and functioning of organisms. The main role of DNA molecules is the long-term storage of information. For example it contains the instructions needed to construct other components of cells, such as proteins and RNA molecules.

DNA consists of two long curves tangled and linked many times. Duplex DNA consists of two backbone strands linked each other. Each strand consists of sugar phosphate backbone with a nitrogenous base attached to each sugar. The DNA of most bacteria and viruses are circular. Although human DNA is linear, it is extremely long and tacked down to a protein scaffold at various points on the DNA. This periodic attachment endows human DNA with topological constraints similar to those for circular DNA. These topological constraints can interfere with vital metabolic cellular processes such as replication and transcription. Enzymes are required to solve these topological entanglement problems that arise through cellular

Key words and phrases: Knots, Links.



<sup>\*</sup> Corresponding Author.

Received May 10, 2012; accepted April 5, 2013.

<sup>2010</sup> Mathematics Subject Classification: 57M25, 57M27.

metabolism and replication. In this case, topoisomerases, which are enzymes that mediate the passage of one segment of DNA through an enzyme-bridged transient break in the backbone strands of another DNA segment, are responsible for unlinking the DNA. Other enzymes called recombinases break two DNA segments and interchange the ends, resulting in an exchange of genetic information.

Tangle calculus has been successfully used to study recombinases. The topological approach to enzymology is an experimental protocol in which the descriptive and analytical powers of topology and geometry are employed in an indirect effort to determine the enzyme mechanism and the structure of active enzyme-DNA complexes in vitro.

The packing, twisting, and topological constraints all taken together mean that topological entanglement poses serious functional problems for DNA. This entanglement would interfere with, and be exacerbated by, the vital life processes of replication, transcription, and recombination. For information retrieval and cell viability, some geometric and topological features must be introduced into the DNA, and others quickly removed. Some enzymes maintain proper geometry and topology by passing one strand of DNA through another by means of a transient enzymebridged break in one of the DNA strands. Other enzymes break the DNA apart and recombine the ends by exchanging them, a move performed by recombinases.

Recently, it has been found that topoisomerases viz. Topoisomerase III and IV also help in DNA recombination where the recombination is nonprocessive. The description and quantization of the three-dimensional structure of DNA and the changes in DNA structure due to the action of these enzymes requires extensive use of geometry and topology in molecular biology. This use of mathematics as an analytic tool is especially important because there is no experimental way to observe the dynamics of enzymatic action directly. The DNA knots and links of the reaction product DNA molecules are observed by gel electrophoresis and electron microscopy.

By observing the changes in geometry (supercoiling) and topology (knotting and linking) in DNA caused by an enzyme, the enzyme mechanism can be described and quantized. The topological approach to enzymology poses an interesting challenge for mathematicians as to how one can deduce enzyme mechanisms from the observed changes of DNA geometry and topology. This requires the construction of mathematical models for enzyme action and the use of these models to analyze the results of topological enzymology experiments. The entangled form of the product DNA knots and links contains information about the enzymes that made them. In addition to utility in the analysis of experimental results, the use of mathematical models forces all of the background assumptions about the biology to be carefully laid out.

We want to analyze the change of the topology of synatosome (enzyme and bound DNA), after an event of enzymatic action. DNA substrate molecule with its two recombination sites can be viewed as a union of circle embeddings. There are two movement of the strands, a global movement and local movement. By an ambient isotopy of  $\mathbb{R}^3$ , the recombination sites are juxtaposed inside a ball. Local movement happens within a ball. Two strands in the ball are cut at the recombination sites and then recombined.

The synatosome is modeled as a 2-string tangle denoted by E. Let  $O_b$  be the all part of the DNA that is bound to the enzyme or to the accessory proteins. Let P be the recombination sites. Then we have a tangle equation  $E = O_b + P$ . Let  $O_f$  be the tangle formed by the complement  $S^3 - E$  and  $O = O_f + O_b$ . Then the whole synaptic complex can be presented as  $N(O + P) = N(O_f + O_b + P)$ . The circular DNA is modeled as a knot or link that intersects the ball in 2 strands. The entangled form of the product DNA knots and links contains information about the enzymes that made them.

The recombinase action corresponds to a tangle surgery where the tangle P is changed to the tangle R.

Now one round of recombination give a system of tangle equations as following

$$\begin{cases} N(O+P) = N(O_f + O_b + P) = K_1, \\ N(O+R) = N(O_f + O_b + R) = K_2, \end{cases}$$

where  $O_f$ ,  $O_b$ , P and R are unknown. When we assume that the tangle P is changed to the tangle R by one round of recombination, it will be changed to  $nR = R + \cdots + R$  after n round of processive recombination. Thus we have a tangle equation  $N(O + nR) = N(O_f + O_b + nR) = K_n$ . See Figure 1.

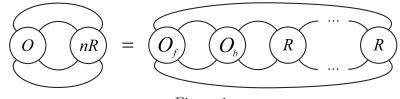


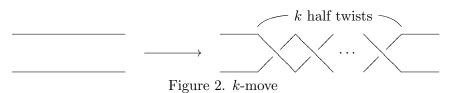
Figure 1.

To analyze enzyme reactions, rational tangle surgery are employed. In particular it is needed to see whether a given type of tangle is embedded in a given DNA knots or links.

Distances between knots/links involving 2-string rational tangle surgery have been used to model biological reactions involving enzymes which perform tangle surgery on DNA knots and links. Unlike topoisomerases which only use the strand passage unknotting operation, recombinases use different tangle surgeries depending upon the type of recombinase. Recombinases are enzymes which cut two segments of DNA and interchange the ends, therefore allowing a reshuffling of genes. Recombination can also cause mutations if the breaks occur within genes or regulatory sequences and is also used in the repair of damaged DNA. Recombination allows viruses to integrate into and excise out of host genomes and can be involved in the regulation of transcription by turning genes on or off by inverting segments of DNA. Biologists are interested in the mathematical question of determining all possible tangle surgeries that a particular recombinase may use. I. Darcy and D. W. Sumners defined a local move of knots and links, called an *R*-move for a tangle *R*. The *R*-move replace a trivial tangle in a link with the tangle *R*. Define the *R*-distance between two given links  $L_1$  and  $L_2$  to be the minimal number of the *R*-move needed to transform  $L_1$  to  $L_2$ . By using rational tangle surgery they gave formulas to determine the *R*-distance of two 4-plat knots or links is less than or equal to 1 for rational tangles *R* as following

**Theorem 1.1.** [1] Let  $R = \frac{t}{w} - tangle$ , (w, t) = 1 and ay - bx = 1. Then there exists s = tj + w for some integer j, such that the following are equivalent for  $|t| \ge 2$ . For  $t = \pm 1$ , (2) and (3) are equivalent and imply (1): (1)  $d(S(a, b), S(u, v)) \le 1$ . (2) If  $w \not\cong \pm 1 \mod t$ , S(u, v) = S(-tb + sa, ty - sx) or S(tx + sa, ty + sb). Else  $s = tpq \pm 1$ , (p,q) = 1, p > 0 and  $S(u, v) = S(-tp^2b + sa, tp^2y - sx)$ . (3) If  $w \not\cong \pm 1 \mod t$ , S(u, v) = N(a/(-b + aj) + (t/w)) or N(a/(x + aj) + (t/w)). Else  $s = tpq \pm 1$ , (p,q) = 1, and S(u, v) = N((ad - be)/(aq - bp) - (e/p) + t) where pd - qe = 1. In 4-plat form when  $s = tpq \pm 1$ ,  $S(u, v) = < c_n, \cdots, c_0 + a_1, \cdots, a_k, \pm t, -a_k, \cdots, -a_1 >$ , where  $S(a, b) = < c_0, \cdots, c_n >$ , n odd, and  $p/(-q) = [a_1, \cdots, a_k]$ , k odd.

The *k*-move on knots is an *R*-move where *R* is a *k*-tangle for an integer k as shown in Figure 2. If the two strands of the *k*-move is oriented parallel (opposite),



then the move is called the  $t_k$ -move ( $\bar{t}_k$ -move). Two links L and L' are said to be k-equivalent if one can transform L to L' by using the  $k^{\pm 1}$ -moves and two oriented links L and L' are said to be  $t_k$ -equivalent and  $\bar{t}_k$ -equivalent if one can transform L to L' by using the  $t_k^{\pm 1}$ -moves and  $\bar{t}_k^{\pm 1}$ -moves respectively.

The k-  $t_k$ - and  $\bar{t}_k$ - equivalences between two knots particularly for a trivial knot has been studied by many knot theorists.

In 1988, J. H. Przytycki ([9]) used the evaluation of the HOMFLYPT polynomial and the Kauffman polynomial to distinguish two  $t_k$ -nonequivalent knots and gave some information on how many  $t_k$ -moves are needed to go from one knot to the other if they are  $t_k$ -equivalent. He also studied  $t_3$ -equivalence and  $\bar{t}_4$ -equivalence of knots ([10]). In a different view point Kim, Park and the first author used some derivatives of the Conway polynomial and the Jones polynomial which are Vassiliev invariants to get some criterions for the  $t_k$ -equivalence and some results on how many  $t_k$ -moves are needed to go from one knot to the other if they are  $t_k$ -equivalent ([2]).

It is natural questions to ask whether two links are k-equivalent, or  $t_k$ -equivalent,

388

or  $\bar{t}_k$ -equivalent and if they are, how many moves are needed to go from one link to the other in each case. In particular, if k = 2 and the second knot is the unknot, it is the question of the classical unknotting number. Kim, Park and the first author studied twist moves of knots by using the second finite type invariant of knots. A twist move is an *R*-move where *R* is an *n*-tangle for an integer *n*.

**Theorem 1.2.** [2] Let K and K' be two  $t_{2i}$ -equivalent knots for a positive integer i. Let  $a_2(K)$  be the coefficient of the Conway polynomial of K. Then

$$\begin{cases} a_2(K) \equiv a_2(K') \quad (mod \ \frac{i}{2}) \quad if \ i \ is \ even \\ a_2(K) \equiv a_2(K') \quad (mod \ i \ ) \quad if \ i \ is \ odd. \end{cases}$$

For a positive integer k, the  $t_k$ -distance between two  $t_k$ -equivalent knots K and K', denoted by  $|K, K'|_{t_k}$ , is defined to be the minimum number of  $t_k^{\pm 1}$  moves needed to go from K to K'. For k > 2, the  $t_k$ -level distance between the  $t_k$ -equivalent knots K and K', denoted by  $|K, K'|_{t_k}^{lev}$ , is defined to be the number of  $t_k$  moves minus the number of  $t_K^{-1}$ -moves needed to go from K to K' ([9]).

**Theorem 1.3.** [2] Let K and K' be two  $t_{2i}$ -equivalent knots for an even positive integer i. Then

$$a_2(K') - a_2(K) \equiv \frac{i(i-1)}{2} |K, K'|_{t_{2i}}^{lev} \pmod{i}.$$

In Section 2, we introduce a module of oriented trivalent graph (k, k)-tangles modulo the Kuperberg's SU(3) skein relation. By using the Euler characteristic of sphere, we find basis for the modules when k = 1, 2. In Section 3, we find necessary condition for a (2,2)-tangle to be embedded in a knot or link by using representation of a tangle as a combination of generators in the module. We can apply this to study enzyme reaction on DNA knots and links.

#### 2. Embedding of a Tangle

D. A. Krebes introduced a necessary condition for a (2, 2)-tangle to be embedded in a link L by using the determinant of links. He showed that the greatest common divisor of the determinants of the numerator and the denominator of t is a divisor of the determinant of L if a (2, 2)-tangle t is embedded in a link L [3].

Silver and Williams ([11]) extended the Krebes result to virtual links by using the Fox coloring for a prime factor. Suppose that a tangle t is embedded in a virtual link L. They showed that n is a divisor of the determinant of L if n is a common prime factor of the determinants of the numerator and denominator of t. Krebes, Silver and Williams developed these results by using the virtual Temperley-Lieb algebra [4].

Now we consider the SU(3) invariant introduced by G. Kuperberg ([5]). Let F be an oriented surface. We consider oriented trivalent graphs possibly with loops with no vertices. Assume that at each vertex the three edges are oriented in the

same way all inner or all outer. A diagram in F is an oriented trivalent graph immersed in F such that the singular points are only transverse double points, to each of which over and under crossing information is associated; this means that locally it is the diagram of an oriented link or a trivalent graph. Throughout the rest of this paper diagrams are considered up to isotopy in F. From now on we assume that the surface F is the 2-sphere  $S^2$ .

We generalize oriented (k, k)-tangles to oriented trivalent graph (k, k)-tangles by allowing them to have trivalent vertices. The orientations of the three edges adjacent to a vertex of an oriented trivalent graph tangle are assumed to be all inner or outer. Similarly to the closure of a braid we define the *closure*  $\overline{T}$  of an oriented trivalent graph (k, k)-tangle T. See Figure 3.

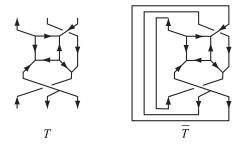


Figure 3. An oriented trivalent graph (3,3)-tangle and its closure.

Kuperberg introduced a linear skein theory for SU(3). We define the SU(3) skein module of oriented surfaces as follows:

$$K1: \bigcirc UD = (A^6 + 1 + A^{-6})D$$

$$K 2: = A^{2} - A^{-1}$$

$$K 3: = A^{-2} - A$$

$$K 4: = (A^{3} + A^{-3})$$

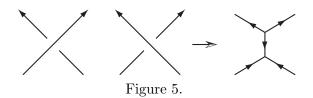
$$K 5: = + (A^{3} + A^{-3})$$

Figure 4. The Kuperberg skein relations.

We assign 1 to the trivial knot diagram. Then it becomes a regular isotopy invariant of oriented links. Denote the polynomial obtained by applying the Kuperberg skein relations to an oriented link L by [L] ([5]). We get an invariant of oriented link L by normalizing it by multiplying  $(A^8)^{-w(L)}$ , where w(L) is the writhe of L([5]). We define  $q(L) = (A^8)^{-w(L)}[L]$ .

For an oriented (k, k)-tangle T, we write the family of oriented trivalent graph (k, k)-tangles with the same boundary orientation with T by  $\mathfrak{T}_T$ . Let R be the ring  $\mathbb{Z}[A, A^{-1}]$  of integral Laurent polynomial ring with indeterminate A. We denote the free R-module generated by  $\mathfrak{T}_T$  by  $R\mathfrak{T}_T$ . Let  $R\mathfrak{T}_T/\sim$  be the quotient module of  $R\mathfrak{T}_T$  by the Kuperberg skein relations.

**Lemma 2.1.** Let T be an oriented graph (k, k)-tangle with n crossings. Let D be the diagram obtained from the closure  $\overline{T}$  of T when we change each crossing as Figure .



Assume that D forms a triangulation of the two sphere  $S^2$ . Let  $\alpha$  and  $\beta$  be the number of bigons and 4-gons in D respectively. Then we have an inequality  $2\alpha + \beta \geq 6$ .

*Proof.* Let F(D), E(D) and V(D) be the number of faces, edges and vertices of D respectively. Since the Euler characteristic  $\chi(S^2) = 2$ , F(D) - E(D) + V(D) = 2 and F(D) = E(D) + 2 - V(D). For each face F of D, we denote the number of edges in F by E(F). Since  $\sum_{F: \text{ face of } D} E(F) = 2E(D)$ , we have  $2\alpha + 4\beta + 6(E(D) + 2 - V(D) - \alpha - \beta) \leq 2E(D)$ . So  $2E(D) + 6 - 3V(D) \leq 2\alpha + \beta$ . Since E(D): V(D) = 3: 2, we see that the inequality  $2\alpha + \beta \geq 6$  holds.

In particular if T is an oriented (k, k)-tangle with n crossings, then we see that  $\alpha + \beta \leq n + 2$  by considering F(D).

**Lemma 2.2.** If T is an oriented (1, 1)-tangle then the quotient module  $RT_T / \sim$  is generated by the trivial tangle without any crossing.

*Proof.* For an oriented trivalent graph (1, 1)-tangle in  $\mathcal{T}_T$ , we have a sum of oriented trivalent graphs with coefficients in the ring R by applying the K2 and K3 relations for each crossing of S. Now it is enough to show that an oriented trivalent graph (1, 1)-tangle S without any crossing is generated by the trivial oriented (1, 1)-tangle. We use mathematical induction argument on the sum n(S) + V(S) of the number of

connected components n(S) of S and the number of trivalent vertices V(S) of S. If n(S) + V(S) = 1 then S is the trivial (1, 1)-tangle. Assume that n(S) + V(S) > 1. If V(S) = 0 then S is union of n(S) - 2 circle components and the trivial tangle. suppose that V(S) > 1. Let S' be a component of S with a trivalent vertex. Then S' or  $\overline{S'}$  forms a triangulation of the two sphere  $S^2$ . By applying Lemma 2.1, we see that there is a bigon or 4-gon in S'. When we apply K4 or K5 to the bigon or 4-gon, we can reduce the sum of the number of connected components and vertices.  $\Box$ 

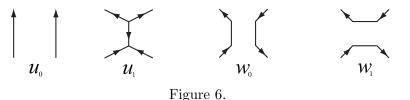
The skein module of the space of (2, 2)-tangles are generated by the following basis depending on the boundary orientation of the tangle.

**Lemma 2.3.** Let T be an oriented (2,2)-tangle with parallel boundary orientation as  $u_0$  as shown in Figure . Then the Kuperberg skein module  $RT_T / \sim$  is generated by the tangles  $u_0$  and  $u_1$ .

Proof. Since any crossing can be spliced by K2 and K3 relations, we consider an oriented trivalent graph (2, 2)-tangle S. We will use mathematical induction on the sum n(S) + V(S) of the number of connected components n(S) of S and the number of trivalent vertices V(S) of S again. If V(S) = 0 then S is generated by  $u_0$ . Suppose that V(S) > 1 and S' is a connected component of S with a trivalent vertex. Since S' or  $\overline{S'}$  forms a triangulation of the two sphere  $S^2$ , by Lemma  $2\alpha + \beta \geq 6$ , where  $\alpha$  and  $\beta$  are the numbers of bigons and 4-gons in S' respectively. If  $\beta$  is 0 then  $\alpha \geq 3$  and S' has a bigon or equals to  $u_1$ . If  $\beta$  is greater than 0 then the sum of the number of components of S and the number of vertices of S can be reduced by applying K5.

Similarly to the argument of the proof of Lemma 2.3, we can show the following

**Lemma 2.4.** Let T be an oriented (2, 2)-tangle with opposite boundary orientation as  $w_0$  as shown in Figure . Then the Kuperberg skein module  $RT_T / \sim$  is generated by the tangles  $w_0$  and  $w_1$ .



### 3. Tangle Embedding and the SU(3) Invariant

In this section we give necessary conditions for a tangle to be embedded in a given oriented knot or link by using the SU(3) invariant based on the Kuperberg's

skein module. We can apply the result to analyze the enzyme reaction.

**Lemma 3.1.** Let T be a (2,2)-tangle with a parallel boundary orientation as  $u_0$ . Assume that  $T = au_0 + bu_1$  for  $a, b \in \mathbb{Z}[A, A^{-1}]$  in the SU(3) skein module  $R\mathfrak{T}_T/\sim$ . Let d be a common divisor of a and b. If T embeds to a link L then d divides q(L).

*Proof.* If T embeds to a link L then there exists a (2, 2)-tangle S such that  $L = \overline{ST}$ . By Lemma 2.3,  $S = a'u_0 + b'u_1$  for some  $a, b \in \mathbb{Z}[A, A^{-1}]$ . Then

$$\begin{split} [L] =& aa'[\overline{u_0u_0}] + ab'[\overline{u_1u_0}] + a'b[\overline{u_0u_1}] + bb'[\overline{u_1u_1}] \\ =& aa'[\overline{u_0}] + \left(ab' + a'b + bb'(A^3 + A^{-3})\right)[\overline{u_1}]. \end{split}$$

Therefore if d|a and d|b, then d|[L] and d|q(L).

Similarly by applying Lemma 2.4 to a (2, 2)-tangle with an opposite boundary orientation, we get

**Theorem 3.2.** Let T be a (2,2)-tangle with an opposite boundary orientation as  $w_0$ . Assume that  $T = aw_0 + bw_1$  for  $a', b' \in \mathbb{Z}[A, A^{-1}]$  in the SU(3) skein module  $RT_T / \sim$ . Let d be a common divisor of a and b. If T embeds to a link L then d divides q(L).

Assume that one round of recombination gives tangle equations as following

$$\begin{cases} N(O+P) = N(O_f + O_b + P) = K_1, \\ N(O+R) = N(O_f + O_b + R) = K_2, \end{cases}$$

If we see the two knots or links  $K_1$  and  $K_2$ , then by applying Theorem and Theorem, we may find candidates for the tangle O. For example if  $O = au_0 + bu_1$ , then the greatest common divisor of a and b should divide  $q(K_1)$  and  $q(K_2)$ . Now we consider processive recombination. Since the orientation of a the knot or link N(O + nR) is compatible, we see that the boundary orientation of the tangle R is opposite.

**Theorem 3.3.** Assume that  $N(O + nR) = K_n$  and  $R = aw_0 + bw_1$ . If d is a common divisor of a and b, then  $d^n$  is a divisor of  $q(K_n)$ .

*Proof.* Let  $\alpha = (A^6 + 1 + A^{-6})$ . The sum of  $w_0$  and  $w_0$  in  $RT_T / \sim$  is illustrated in Figure . If  $R = aw_0 + bw_1$ , then

$$2R = a^{2} \alpha w_{0} + abw_{0} + abw_{0} + b^{2} w_{1}$$
  
=  $(a^{2} \alpha + 2ab)w_{0} + b^{2} w_{1}$   
=  $\frac{1}{\alpha} ((a\alpha + b)^{2} - b^{2}) w_{0} + b^{2} w_{1}.$ 

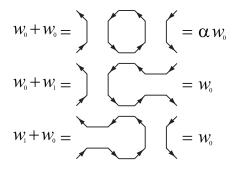


Figure 7.

Inductively, we can see that  $nR = \frac{1}{\alpha} ((a\alpha + b)^n - b^n) w_0 + b^n w_1$ . If *d* is a common divisor of *a* and *b* then  $d^n$  is a common divisor of  $\frac{1}{\alpha} ((a\alpha + b)^n - b^n)$  and  $b^n$ . Therefore  $d^n$  is a divisor of  $[K_n]$  and  $q(K_n)$ .

In a tangle model, The circular DNA substrates and products are regarded as knots or links. Site-specific recombination arising from enzyme reaction affects the topology of the knots or links. By using this result, we may guess how many enzyme reaction happened and the topological change by using the types of the tangle *R*. Then we may analyze enzymatic mechanisms based on the experimental data better and get a map to manipulate DNA topologically.

Acknowledgement. The authors would like to thank the referee for careful reading and kind comments.

# References

- I. K. Darcy and D. W. Sumners, *Ratioal tangle distance on knots and links*, Math. Proc. Camb. Phil. Soc., 128, 2000, 497-510.
- [2] M.-J. Jeong, Eun-Jin Kim and C.-Y. Park, Twist moves and Vassiliev invariants, to appear in J. of Knot Theory and Its Ramifications.
- [3] D. A. Krebes, An obstruction to embedding 4-tangles in links, J. Knot Theory Ramif., 8(1999), 321-352.
- [4] D. A. Krebes, D. S. Silver and S. G. Williams, Persistent invariants of tangles, Preprint.
- [5] G. Kuperberg, The quantum  $G_2$  link invariant, Preprint.

- [6] J. C. Misra, S. Mukherjee and A. K. Das, A mathematical model for enzymatic Action on DNA knots and links, Mathematical and Computer Modelling, 39, 2004, 1423-1430.
- Y. Ohyama, A new numerical invariant of knots induced from regular diagrams, Topology and Its Applications 37(1990), 249-255.
- [8] Y. Ohyama, Vassiliev invariants and similarity of knots, Proc. Amer. Math. Soc. 123(1993), 287-291.
- [9] J. H. Przytycki,  $t_k$  moves on links, Braids, Comtemp. Math. 78, Amer. Math. Soc., 1988, 615-656.
- [10] J. H. Przytycki,  $t_3$ ,  $\bar{t}_4$  moves conjecture for oriented links with matched diagrams, Math. Proc. Cambridge Phil. Soc., 108, 1990, 55-61.
- [11] D. S. Silver and S. G. Williams, Virtual tangles and a theorem of Krebes, J. Knot Theory Ramifications 8(1999), 941-945.