A TOPOLOGICAL CHARACTERIZATION OF THE KNOTTED AND LINKED PRODUCTS ARISING FROM SITE-SPECIFIC RECOMBINATION ON $T(2,n)\#C(2,r)$ DNA SUBSTRATES

JORDAN KATZ
MENTOR: DIANA HUBBARD

Abstract. A mathematical knot is a simple closed curve in 3-space. We model circular DNA as a knot to predict and classify the possible knotted products which can arise as a result of site-specific recombination. We are particularly interested in DNA substrates of the form $T(2,n)\#C(2,r)$. We show that, given some reasonable biological assumptions, all of the possible products are contained in one of two families.

1. Introduction

Circular DNA is a double-helix DNA molecule that forms a closed loop so that it has no free ends. It is found in bacterial chromosomes, mitochondria, and chloroplasts, and it is used by scientists in molecular biology and genome engineering experiments. In this paper we model the central axis of a molecule of circular DNA as a mathematical knot. A *knot* is a simple, closed curve in 3-space; simple meaning non-self-intersecting and closed meaning having its two loose ends glued together. Closely related to a knot is something called a *link*, or a collection of knots with mutual entanglements.

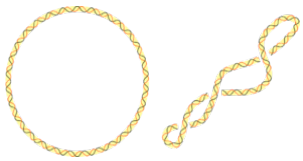


Figure 1. Circular DNA
Image: https://en.wikipedia.org/wiki/Topoisomer

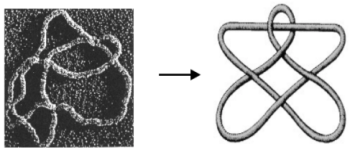


Figure 2. DNA as a knot
Image: https://julianpark.me/posts/topology

Site-specific recombination is a naturally-occurring biological process that alters the structure of a DNA molecule. In this process, two recombinase *enzymes* attach to a DNA *substrate* at two different *sites* and bring them close together. We refer to this region of space where the sites and enzymes interact as the *recombinase complex*. The sites are then cleaved, exchanged, and resealed in a way that is determined by which recombinase subfamily - *serine* or *tyrosine* - our enzyme belongs to. We refer to the post-recombinant DNA as the *product*.

The underlying mechanism of this process is not well understood [FGL$^+$13]. To solve this problem, the molecular biologists let the enzymes act on a circular DNA substrate [FGL$^+$13]. The product then, a knotted or linked strand of circular DNA, is used to deduce the underlying mechanism which could have caused the knot or link [BV11]. (If the DNA were to have loose ends, one could simply unwind it post-recombination losing all of the relevant information about the underlying reaction.)

The ability to identify and categorize these possible knotted and linked products would help the biologist's understanding of the underlying reaction [BF07, FGL$^+$13]. This is precisely the aim of this paper. Current methods for doing this, such as electron microscopy and gel electrophoresis, have proven to be difficult [FGL$^+$13]. Instead, we develop a topological model similar to that seen in [FGL$^+$13, BF07, BV11]. We take a certain set of biologically reasonable assumptions and formulate them in a mathematically precise way so that we can deduce such products given the configuration of our initial substrate.

Specifically we are interested in the case where our substrate is represented by a knot of the form $T(2,n)\#C(2,r)$; that is, the connected sum of a torus knot $T(2,n)$ and a twist knot $C(2,r)$. The *connected sum* operation $\#$ takes in two knots and outputs a single one; it can be thought of as "knot addition."

In general the operation is not well-defined, meaning that the output for a given input is not necessarily unique. Since $T(2,n)$ and $C(2,r)$ are both invertible, however, we do not have a problem [Ada94]. The connected sum $T(2,n)\#C(2,r)$ is indeed well-defined, as are both $K_1\#T(2,n)$ and $K_1\#C(2,r)$ for some arbitrary knot K_1. We denote some integer number of crossings by an ellipsis.

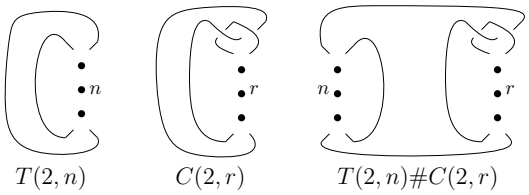


FIGURE 3. Substrates of interest

Previous work has been done on simpler substrates. [BF07] determined the knotted and linked products of site-specific recombination on unknotted, unlinked, and $T(2,n)$ substrates, [BV11] on $C(2,r)$ substrates, and [FGL$^+$13] on $T(2,n)\#T(2,m)$ substrates. We choose to focus on $T(2,n)\#C(2,r)$ substrates since these are one of the the next most trivial products. It only takes two rounds of recombination on an unknotted substrate to get to a product of the form $T(2,n)\#C(2,r)$.

Throughout the paper we will refer to the knot representing our DNA substrate as J, and the two recombination sites as α_1 and α_2, which are two small arcs living on J. Further, we will model the recombinase complex as a topological ball $B \subset \mathbb{R}^3$.

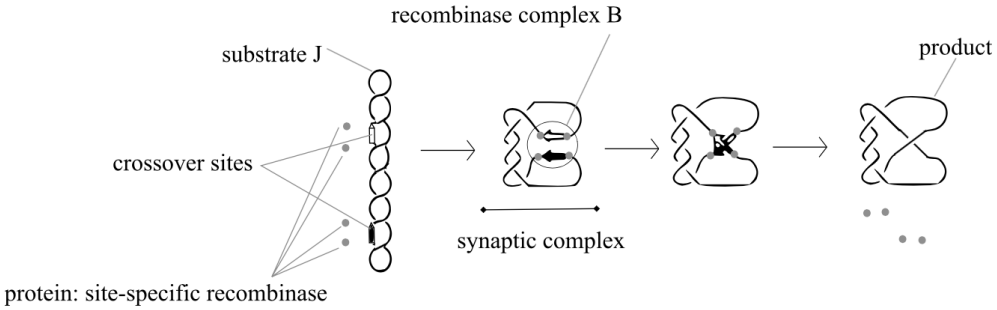


FIGURE 4. An example of site-specific recombination
Image: [BV11]

Before we derive the possible knotted and linked products, we lay out some mathematical assumptions based on what is experimentally observed to occur [BF07]. The first assumption is about the "niceness" of the recombinase complex, the second assumption is about the "niceness" of the substrate outside of the recombinase complex, and the third assumption is about how the recombinase enzyme alters the substrate within the recombinase complex.

2. Assumptions

Let B be a ball such that $B \cap J = \alpha_1 \cup \alpha_2$.

Definition 1. A *projection* is the image of a mapping $P : \mathbb{R}^3 \mapsto \mathbb{R}^2$ such that $P^2 = P$.

Remark. Intuitively, a projection is a 2-dimensional shadow cast by a 3-dimensional object.

Assumption 1. *There is a projection P_0 of $B \cap J$ such that α_1 and α_2 do not cross and neither α_1 nor α_2 self-cross.*

The following elaborates on an idea considered in [FGL$^+$13]. Let R be a surface in $\mathbb{R}^3$ such that $\partial R = J$, and let S be a thrice-punctured planar surface in $\mathbb{R}^2 \subset \mathbb{R}^3$ decorated with finitely many arcs whose boundaries lie in the boundary of S (see Figure 6). We denote an arbitrary number of parallel arcs grouped together by an ellipsis.

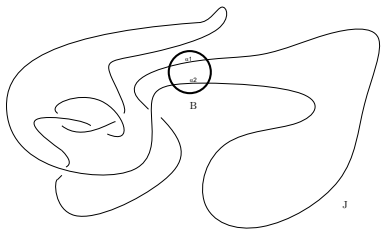


FIGURE 5. An example of $B \cup J$ satisfying Assumption 1

Caution. These decorating arcs are different from the arcs α_1 and α_2 we use to represent the two recombination sites.

Remark. We could annotate each decorating arc with a $-$ or $+$ sign to denote the direction of the twisting. In particular, the two arcs representing the clasp on the $C(2, r)$ component must both be of the same sign. However, for sake of simplicity in our drawings, we generally omit this information.

We view S in the projection P, which can be thought of as what is seen in the plane of this page.

Definition 2 ([FGL$^+$13])**.** We say that R *is represented by* S when R can be obtained from S by replacing each of the decorating arcs in S by a half-twisted band contained in the neighborhood of the arc.

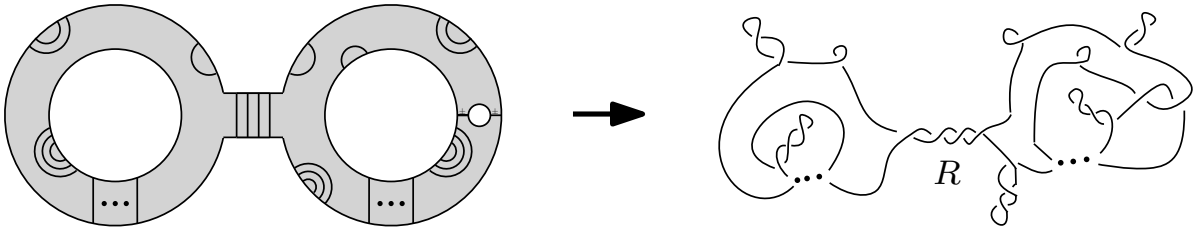


FIGURE 6. The decorated planar S represents the surface R

Definition 3. For two knots K_1, K_2, an *ambient isotopy* from K_1 to K_2 is a mapping which continuously, and without cutting or self-intersecting, deforms K_1 into K_2.

Assumption 2. *There is an ambient isotopy, pointwise fixing B, from J to the boundary of a surface R which can be represented by S.*

We now isotope J accordingly and continue to view S in the projection P.

Remark. Since P and P_0 may be different, it is possible that α_1 and α_2 admit a single crossing in the projection P while maintaining zero crossings in the projection P_0. The following figure illustrates the possible forms that $B \cap J$ can take in the projection P.

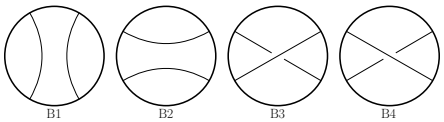


FIGURE 7. Pre-recombinant $B \cap J$ in the projection P

Remark. $B1$ and $B2$ are prima-facie equivalent by a 90° rotation, as are $B3$ and $B4$. We list them separately for the following reason: If we think of P as what is seen on this page, we are implicitly choosing an orientation by which we look at the page, and thus lose the rotational symmetry.

Without loss of generality we henceforth conflate R and S. We refer to "B interacting with R" as "B interacting with S" in order to make the proof more straightforward.

Assumption 3. *Site-specific recombination only affects the substrate J inside of the recombinase complex B. In particular,*

(1) *After $|k|$ rounds of processive recombination, a **serine** recombinase adds a row of k identical crossings between α_1 and α_2.*
(2) *A **tyrosine** recombinase replaces α_1 and α_2 with arcs that cross at most twice.*

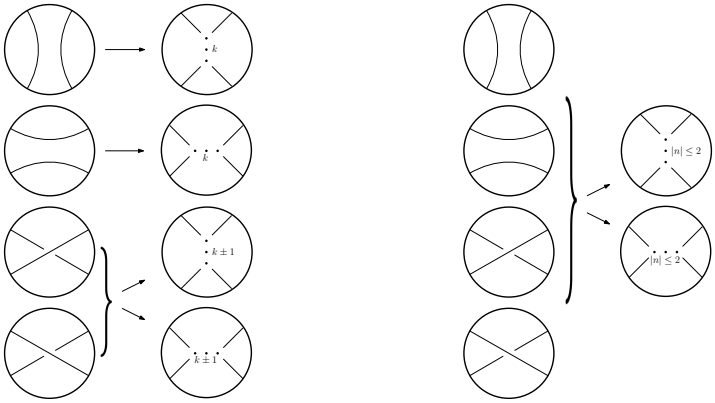


FIGURE 8. Post-recombinant forms of $B \cap J$ in the projection P, as mediated by serine (left) and tyrosine (right)

3. Lemmas

Lemma 1. *$B \cap S$ is either a (possibly twisted) strip or two disjoint disks.*

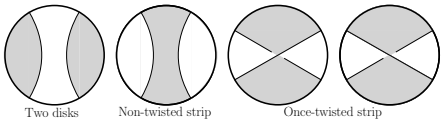


FIGURE 9. The different forms that $B \cap S$ can take, up to rotation

Proof. If pre-recombinant $B \cap J$ is of the form $B1$ or $B2$, then $B \cap S$ can be a strip or two disks, depending on how one chooses to shade it. If pre-recombinant $B \cap J$ is of the form $B3$ or $B4$, then $B \cap S$ can only be a once-twisted strip. □

Remark. When we enumerate all possible configurations of $B \cup S$, we assume (unless otherwise specified) that in all cases where $B \cap S$ is a strip, the strip may be once-twisted. Doing so allows us to restrict the number of drawings to a minimum without loss of generality.

Definition 4. A decorating arc *separates* $B \cup S$ if cutting along it divides $B \cup S$ into two disjoint pieces.

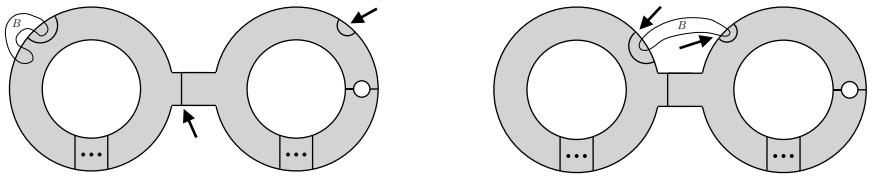


FIGURE 10. Two decorating arcs which each separate $B \cup S$ (left) and two decorating arcs which *together* separate $B \cup S$ (right)

Lemma 2 ([FGL$^+$13]). *No decorating arc on $S - B$ separates $B \cup S$.*

Lemma 3 ([FGL$^+$13]). *Any two decorating arcs on $S - B$ which together separate $B \cup S$ belongs to a single group of parallel arcs.*

Remark. We utilize Lemmas 2 and 3 to make our drawings simpler.

4. Main Proof Strategy

For the remainder of the paper we will assume that our substrate and enzyme satisfy Assumptions 1, 2, and 3. To determine all possible products, we enumerate all possible ways that the ball B can interact with the surface S without violating any of the assumptions. Doing so will provide a list of "legal" configurations of $B \cup J$ from which we can then apply Assumption 3 to obtain the desired result.

To help us enumerate, we break up the possible configurations into three exhaustive cases. Case 1 covers when α_1 and α_2 are both on the $T(2, n)$ component, Case 2 covers when α_1 and α_2 are both on the $C(2, r)$ component, and Case 3 covers when one of α_1 and α_2 is on the $T(2, n)$ component and the other is on the $C(2, r)$ component. Each case can be further broken down into two exhaustive sub-cases, depending upon whether $B \cap S$ is a (possibly twisted) strip or a set of two disjoint disks. We implicitly invoke Lemmas 2 and 3 so that our drawings can be as simple as possible.

Once we enumerate these, we isotope out duplicates by applying moves M1, M2, and M3. This ultimately provides us with a minimal set of configurations with maximal predictive power, from which we are free to apply Assumption 3 to determine the post-recombinant form of our substrate.

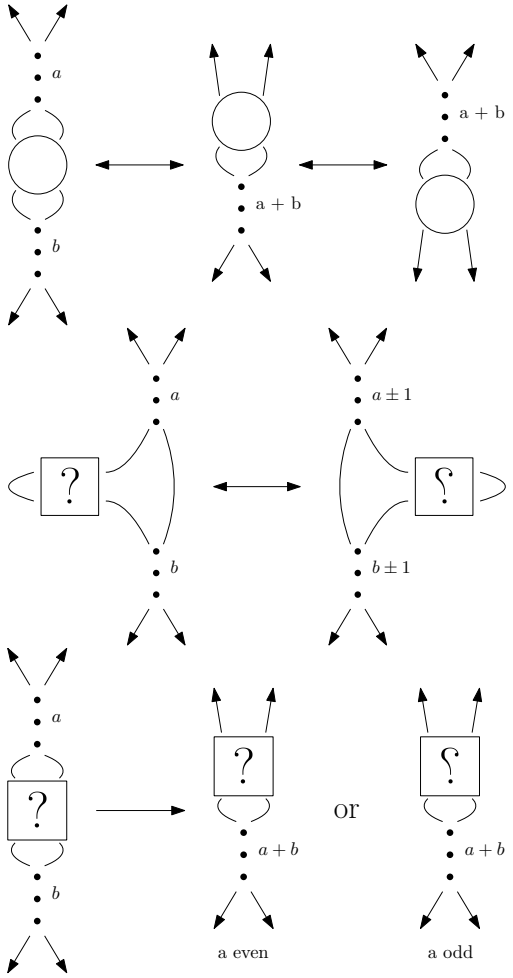


Figure 11. Move 1 (M1), Move 2 (M2), and Move 3 (M3)

Remark. M1 can be visualized by rotating the ball (used to represent B) by multiples of π around the vertical axis of the figure. M2 can be visualized by "lifting" the box and attached strand out of the page and then swinging in front of the rest of the figure (or equivalently by "pushing" the box and attached strand into the page and then swinging it behind the rest of the figure), while keeping the rest of the figure glued to the page. M3 can be visualized by rotating the box by multiples of π around the vertical axis of the figure.

5. Dividing into cases

Here we consider three cases depending upon the locations of α_1 and α_2 in ∂S. For each of the cases, we use a combinatorial argument by considering all possible locations for α_1 and α_2 except for those which violate Assumption 2. We number sections of ∂S to make this easier. By "14," for example, we mean the instance where α_1 or α_2 lives in $1 \subset \partial S$ and the other lives in $4 \subset \partial S$. When α_1 or α_2 lives in an ellipsis section, we break up the ellipsis into two sets of ellipses and place α_1 or α_2 between them, as illustrated below.

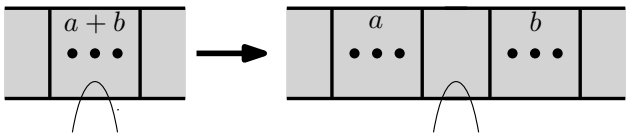
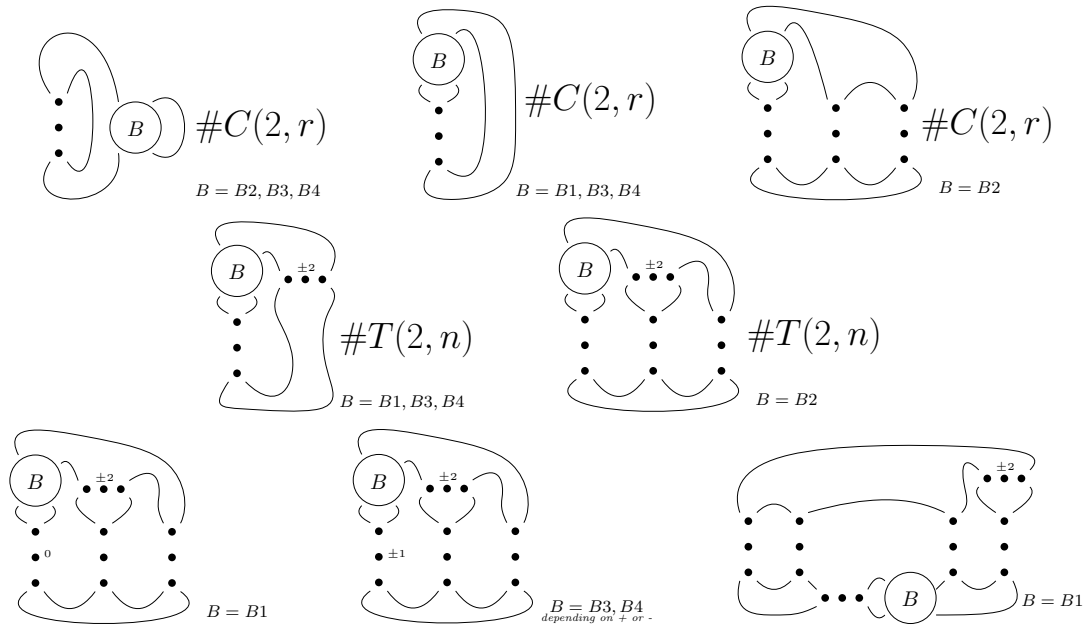


FIGURE 12. What it means for α_1 or α_2 to live in an ellipsis section

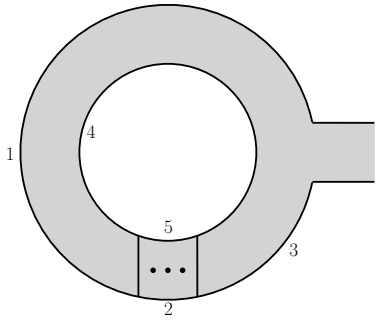
After this enumeration, we eliminate isotopically equivalent instances using our three moves, or instances which are limiting cases of others. Limiting cases can only occur when one of α_1 and α_2 lives in an ellipsis section. By *limiting case*, we mean a case where a or b (as in the figure above) evaluates to 0. For example, in Case 1 when $B \cap S$ is two disks, 13 is a limiting case of 12. All information about 13 is contained in that about 12, which is why it is safe to eliminate. In Case 2, some instances are repeated to make the logic consistent despite them already being covered in Case 1. Additionally, some Case 2 instances can be eliminated by continuously sliding (via isotopy) α_1 or α_2 out of the clasp, as in, for example, 44.

Theorem 1. *All possible forms that $B \cup J$ can take pre-recombination are illustrated below:*

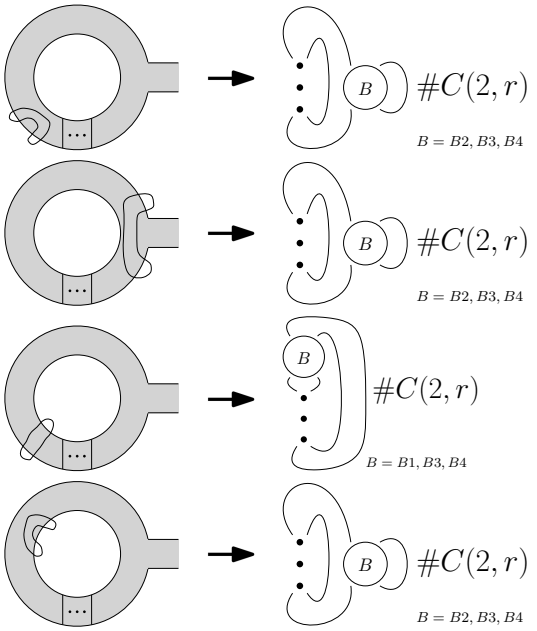


Proof. We break $B \cup S$ into cases.

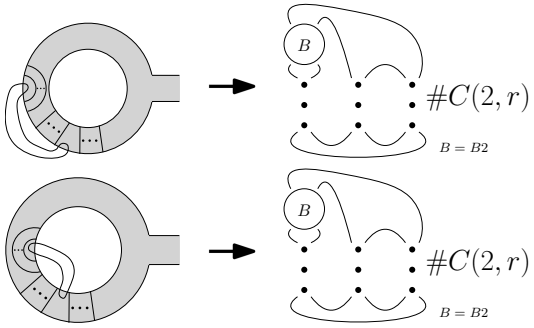
Case 1: α_1 and α_2 are both on the $T(2,n)$ component.



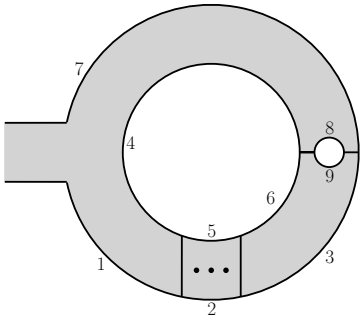
Subcase 1.1: $B \cap S$ is a strip. The only possible configurations which don't violate Assumption 2 are 11, 13, 14, 22, 25, 33, 34, 44, and 55. We can simplify this list further: 33 is a limiting case of 22. We can apply M1 to isotope 25 and 34 to 14. We can apply M2 to isotope 55 to 22. We can apply M3 to isotope 22 to either 11 or 14. This leaves us with 11, 13, 14, and 44 as illustrated below.



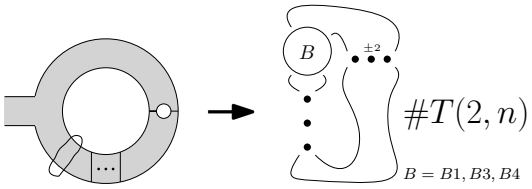
Subcase 1.2: $B \cap S$ is two disks. The only possible configurations which don't violate Assumption 2 are 11, 12, 13, 22, 23, 33, 44, 45, and 55. We can simplify this list further: 11 and 13 are limiting cases of 12. 33 is a limiting case of 23. 44 is a limiting case of 45. We can apply M2 to isotope 55 to 22. We can apply M3 to isotope both 22 and 23 to either 12 or 45. This leaves us with 12 and 45 as illustrated below.



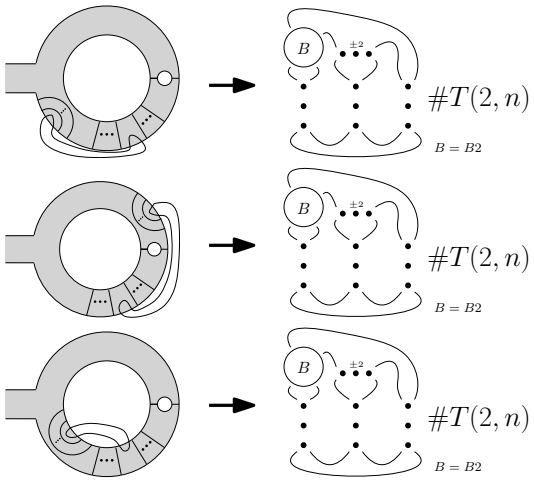
Case 2: α_1 and α_2 are both on the $C(2,r)$ component.



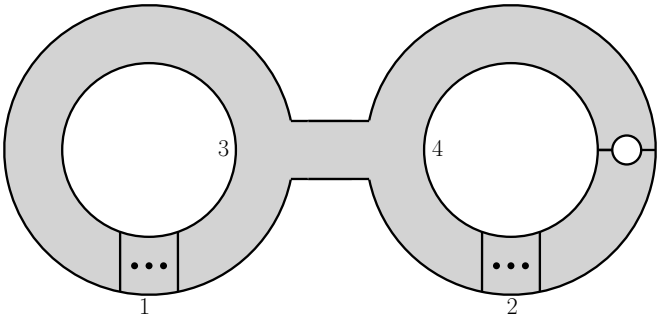
Subcase 2.1: $B \cap S$ is a strip. The only possible configurations which don't violate Assumption 2 are 11, 14, 17, 22, 25, 33, 36, 44, 47, 55, 66, 77, 88, and 99. We can simplify this list further: 11, 17, and 77 are already covered in Case 1. 33 is a limiting case of 22. We can apply M1 to isotope 25, 36, and 47 to 14. We can apply M2 to isotope 55 to 22. We can apply M3 to isotope 22 to either 11 or 44. We can isotope both 44 and 99 to 77, and both 66 and 88 to 33, by continuously moving both α_1 and α_2 through the clasp. This leaves us with 14 as illustrated below.



Subcase 2.2: $B \cap S$ is two disks. Any arc living in 8 or 9 can be continuously moved through the clasp to 3 or 7 respectively, so we don't list them. (More justification for this in [BV11].) Then, the only possible configurations which don't violate Assumption 2 are 11, 12, 13, 17, 22, 23, 27, 33, 37, 77, 44, 45, 46, 55, 56, and 66. We can simplify this list further: 11, 17, and 77 are already covered in Case 1. 13 is a limiting case of 12. 33 is a limiting case of 23. 37 is a limiting case of 27. 46 is a limiting case of 45. We can apply M2 to isotope 55 to 22. We can apply M3 to isotope 22 to either 11 or 44, and both 23 and 56 to either 12 or 45. We can isotope 44 to 77 and 66 to 33 by continuously moving both α_1 and α_2 through the clasp. This leaves us with 12, 27, and 45 as illustrated below.

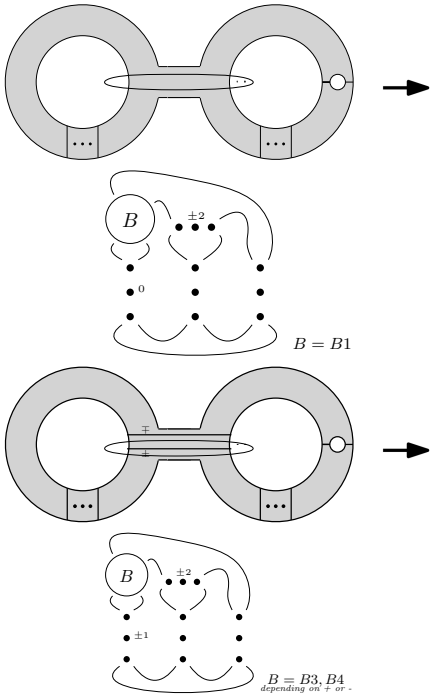


Case 3: One of α_1 and α_2 is on the $T(2,n)$ component and the other is on the $C(2,r)$ component.

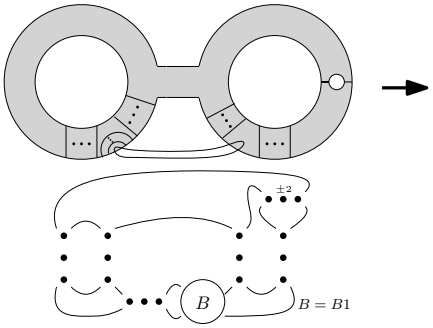


Remark. We only annotate sections 1, 2, 3, and 4 for sake of simplicity. Any instances where α_1 or α_2 lives in one of the non-annotated sections have been covered in previous cases or violate Assumption 2.

Subcase 3.1: $B \cap S$ is a strip. 34 is the only option, as illustrated below. Any other choice for α_1 or α_2 would either violate Assumption 2 or would have been covered in one of the previous two cases.



Subcase 3.2: $B \cap S$ is two disks. 12 is the only option, as illustrated below. Any other choice for α_1 or α_2 would either violate Assumption 2, be a limiting case of 12, or would have been covered in one of the previous two cases.

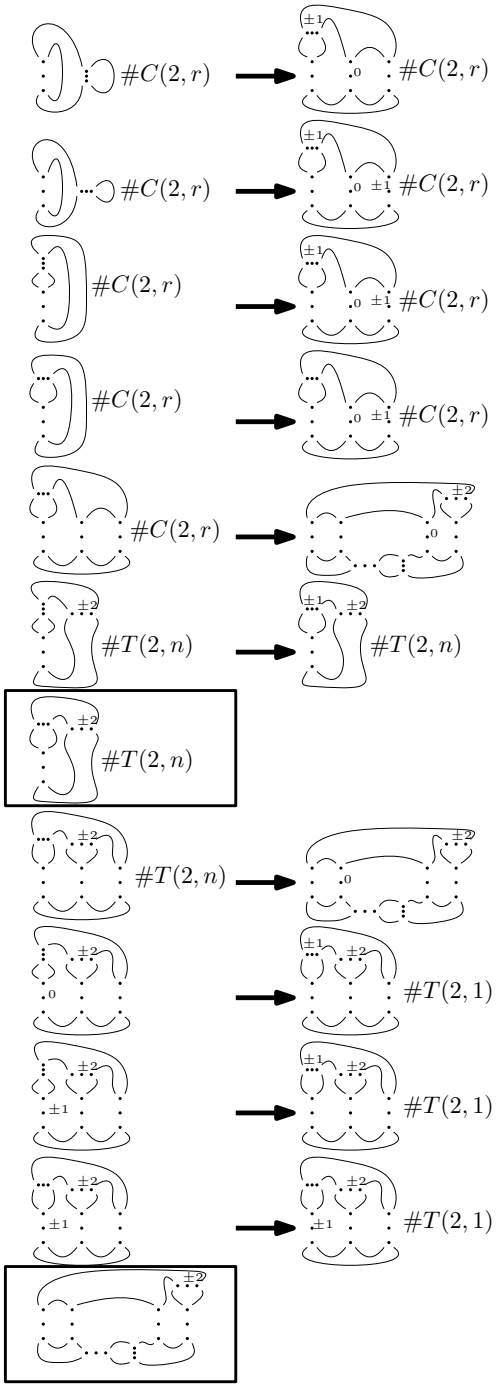


□

6. Main Results

Theorem 2 (Serine)**.** *All knotted and linked products that can arise as a result of site-specific recombination mediated by a serine recombinase on a substrate of the form $T(2,n)\#C(2,r)$ fall into one of the two boxed families illustrated below.*

Proof. We apply Assumption 3 (see Figure 8) to our list of legal configurations for pre-recombinant $B \cup J$ determined in Theorem 1. Depending upon what forms pre-recombinant $B \cap J$ can take $(B1, B2, B3, B4)$ we fill in the appropriate post-recombinant form. The arrows denote isotopies. Every knotted or linked product not boxed can be isotoped to one of the two boxed families.



□

Theorem 3 (Tyrosine)**.** *All knotted and linked products that can arise as a result of site-specific recombination mediated by a tyrosine recombinase on a substrate of the form $T(2,n)\#C(2,r)$ fall into one of the two boxed families illustrated above.*

Proof. We use the exact same reasoning as in the proof of Theorem 1. The difference is in how we apply Assumption 3 to our list of legal configurations for pre-recombinant $B \cup J$. In particular, Assumption 3 for tyrosine treats all pre-recombinant $B \cap J$ $(B1, B2, B3, B4)$ as equivalent. □

Remark. The possible products from a tyrosine recombination will have more restrictions on the number of crossings for certain ellipsis sections due to Assumption 3 for tyrosine.

7. Acknowledgments

I would like to thank my mentor, Dr. Diana Hubbard, for her patience, guidance, and insight. I would also like to thank the University of Michigan REU program for allowing this research experience to take place, and the NSF for helping fund it. All figures were made using IPE.

References

[Ada94] Colin C. Adams. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots*. W.H. Freeman and Company, 1994.

[BF07] D. Buck and E. Flapan. A topological characterization of knots and links arising from site-specific recombination. *Journal of Physics A Mathematical General*, 40:12377–12395, October 2007.

[BV11] D. Buck and K. Valencia. Characterization of knots and links arising from site-specific recombination on twist knots. *Journal of Physics A Mathematical General*, 44(4):045002, January 2011.

[FGL+13] E. Flapan, J. Grevet, Q. Li, C. Sun, and H. Wong. Knotted and linked products of recombination on $T(2,n)\#T(2,m)$ substrates. *ArXiv e-prints*, November 2013.

Department of Mathematics, University of Michigan, Ann Arbor, MI 48109
E-mail address: jkatzm@umich.edu