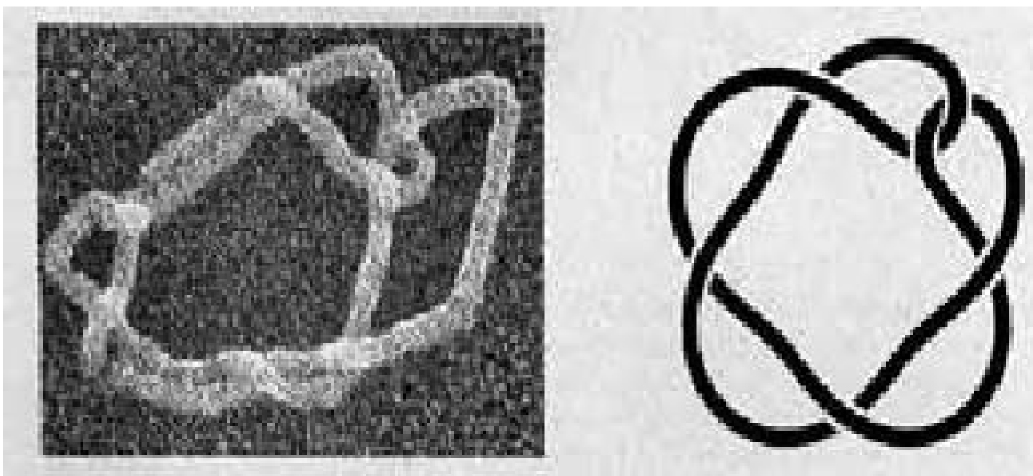
The role of knot theory in DNA-research

Planqué, R.

The role of knot theory in DNA research



Robert Planqué

14th January 2000

Contents

1 Introduction

In this thesis an overview is given of the recent developments of the use of topology in DNA research. Apart from this expository feature new results are given for the mathematical structures of DNA knots.
We start with an introduction discussing relevant parts of knot theory which are used in this paper. Here we introduce some well-known invariants which distinguish different knots. One invariant will receive extra attention: the bridge number. Two subclasses of knots, 2-bridge and torus knots, are especially relevant for further applications in DNA research and will be dealt with in more detail.
With these mathematical tools at our disposal we proceed by discussing the biochemical properties of DNA structures and giving a short introduction to a field which has received much attention during the last decade from topologist, the field of site-specific recombination processes. The main goal of these sections is to show that topology may be used to gain information about enzyme actions which mediate such recombinations of DNA strands. Two cases are considered: recombination mediated by an enzyme called Tn3 resolvase, and one called phage λ integrase. Both have different characteristics which can be described with the use of knot theory. The state of the art in topological enzymology is given.
The model which is presented by previous work in this field is used to prove a theorem which describes all possible knot structures produced by a site-specific recombination event in the case of Tn3 resolvase. This proves to be a useful result providing us with a method to predict where recombination might take place on the knot. Knowing the various types of product knots with this classification theorem, we study the effect of site-specific recombination on the various knot invariants. The genus and graph of a knot are discussed first. We prove a proposition which states that the genus of the product knot is a lower bound for the number of times it has been recombinated. The other two important invariants are the Alexander and Jones polynomials. Although the Alexander polynomial is easily computable for a given knot, the structure between the knot and its Alexander polynomial is too weak for the study of site-specific recombination processes. We proceed by giving a general scheme to calculate Jones polynomials for arbitrary 2-bridge knots. Several subtleties involved in these calculations are discussed. As with the Alexander polynomial, the Jones polynomial does not have the right structure to gain much information on site-specific recombination processes.
For completeness an up to date account of additional problems in DNA research is given in which topology plays a role or in which it might do so in

the future. One of the most exciting frontiers of DNA topology may be the recently found structure of the kinetoplasts, DNA rings which form a network which is very much unlike the forms of DNA found in almost all other living cells.

2 A knot theory primer

2.1 Introduction

In this chapter we discuss the relevant parts of knot theory needed for this paper. In knot theory we consider an embedding $k\colon S^1 \longrightarrow \mathbb{R}^3$ or S^3, or in more general terms embeddings $k\colon S^{n-2} \longrightarrow S^n$. The most important question one tries to solve is to decide whether two given embeddings are the same or not, according to a precise mathematical notion of equivalence.

This classification problem dates from the early nineteen hundreds, when Tait, Kirkmann and Little gave the first tables of knots. The methods in those days were rather empirical and combinatorial. With Poincaré a new branch of mathematics, the ideas and methods of algebraic topology, emerged, which on itself has also been developed to handle problems in knot theory. The mathematicians associated with the early rise of knot theory in the beginning of this century are J.W. Alexander, M. Dehn, W. Burau, O. Schreier, E. Artin, K. Reidemeister, E.R. van Kampen, H. Seifert, J.H.C. Whitehead, H. Tietze and R.H. Fox.
The basic idea used to classify knots is to define an appropriate equivalence relation with which one defines which knots are equal and which are not. It has proved to be useful to introduce quantities which remain invariant under deformations of a knot within the boundaries of the equivalence relation.
These ideas will be developed in the next sections.

2.2 First definitions

For clarity we review some elementary definitions from topology.

Definition Let $\mathbb{R}^n$ be the real n-dimensional space, equipped with the standard metric, $d(x,y) = |x-y|$. Then the **n-dimensional unit disc D^n** is defined to be

$$D^n := \{x \in \mathbb{R}^n \ s.t.\ |x| \leq 1\}.$$

The boundary of D^n, the **$(n-1)$-dimensional sphere** is defined by

$$S^{n-1} := \{x \in D^n \ s.t.\ |x| = 1\} \subset \mathbb{R}^n.$$

S^n can be seen as a 'compact version' of $\mathbb{R}^n$, i.e. if we consider $\mathbb{R}^n \cup \{\infty\}$, the result is diffeomorphic to S^n, where the diffeomorphism in question is a stereographic projection with ∞ as the centre of projection. S^n is therefore the *one-point compactification* of $\mathbb{R}^n$. See Figure 1 below for the identification $\mathbb{R}^2 \cup \{\infty\} \cong S^2$.

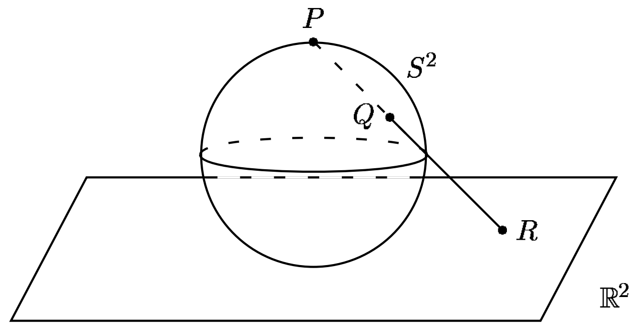


Figure 1: *Stereographic projection of* $\mathbb{R}^2$

Definition A subset k of a space X, which can be either $\mathbb{R}^n$ or S^n is called a **knot** if k is homeomorphic to an $(n-2)$-sphere. More generally, k is called a **link** of p components, if k is homeomorphic with a disjoint union $S_1^{n-2} \coprod \ldots \coprod S_p^{n-2}$. Two knots or links k_1, k_2 are said to be **equivalent** if there is a homeomorphism $h\colon X \longrightarrow X$, such that $h(k_1) = k_2$. This equivalence-relation divides the set of knots up into classes of a certain **knot type**. When two knots are equivalent we write $k_1 = k_2$ or $k_1 \sim k_2$.

For this paper it will be sufficient to consider diffeomorphic images of S^{n-2}, so we assume the knots do not have any singularities. This has some consequences for the types of knots we will study. In the smooth category so called 'wild knots' are excluded. An example of such a knot is given in Figure 2. Although these knots have some remarkable properties they are beyond the scope of this paper. We will rather limit ourselves to 'tame knots'. A more precise definition will be given in a moment.

We often also denote by k the map $k\colon S^{n-2} \longrightarrow S^n$, instead of the homeomorphic image of S^{n-2} in S^n. Then k is an **embedding** into S^n, i.e. k is a homeomorphism $S^{n-2} \xrightarrow{\cong} k(S^{n-2}) \subset S^n$. In the case that we consider k_1 and k_2 to be maps rather than sets, we have that k_1 is equivalent to k_2 if there exists a homeomorphism such that $h \circ k_1 = k_2$.

Definition Two maps $f_1, f_2\colon X \longrightarrow Y$ are called **isotopic** if there is an embedding

$$F\colon X \times [0,1] \longrightarrow Y \times [0,1] \ s.t. \ \begin{cases} F(x,0) = f_0(x) \\ F(x,1) = f_1(x) \end{cases}$$

and $F(x,t) = (f(x,t),t)$, for $x \in X, t \in [0,1]$. F is then called an **isotopy**.

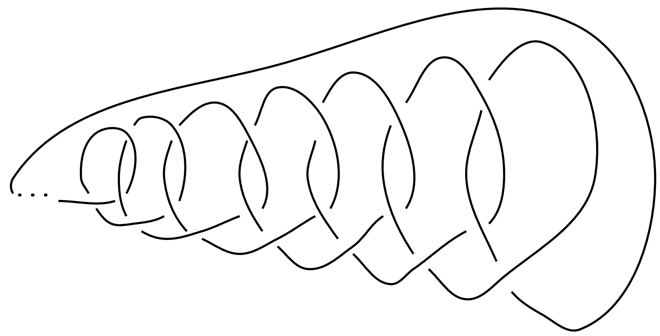


Figure 2: *A wild knot with infinitely many meshes*

For the rest of this paper we confine our knots to those living in $\mathbb{R}^3$ or S^3.

Definition A knot k is called **tame** if there exists an isotopy to a simple closed curve in $\mathbb{R}^3$ or S^3. A knot is called **wild** if it's not tame.

We usually visualise knots by drawing pictures on paper, with over- and undercrossings. A presentation of a knot in a plane P by the orthogonal projection $p\colon \mathbb{R}^3 \longrightarrow P$ is called **regular** if $p^{-1}(x), x \in P$ consists of up to two points in k. It's a simple exercise to prove that one can always make a **regular projection** for a given tame knot in $\mathbb{R}^3$. Here we mean that the knot is projected in such a way that it satisfies the following condition: there are only finitely many multiple points (points with more than one pre-image), and all these points are double points. The minimal number of crossings in a regular projection is called the **order** of the knot (projection). Such regular presentations of knots are also called **knot diagrams**.

We will give some more definitions concerning oriented knots in $\mathbb{R}^3$ or S^3.

Definition Let k be a knot in $\mathbb{R}^3$ or S^3. The knot obtained by inverting its orientation is called the **inverted knot** and denoted by $-k$. The **mirror image** of k is obtained by reflection of k in a plane, and denoted by k^*.

A knot is called **invertible** if $k = -k$ and **ampicheiral** if $k = k^*$.

A knot projection is called **alternating** if under- and overcrossings alternate while following the orientation of the knot. A knot is called alternating if it admits an alternating projection.

Another concept in knot theory is the genus of a knot. To introduce this notion, one first defines the concept of a Seifert surface of a knot. The proof of the existence of Seifert surfaces can for instance be found in (Burde & Zieschang 1985).

Proposition 2.1 *A simple closed curve $k \subset \mathbb{R}^3$ is the boundary of an orientable surface S, embedded in $\mathbb{R}^3$. Such a surface is called a* **Seifert surface** *for k.* □

From classical geometry one can define the genus of such Seifert surfaces in the following way: the surface is a compact 2-dimensional manifold with a disc removed. By closing the Seifert surface by adding a disc we may resort to the well-known classification theorem for compact surfaces which tells us that, up to a homeomorphism, this compact surface is a connected sum of g tori. Now we define the genus of the Seifert surface to be the genus of its closed 2-manifold.
Since a knot k usually has more than one of these surfaces with different genera, the genus of one Seifert surface is not an invariant for the knot. But the minimal number of these genera is, and this number is called the **genus of a knot**. Evidently, the genus of a knot does not depend on the choice of the representative curve in its equivalence class. The genus is our first invariant for a knot, and was introduced by Seifert in (Seifert 1934).

Another concept was introduced by Schubert, in (Schubert 1949): the product of knots.

Definition Let $k \subset \mathbb{R}^3$ meet a plane E in two points P and Q. The arc of k from P to Q is closed by an arc in E to obtain a knot k_1; the other arc (from Q to P) is closed in the same way and defines a knot k_2. The original knot k is called the **product** of k_1 and k_2, in notation $k = k_1 \# k_2$. An example is shown in Figure 3.

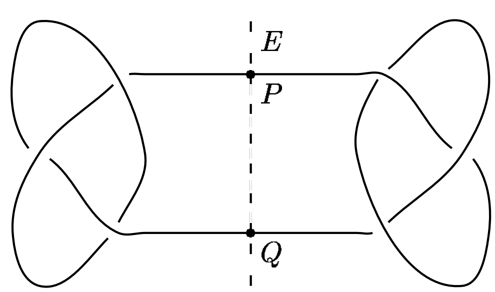


Figure 3: *The product of two trefoils*

Definition A knot is called **prime** if it cannot be written as the product of two non-trivial knots.

For any given two knots, their product can be defined by the inverse procedure. The product will not depend on the plane E chosen in the construction, nor on the choice of representatives from the equivalence-classes of k_1 and k_2. An important result concerning products of knots is the uniqueness of prime decomposition, which plays much the same role as the prime-factorisation theorem in number theory. The precise statement is shown below. A proof can be found in the same article by Schubert.

Theorem 2.2 *For any tame knot k we have the following two statements:*
(1) k can be decomposed into a finite number of prime knots.
(2) This decomposition is unique up to the order of factors. That is, suppose we can compose k in two ways, $k_1\#\cdots\#k_n$ and $k'_1\#\cdots\#k'_m$, then $n = m$ and if we choose suitable subscripts of the decompositions, we have that $k_1 \sim k'_1, \dots, k_n \sim k'_n$. □

Definition A tubular neighbourhood $V(k)$ of a knot $k \subset S^3$ is homeomorphic to a solid torus. There is a simple closed curve m on the boundary of $V(k)$, denoted by $\partial V(k)$, which is nullhomologous in $V(k)$ but not in $\partial V(k)$. We call m a **meridian** of k. Any two meridians of k are isotopic.
A Seifert surface will meet $\partial V(k)$ in a simple closed curve l, if V is suitably chosen which is called a **longitude** of k. This curve is isotopic to k in $V(k)$. These two curves correspond to the generators of the fundamental group of the torus.

If we consider a tubular neighbourhood of the trivial knot in S^3, and we take the complement of this neighbourhood, then the result is homeomorphic to a solid torus $S^1 \times D^2$. In notation: $S^3 - S^1 \times D^2 \cong S^1 \times D^2$. The **complement of a knot** is then defined by $\overline{S^3 - k} \cong \overline{S^3 - V(k)}$. The study of complements of knots is an important area of research in knot theory. This is stimulated by the following theorem:

Theorem 2.3 *If two knots k_1 and k_2 in S^3 are equivalent, then their complements $S^3 - k_1$ and $S^3 - k_2$ are homeomorphic.* □

Remark The converse is also true, which has been proved by Gordon and Luecke in the late 1980's. But unfortunately, non-equivalent *links* may have homeomorphic complements. An example is given in Figure 4.

This notion of complements of knots allows us to define the concepts of companionship and satellites of knots.

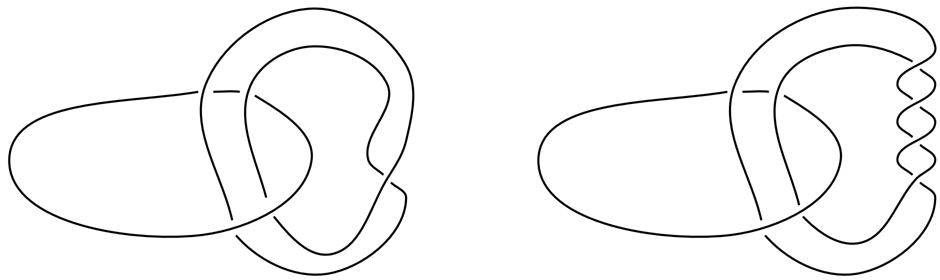


Figure 4: *Two non-equivalent links with equivalent complements*

Definition Let k be a knot in S^3, and V an (unknotted) solid torus in S^3, such that $k \subset V \subset S^3$. Assume that k is not contained in a 3-ball of V (i.e. k goes at least once around the torus). A homeomorphism $h\colon V \longrightarrow W \subset S^3$ onto a tubular neighbourhood W of a non-trivial knot l, which maps a meridian of $S^3 - V$ onto a longitude of W maps k onto a knot $k' = h(k) \subset S^3$. This new knot k' is called a **companion** of k, and k is called a **satellite** of k'.

Another way of looking at certain types of knots is by interpreting them as braids. This concept was introduced by Artin in (Artin 1925).

Definition A finite collection of strings which do not have local maxima or minima is called a **braid**. One can visualise them by drawing a rectangle, with at two opposite sides n points. One then draws curves from one side with points to the other, braiding the strings. The result is called an n-**braid**.

Definition The isotopy classes of n-braids form a group called the **braid group** $\mathfrak{B}_n$. The braid group can be represented by $n-1$ generators $\sigma_i, i = 1, \ldots, n-1$, where the σ_i are illustrated in the Figure 5.

A braid may be closed with respect to an axis h (cf. Fig 6).

2.3 The bridge number

The concept of the n-braid gives us the opportunity to define the notion most important in this paper:

Definition Let k be a knot or link in $\mathbb{R}^3$ which meets a plane E in $2m$ points, such that the *arcs* of k contained in each halfspace (i.e. curves without any knotting but mere lines joinig two points) are transverse

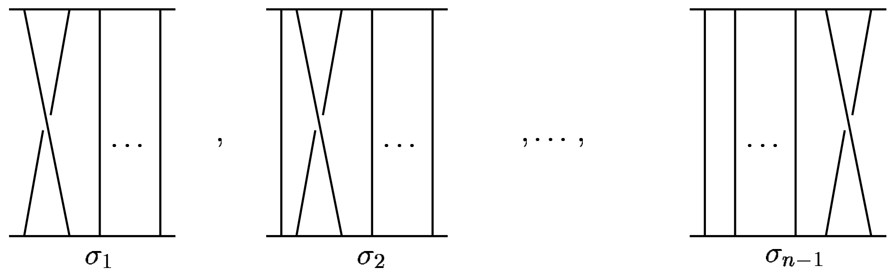


Figure 5: *The generators of the braid group* $\mathfrak{B}_n$.

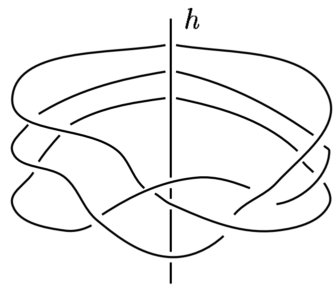


Figure 6: *A closed braid with axis* h

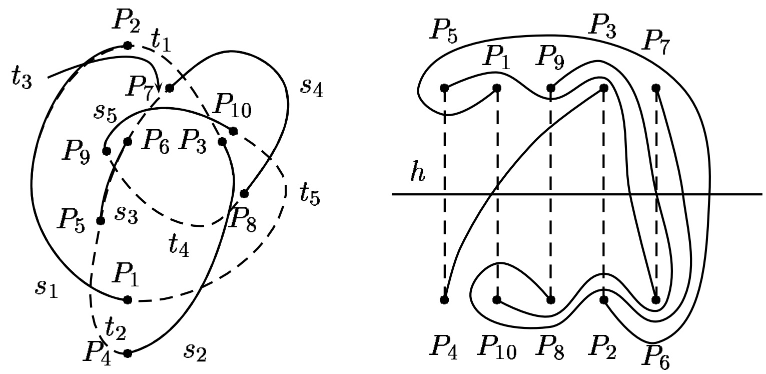
to E. Then the pair (k, E) is called an m-**bridge representation** of k, and m is called the **bridge number** for this representation. Of course, different planes may give a different number of intersection points, and analogously to the genus of k, we define the **bridge number** of k, $br(k)$, to be minimum bridge number in any representation of k.

We can also define the bridge number alternatively:

Definition Let k be represented by a regular diagram. At each crossing point, remove a small segment from the diagram that passes over the crossing point, until you end up with a collection of disjoint curves. The removed curves are called **bridges**, and the number of times you have to remove the small segments to obtain the simple arcs is the bridge number of the diagram. As before, define the **bridge number of the knot** to be the minimal number of these bridge numbers for all regular diagrams of the knot. It is a knot invariant for k.

An example can be found in Figure $7a$.
Hence a regular projection of order n admits an n-bridge representation.

Proposition 2.4 *An m-bridge knot k can be represented as a closed m-braid.*



a. A 5-bridge presentation

b. The same presentation with parallel bridges

Figure 7: *Proof of Proposition 2.4*

PROOF Choose $2m$ points P_i in a regular projection of order m of k, one on each arc between undercrossing and overcrossing. This defines an m-bridge representation, with arcs s_i, $1 \leq i \leq m$ between P_{2i-1} and P_{2i} in the upper half space and arcs t_i, $1 \leq i \leq m$ joining P_{2i} and P_{2i+1} in the lower half space (cf. Fig. 7*a*). We may arrange the projection of k by performing an isotopy to form m parallel straight line segments bisected by a staight perpendicular line h such that all P_i, i odd are contained above h and all other points below h (cf. Fig. 7*b*). We may assume that each arc s_i or t_i meets h only once by adding extra arcs if necessary and subdividing the former arc into smaller ones. If we now deform the arcs such that they have no local minima or maxima, going from one side of h to the other, we have a closed braid with axis h. □

A $2m$-braid completed by $2m$ simple arcs to make a link as in Fig. 8 is called a **plat** or **$2m$-plat**. Observe that a closed m-braid can be presented as a $2m$-plat. Hence we know that an m-bridge knot can be presented by a $2m$-plat.
A canonical way of drawing an m-bridge knot, is by braiding $2m$ strands without local minima or maxima (w.r.t. some height function while drawing the knot vertically). The $4m$ ends of all the strings are then glued together as in the example for a 2-bridge knot in Figure 8.

A regular projection of a knot admits an m-bridge representation, and an important fact for this paper is that alternating representations yield minimal bridge number for (k, E), which means, that by drawing an alternating diagram for k, and finding a way to determine the bridge number of this

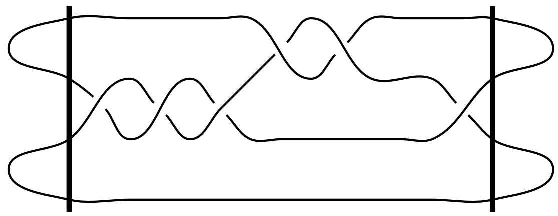


Figure 8: *A 2-bridge knot regarded as a 4-plat*

drawing, one has the bridge number of the knot.

We state some simple results:

By convention, the unknot has bridge number 1.

Proposition 2.5 *The only knot with bridge number equal to 1 is the unknot.*

PROOF The proof comes from (Schubert 1954). Let E_1 and E_2 be two horizontal planes, with E_1 the upper one. Let k be a knot consisting of line segments in E_1 and E_2 and orthogonal lines between the two planes which connect points in E_1 from k with those in E_2 from k, to form a knot. We call such a representation of k a **bridge representation** (Brückedarstellung) and its equivalent to our first definition of the bridge number. A bridge in this representation is a subset of the line segments of k consisting of one line segment in E_1, and the adjoining line segments orthogonal to the planes. The endpoints of these adjoining line segments in E_2 are called the base points (Fußpunkte) of the bridge.
Now let k be a knot with a bridge representation, also denoted by k. Since k is a 1-bridge knot, we may assume that E_1 contains one line segment. In E_2 we have one line segment l with two base points coming from the bridge as endpoints. l is contained in a sufficiently large 2-simplex in E_2, which can be extended to a 2-sphere which intersects the bridge only in the base points. Therefore l is isotopic to the straight line connecting the base points in S^2, which shows that k is isotopic to the unknot. □

The following goes without proof:

Proposition 2.6 *If k is an n-link, then $br(k) \geq n$.* □

The following theorem gives an idea of the power of the bridge number as invariant: it's nearly linear under products of knots:

Theorem 2.7 *Suppose k_1 and k_2 are two knots or links. Then*

$$br(k_1 \# k_2) = br(k_1) + br(k_2) - 1.$$

□

This theorem has been proved by Schubert, in (Schubert, 1954), the article in which also the bridge number has been introduced. It involves a detailed and rather technical proof so we will not repeat it here.

The bridge number of a knot is not an easily computable quantity. In general, there does not exist an algorithm which determines the bridge number for a given knot k (Murasugi 1996).

Before we state the next proposition, let $\mathfrak{G}$ denote $\pi_1(S^3 - k)$, the fundamental group of the knot complement of k. This group is called the **knot group** of k.

Proposition 2.8 *If $br(k) = n$, then the knot group $\mathfrak{G}$ is a group with n generators and $n-1$ relations.* □

We can also combine bridge number and companionship to find for instance:

Proposition 2.9 *If k_1 is a proper companion of k_2 (i.e. they are not equivalent), then $br(k_1) < br(k_2)$.* □

The proof can be found in (Schubert 1954).

2.4 2-bridge knots

The class of 2-bridge knots has been studied extensively for the last seventy years. It is the only class of n-bridge knots which has been completely classified. This makes studying of these knots a great deal easier compared to studying knots with different bridge number. We will state some important results from this field.

We first introduce a bit of notation to handle 2-bridge knots. We draw a knot horizontally as a 4-braid and number the strings from bottom to top from one to four. We let the a_{2i+1}, $i = 0, \ldots, m$, denote the crossings between the second and third string, and the a_{2i}, $i = 1, \ldots, m$, the crossings between the third and fourth string. The ends of the first and second string on the left are glued together, as are all other pairs of strings, to make a knot or 2-link. See Figure 9.

One can always project a 4-braid in such a way, that the first string does not contain any crossings. The idea to get rid of the crossings between the lower two strings is to stretch up the knot such that the crossings $a_{2i+1},\ i = 0, \ldots, m$ between the middle two strings are aligned vertically. On top of that, the crossings in the lower strings are aligned horizontally, left of the vertical line of middle-crossings. Analogously, the crossings between the upper strings are aligned right of the middle-crossings. Now one can eliminate the crossings between the lower strings by unwinding the outer half-circle a suitable number of times. This unwinding on the left will result in a positive contribution of crossings on the right. Every winding on the left is eliminated in this way, and the result is a knot with no crossings between the lower string, as depicted in Figure 9 (after transformation to its original 4-braid representation).

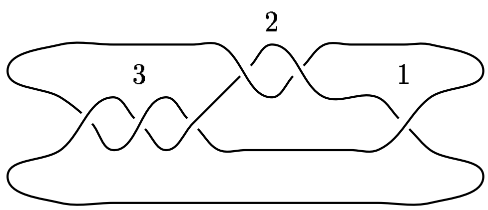


Figure 9: *The canonical projection of the 2-bridge knot* $< 3, 2, 1 >$

Now one has a 4-plat, with alternating crossings between the middle and the upper strings. If one has successive crossings of opposite sign one can locally unwind them until this is no longer the case. By this process all 'trivial winding' is eliminated and in the end one has a 2-bridge knot with all crossings positive or all negative. We can thus represent a 2-bridge knot by a vector with integer entries of odd length (since both beginning and end crossings have to be between the middle strings), and the claim is proved, up to working out the details of the unwinding of the lower crossings. For more details, see the excellent exposition by Murasugi (1996). We have therefore proved the following Proposition:

Proposition 2.10 *Any 2-bridge knot can be written in a canonical form denoted by* $< a_1, \ldots, a_{2m+1} >$ *with* a_i *all strictly positive or strictly negative integers:*

The vector-notation invites new inquiries in 2-bridge knots. For instance, one can make a continued fraction from the entries of the vector, and create a number

$$\frac{\alpha}{\beta} = [a_1, \dots, a_n] := a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots a_{n-1} + \frac{1}{a_n}}} \tag{1}$$

such that $gcd(\alpha, \beta) = 1$. Conversely, for any fraction p/q one can find a continued fraction $[c_1, \dots, c_m]$. This presentation is not at all unique: for instance we have the following identities:
If m is even and $c_m > 1$, then

$$[c_1, \dots, c_m] = [c_1, \dots, c_m - 1, 1]$$

and if $c_m = 1$, we may write

$$[c_1, \dots, c_m] = [c_1, \dots, c_m + 1]$$

We can therefore assume m to be odd, and a 2-bridge can be constructed from the continued fraction. A 2-bridge knot $k =< a_1, \dots, a_n >$ is said to be of type (α, β), if $\alpha/\beta = [a_1, \dots, a_n]$. This fraction α/β completely classifies the knot, as stated in the following theorem:

Theorem 2.11 *Suppose that k and k' are 2-bridge knots of type (α, β) and (α', β') respectively. Then k and k' are equivalent up to orientation of the knots if and only if:*
(1) $\alpha = \alpha', \ \beta \equiv \beta' \pmod{\alpha}$
or
(2) $\alpha = \alpha', \ \beta\beta' \equiv 1 \pmod{\alpha}$.
Further, the mirror image k^* of k is a 2-bridge knot of type $(\alpha, -\beta)$. Therefore, a necessary and sufficient condition for k to be ampicheiral is that
(3) $\beta^2 \equiv -1 \pmod{\alpha}$. □

A proof of this theorem can be found in (Schubert 1956). The following proposition follows from Prop. 2.5 and Thm. 2.7.

Proposition 2.12 *Any 2-bridge knot is prime.* □

Let k be an n times twisted unknot. If we knot the ends together, we get something like the knot shown in the figure below. Such a knot is called a **twist knot**. We have the following proposition:

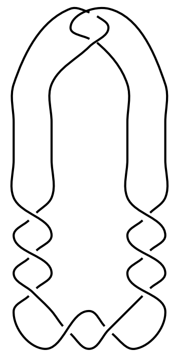


Figure 10: *A twists knot with 8 twists*

Proposition 2.13 *Every non-trivial twist knot is a 2-bridge knot.* □

We can give a precise characterisation of 2-bridge knots in terms of the genus. To do this, we first determine the structure of the 2-bridge knot, seen as a 4-plat. Since we can always project a 4-plat such that one of the strings does not cross any of the others, the set of 4-plats is generated by two elements σ_1 and σ_2. In the following proposition we see that there is a nice relation between the coefficients of the vector $< a_1, \dots, a_{2m+1} >$ and the braid-presentation in the generators σ_1, σ_2 (Burde & Zieschang 1985):

Proposition 2.14 *The 2-bridge knot k of type (α, β) has a presentation as a 4-braid as*

$$\zeta = \sigma_2^{a_1}\sigma_1^{-a_2}\dots\sigma_2^{a_{2m+1}},$$

with $a_i > 0$, such that the continued fraction $[a_1, \dots, a_{2m+1}]$ equals α/β. □

With this bridge-presentation it is possible to state the result on genera of 2-bridge knots, also proved in (Burde & Zieschang 1985):

Proposition 2.15 *Let $\sigma_2^{a_1}\sigma_1^{-a_2}\dots\sigma_2^{a_{2m+1}}$ represent a 2-bridge knot k of type (α, β). Then the genus of k is*

$$\frac{1}{2}[\sum_{j=1}^{m}|a_j| - \mu]$$

where μ is 1 if k is a knot, and 2 if k is a 2-link. □

2.5 Torus knots

In this section we discuss another group of knots which has been studied exhaustively the last decades, since they occur in various branches of science. They are the torus knots. We begin by defining the concepts of handlebodies and Heegaard splitting.

Definition A **handlebody** V of **genus** g is obtained from a 3-ball B by attaching g handles $D^2 \times [0, 1]$, such ∂V is an orientable closed surface of genus g. The composition of a closed orientable 3-manifold M^3 into two handlebodies V, W such that $M^3 = V \cup W$ and $V \cap W = \overset{\circ}{V} = \overset{\circ}{W}$ is called a **Heegaard splitting of genus** g.

For S^3 we have the following theorem.

Theorem 2.16 *Any two Heegaard splittings of of the same genus of S^3 are homeomorphic. To be more precise: If (W, W) and (V', W') are two Heegaard splittings of the same genus, then there is an orientation preserving homeomorphism $h\colon\ S^3 \longrightarrow S^3$ such that $h(V) = V'$ and $h(W) = W'$.* $\square$

A proof of this can be found in (Waldhausen 1968). We can directly give an application of this result to knot theory.

Proposition 2.17 *Every knot k in S^3 can be embedded in the boundary of the handlebodies of a Heegaard splitting of S^3.*

Although the proof can be found in general textbooks on knot theory, we repeat it here since we will need it later on.

PROOF We can colour the knot by projecting it onto a plane, and giving the distinct regions alternatingly black and white colours, in chess board manner. This is always possible because the projected knot is a closed curve in the plane (For the links which occur in this paper it's also possible, since we only need n-bridge links, which admit an alternating regular diagram, making the chess board colouring possible). By convention, we mark the region outside the projected curve with white. Next we put points inside the white regions, and connect points when their corresponding regions are separated by a crossing point of the projected knot. For every crossing point we get an edge of the graph we produce in this manner. The graph which belongs to the original knot will be denoted by Γ, with the set of vertices of Γ denoted by V and the set of edges of Γ by E. An example is shown in Figure 11.

We now take a tubular neighbourhood W of the graph. We claim that W is a handlebody. This is seen by taking a tree of Γ (i.e. a subset of (V, E) such that all vertices of Γ are in this subset, and there are no loops) and for every $e \in E$ we put a handle $D^2 \times [0,1]$ onto the tubular neighbourhood of the tree, and by induction we have a handlebody of genus g, where g is the number of edges in E minus the number of edges of the tree (which is well defined).
Now we can find our knot k back, by drawing twists between the vertices of E in W as in Figure 12. The result is the knot k. □

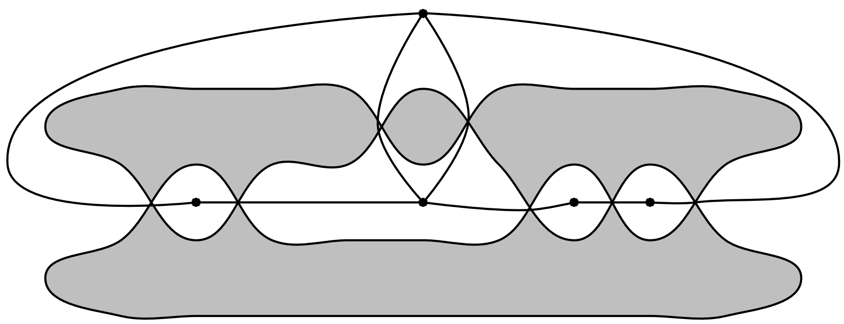


Figure 11: *Construction of a graph of a 2-bridge knot*

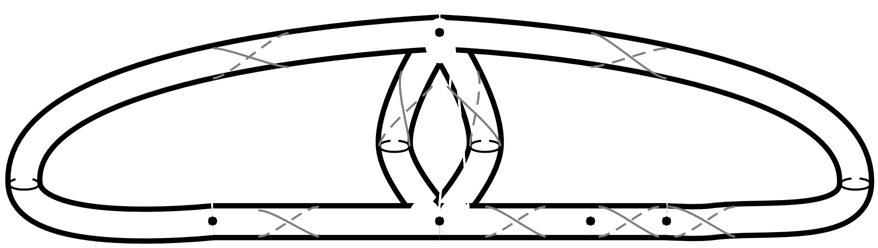


Figure 12: *Construction of the knot from its handlebody*

Let $S^3 = V \cup W$ be the genus 1 Heegaard splitting of S^3. We may assume that V is an unknotted solid torus in $\mathbb{R}^3$, and $T := V \cap W$ a torus T^2. There are meridians μ and ν of V and W on T which intersect in a base point with intersection number 1 (i.e. they intersect once). Any closed curve on T is homotopic to a curve $\mu^a \cdot \nu^b$, with $a, b \in \mathbb{Z}$. If a and b are relatively prime, then the homotopy class of this curve $\mu^a \cdot \nu^b$ contains a simple closed curve. Such a curve intersects the curve μ and ν a and b times respectively.

Definition A **torus knot** of **type** (a, b), $t(a, b)$, is a simple closed curve on $T = V \cap W$ which intersects μ and ν a and b times respectively.

We can give a simple characterisation of torus knots by their knot groups:

Proposition 2.18 *The knot group $\mathfrak{G}$ of a torus knot of type (a, b) can be presented as follows:*
$\mathfrak{G} =< u, v | u^a v^{-b} >$, *where μ, ν represent u, v.* □

A proof can be found in (Burde & Zieschang 1985). This presentation of $\mathfrak{G}$ can be used to give a classification of all torus knots, which is also proved in (Burde & Zieschang 1985).

Theorem 2.19 *Two torus knots $t(a, b)$ and $t(a', b')$ are equivalent if and only if (a', b') is one of the following: (a, b), (b, a), $(-a, -b)$, $(-b, -a)$.* □

Furthermore this result can be used to prove that torus knots are invertible but not ampicheiral. A question which is not often asked in topology is whether we can obtain insight in the relative position of a subset of objects to the greater set of objects. In knot theory invariants are often not easily computable (eg. the bridge number), and being able to decide whether a given knot belongs to some class of knots is not easy. Fortunately for torus knots one can give some information concerning this question. Before we can state the result we review a concept from algebra, the centre of a group.

Definition Let $\mathfrak{G}$ be a group. Define the **centre** of $\mathfrak{G}$ to be the subset

$$\{y | xyx^{-1} = y, \quad x, y \in \mathfrak{G}\}.$$

In other words, the centre of $\mathfrak{G}$ consists of those elements of $\mathfrak{G}$ which commute with all elements of $\mathfrak{G}$.

Theorem 2.20 *A non-trivial knot whose knot group $\mathfrak{G}$ has a non-trivial centre is a torus knot.* □

We can also give some nice characterisations of torus knots in bridge numbers, **crossing numbers** (i.e. the minimum number of crossings one finds in a regular diagram of the knot, denoted by $c(k)$). We state the results:

Proposition 2.21 *The bridge number of a torus knot of type* (a, b) *is*

$$br(t(a,b)) = \min\{|a|, |b|\}.$$

□

Proposition 2.22 *The crossing number of a torus knot of type* (a, b) *is*

$$c(t(a,b)) = \min\{|a|(|b|-1), |b|(|a|-1)\}$$

□

The proofs can be found in (Murasugi 1996). A far deeper result has been obtained for the **unknotting number** of a knot, which is defined as the minimal number of transient breaks one has to perform to obtain the unknot. A transient break is performed by making a cut in the curve at some crossing (w.r.t. a projection onto a plane) and transforming it into its negative counterpart. The unknotting number is denoted by $u(k)$. The proof of this theorem can be found in (Kronheimer & Mrowka 1993, 1995).

Theorem 2.23 *For* $a, b > 0$ *and* $gcd(a,b) = 1$*, we have*

$$u(t(a,b)) = \frac{(a-1)(b-1)}{2}.$$

□

Investigation of the prime factorisation of torus knots gives the following result:

Theorem 2.24 *Torus knots are prime.* □

Two proofs can be found in (Burde & Zieschang 1985).

2.6 Knot polynomials

The classification of knots has always been the major problem in knot theory. As we have already seen, many invariants which distinguish certain knots have been constructed over the years, and methods have become increasingly sophisticated. Among these invariants there is a group which can be used for our purposes in DNA topology: the knot polynomials.
Since Alexander has introduced the first polynomial in 1928 (Alexander

1928), many others have been constructed. Among them are the Jones, Kauffman, Conway and HOMFLY[1] polynomials. All have their own strengths and weaknesses and until now no invariant has been found which classifies all knots up to isotopy. In this paper the two most important polynomials will be introduced and a discussion on their use in DNA topology will be given in a later chapter. We start with the more classical invariant of the two: the Alexander polynomial.

2.6.1 The Alexander polynomial

The Alexander polynomial is generally considered in two rather different manners: the first is purely algebraic, the second more constructive using linear algebra. Getting a good grip on the notion of the Alexander polynomial requires both view points. We begin with the more theoretic approach.

Let Λ be the ring of Laurent polynomials with integer coefficients, sometimes also denoted by $\mathbb{Z}[t, t^{-1}]$. A typical element of Λ has the form

$$c_{-r}t^{-r} + \cdots + c_0 + c_1 t + \cdots + c_s t^s$$

with the c_i integers. Λ is a principal ideal ring, which means that for every ideal $\mathfrak{I}$ of Λ there exists an element $r \in \Lambda$ such that $\mathfrak{I} = r\Lambda$. In our case r is a polynomial.

Now consider a knot $k \subset S^3$. The knot complement X has an infinite cyclic covering $\tilde{X}$.[2] We may construct it by the following method: let M be a Seifert surface for k and let $N\colon\ \dot{M} \times (-1, 1) \longrightarrow S^3$ be an open *bicollar* of the interior of M. Thus we have $\dot{M} = N(\dot{M} \times 0)$. Denote:

$$\begin{aligned} N &= N(\dot{M} \times (-1, 1)), \\ N^+ &= N(\dot{M} \times (0, 1)), \\ N^- &= N(\dot{M} \times (-1, 0)), \\ Y &= S^3 - M. \end{aligned}$$

Thus we have two triples (N, N^+, N^-) and (Y, N^+, N^-). Take countable many copies of these triples, denoted (N_i, N_i^+, N_i^-) and (Y_i, N_i^+, N_i^-), $i \in \mathbb{Z}$. We now form an identification space by identifying $N_i^+ \subset Y_i$ with $N_i^+ \subset N_i$ via the identity homeomorphism, and similarly $N_i^- \subset Y_i$ with $N_{i+1}^- \subset N_{i+1}$.

[1] The name HOMFLY is an acronym for the first letters of the six persons who constructed the invariant

[2] Much can be said about coverings of knots in general. For thorough accounts on this topic, see (Rolfsen 1976, Burde & Zieschang 1985).

The resulting space, $\tilde{X}$, is a path-connected 3-dimensional manifold. We have a regular covering map $p\colon \tilde{X} \longrightarrow X$. There is a covering automorphism $\tau\colon \tilde{X} \longrightarrow \tilde{X}$ which takes Y_i to Y_{i+1} and N_i to N_{i+1}, and τ generates the group $Aut(\tilde{X})$, which is isomorphic to $\mathbb{Z}$. The first stage of the method described above is also referred to as *cutting the knot complement along the surface* .
The next step is to consider the first homology group of $\tilde{X}$, which is called the *Alexander invariant* of k. This group is equal to $\pi_1(\tilde{X})$ divided by its commutator subgroup and thus abelian. The covering transformation $\tau\colon \tilde{X} \longrightarrow \tilde{X}$ induces an automorphism $\tau_*\colon H_1(\tilde{X}) \longrightarrow H_1(\tilde{X})$ and this τ_* will be the key to the Alexander polynomial.
For a commutative ring with identity L, let A be an $m \times n$ matrix with entries from L. Write L^n for the free module $L[x_1, \ldots, x_n]$ and $L^m = L[y_1, \ldots, y_m]$. Let $f\colon L^n \longrightarrow L^m$ be the L-module homomorphism determined by matrix multiplication by A. and define K to be the quotient module $L^n/f(L^m)$. The matrix A is called a *presentation matrix* or *Alexander matrix* for the module K. Observe that two matrices A, B yield isomorphic L-modules if we can convert A into B by one of the following steps:

(1) interchange two rows or two columns;
(2) multiply a row or a column by a unit of Λ;
(3) add any multiple of on row to another row, or a multiple of any column to another column;
(4) add or remove a column of zeros;
(5) interchange A with

$$\begin{pmatrix} A & 0 \\ 0 & 1 \end{pmatrix}$$

or vice versa.

Moreover, if A is a square matrix, the *determinant* of A is an isomorphism invariant of the module K. By the robservation above this determinant is determined up to a unit of L.
More concretely, τ_* makes $H_1(\tilde{X})$ into a Λ-module, since we can define the product of a polynomial $p(t) \in \Lambda$ with a homology class $[z] \in H_1(\tilde{X})$ by

$$p(t)[z] = c_{-k}\tau_*^{-k}[z] + \cdots + c_l\tau_*^l[z]$$

This polynomial functions as a 'separator' for the 1-cycles $c_i[z]$ for each subspace Y_i of the infinite cyclic covering. The power of the variable in $p(t)$ determines the place of $[z]$ in the corresponding Y_i. Since Λ is a principal ideal ring there is a Laurent polynomial $\Delta(t)$ such that $H_1(\tilde{X}) = \Lambda/\Delta(t)\Lambda$.

This polynomial is called the **Alexander polynomial**, and is uniquely determined up to a unit of Λ, i.e. up to a factor $\pm t^k,\ k \in \mathbb{Z}$.

Before we turn to the more constructive approach to determine Alexander polynomials for given knots, we make an important remark. The existence of the Alexander polynomial relies on the fact that Λ is a principal ideal ring. If we try to construct a similar invariant for links rather than knots, the corresponding polynomial rings are not principal ideal rings, and ambiguity in the definition of an Alexander polynomial arises. There is no consensus which definition is most suitable and in general only the Alexander polynomial of true knots are considered. Studying the Alexander invariant of the link is still possible though. The universal abelian cover of X is not infinite cyclic however, but a free abelian group on p generators, where p is the number of components of the link. The ring with Laurent polynomials now becomes $\mathbb{Z}[x_1, \ldots, x_k]$. For more on the Alexander invariant of links see (Rolfsen 1976). We now return to the constructive approach to determine Alexander polynomials of true knots.

To study $H_1(\tilde{X})$ it is necessary to find a presentation matrix A for $H_1(\tilde{X})$. We have already seen that the determinant of A is an isomorphism invariant for $H_1(\tilde{X})$, and it is our primary goal to define a presentation matrix A. This can be done in several ways, and the equivalence of the definitions relies on the fact that the different matrices are all presentation matrices for $H_1(\tilde{X})$. We give two definitions of a presentation matrix A of a knot k. The justification of these definitions is rather involved and beyond the scope of this paper. The interested reader may find thorough accounts in (Rolfsen 1976, Burde & Zieschang 1985).

The first matrix is famous for its many applications throughout knot theory and is usually called the **Seifert matrix** of a knot. For a knot k in S^3, choose a Seifert surface M of k and a bicollar $N = \dot{M} \times [-1, 1]$ in the knot complement $S^3 - k$. We take a representative of an element in the first homology group of $\dot{M}$, both denoted by x, and denote the 1-cycle $x \times 1$ by x^+. These two 1-cycles x, x^+ have a well-defined linking number. If we generate $H_1(\dot{M})$ by a basis $e_1, \ldots, e_{2g}$, where g is the Seifert genus of M, we may set up a matrix of linking numbers of the pairs of 1-cycles in the basis in the following way:

Definition The Seifert matrix associated to a pair (k, M) and its bicollar N is defined by $V = (v_{ij})_{i,j},\ 1 \leq i, j \leq 2g$ where $v_{ij} = lk(e_i, e_j^+)$.

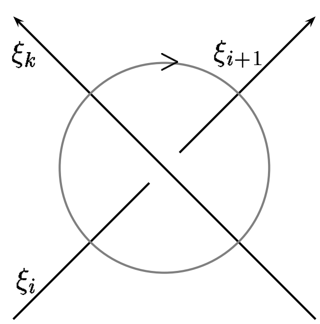
Theorem 2.25 *If V is a Seifert matrix for a knot k in S^3, then $V - tV^T$ is an Alexander matrix for k. Its transpose, $V^T - tV$, as well.* □

A proof can be found in (Rolfsen 1976). The presentation matrix of the Alexander invariant of a knot in S^3 is square and its determinant is thus defined.

Corollary 2.26 *For a knot k in S^3, the Alexander polynomial of k, $\Delta(t)$, is equal to $\det(V - tV^T)$.*

It's usually quite easy to write down a Seifert matrix for a given knot using the algorithm described above. In a later chapter we will give an example and try to characterize the recombination processes in circular DNA using the Alexander polynomial. We now turn to another easy method to form a presentation matrix.

The second constructive approach to form a presentation matrix uses the concept of knot groups. Unfortunately, by its algebraic nature, the precise geometric information is not well readible in this second presentation, which makes a comparison between the two matrices on the geometric level rather difficult. To set up general methods to determine the knot group for arbitrary knots is quite envolved, and can be found in (Burde & Zieschang 1985, Fox 1962). In short, we may write $\mathfrak{G} = \{x_1, \ldots, x_n | r_1, \ldots, r_n\}$. Here we have projected the knot and we've labeled the overpasses in this projection with x_i. For each crossings we may set up a relation r_i which is generally of the form $x_k x_i x_k^{-1} x_{i+1}^{-1}$. A generator x_i comes from an arc ξ_i in the projection. See Figure 13 for details.



The local situation to get the relation $x_k x_i x_k^{-1} x_{i+1}^{-1}$

Figure 13: *The relations of the knot group*

Let $\mathfrak{G}' = \pi_1(\tilde{X})$ and $\mathfrak{G}''$ its commutator subgroup. Then we already know that $H_1(\tilde{X}) \cong \mathfrak{G}'/\mathfrak{G}''$ and this last group becomes a Λ-module by our previous discussions. If we abelianise the generators of $\mathfrak{G}$ they all become equal and give a generator of $\mathbb{Z}$. By changing to a new set of generators

$$x = x_n, \alpha_1 = x_1 x^{-1}, \ldots, \alpha_{n-1} = x_{n-1} x^{-1}$$

the elements $\alpha_1, \ldots, \alpha_{n-1}$ all lie in $\mathfrak{G}'$ and together with all their conjugates under powers of x they generate $\mathfrak{G}'$. The ideals $u_i = \alpha_i \mathfrak{G}''$ generate $\mathfrak{G}'/\mathfrak{G}''$ as a module over Λ and we have new relations R_i corresponding to the r_i.
If we want to get the presentation matrix for this module we simply have to write down the R_i in the u_i as linear combination and put all these entries in a $(n-1) \times (n-1)$ matrix. The determinant of this matrix then is the desired Alexander polynomial.
A very easy way to write down this matrix is by projecting the oriented knot onto a plane, and label the overpasses by $\xi_1, \ldots, \xi_n$. Now construct an $n \times n$ matrix A by filling all non-zero entries of a column in the following way: choose a crossing and consider the different overpasses which come together in that crossing. Then write in the corresponding column of A:

$1-t$	in place k,
t	in place i,
-1	in place j.

By Proposition 2.8 we know that one of the relations is redundant. Since the number of generators may be limited to $n-1$, we may take any $(n-1)$-minor of A. All these minors are polynomials which generate a principal ideal. The generator of this ideal is the desired Alexander polynomial. Remark that A is not a presentation matrix for $H_1(\tilde{X})$ but for $H_1(\tilde{X}) \oplus \Lambda$. Therefore if we denote the pth principal ideal generated by the $(n-p)$-minors by $\mathfrak{J}_p$, then $\mathfrak{J}_p$ of the former Alexander matrix equals $\mathfrak{J}_{p+1}$ of the latter Alexander matrix.

Note that the dimensions of the two presentation matrices are different: in the former case the genus of the Seifert surface determines the number of closed curves which generate $H_1(\dot{M})$. More precisely, this number is twice the genus of M. The number of columns in the latter matrix is determined by the number of crossings in the chosen regular projection. As we will see in a later chapter, for 2-bridge knots both these numbers can be determined precisely.

2.6.2 The Jones polynomial

The **Jones polynomial** has proved to be a very successful invariant to distinguish certain types of knots and links. To every knot or link k we associate a polynomial in the variable $\sqrt{t}$, which is calculated inductively in the number of crossings according to the following scheme: We first choose an orientation

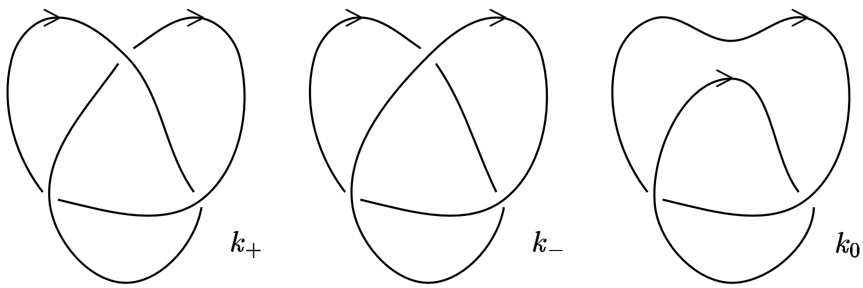


Figure 14: *The three knots needed to compute the Jones polynomial of the positive trefoil*

and project k onto a plane. We can now make a simpler link by choosing a suitable positive crossing and changing the over- in the undercrossing. We denote the first knot k by k_+ and the transformed one by k_-. We can also make a third knot by doing a transformation as in Figure 14. This knot will be denoted by k_0. Obviously, also k_0 has less crossings than k_+. We now calculate the Jones polynomial of $k = k_+$ from the knots k_- and k_0 by the so-called *skein relation*

$$V_{k_+}(t) = t^2 V_{k_-}(t) + t(\sqrt{t} - \frac{1}{\sqrt{t}})V_{k_0}(t)$$

This procedure is done until we are left with n unknots, where n is the number of S^1's from which k is made, i.e. k is an n-link. By convention the Jones polynomial of the unknot is 1.By induction on the number of unknots, we can show that the Jones polynomial of n unknots is

$$V_n(t) = [-(\sqrt{t} + \frac{1}{\sqrt{t}})]^{n-1}.$$

Example Let k be the positive trefoil. After having constructed the knots k_- and k_0, cf. Figure 14, we can write down the skein relation for k and see this to be:

$$V_k(t) = t^2 \cdot 1 + t(\sqrt{t} - \frac{1}{\sqrt{t}})V_{k_0}.$$

If we apply the skein relation once more to $k_0 := k_{0+}$ we obtain

$$V_{k_0}(t) = t^2[-(\sqrt{t} - \frac{1}{\sqrt{t}})] + t(\sqrt{t} - \frac{1}{\sqrt{t}}) \cdot 1$$

which yields

$$V_k(t) = -t^4 + t^3 + t.$$

It can be shown that equivalent knots yield the same polynomial, and one of the simple but important result is that the Jones polynomial distinguishes the positive and negative trefoil by the following theorem.

Theorem 2.27 *Let k be a knot and k^* be its mirror image. Then we have*

$$V_k(t) = V_{k^*}(t^{-1}).$$

□

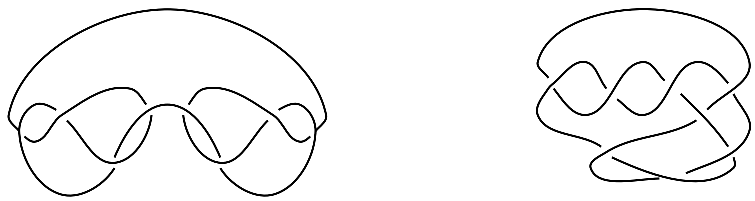


Figure 15: *Two non-equivalent knots with the same Jones polynomial*

A distinction between knots and their mirror images (if non-equivalent) could not be achieved with the other polynomials or other invariants which had been devised, although it's one of the simplest problems one first thinks of. To show that the classification of knots with Jones polynomials is not complete is shown by the two knots below in Figure 15: although they are not equivalent, their polynomials are the same. A paper such as (Kanenobu & Sumi 1992) shows that such a situation are not unique or rare in any case. Here we can find a table of 2-bridge knots with up to 20 crossings with remarks on equivalent Jones polynomials. More often than not two 2-bridge knots of type (α, β) and (α, γ) share the same Jones polynomial. We will study the use of Jones polynomials in a later section.

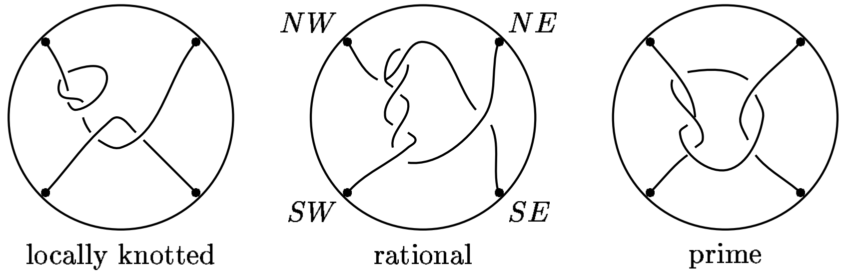


Figure 16: *Different types of tangles*

2.7 Tangle calculus and its relation to knot theory

In this section we will survey a part of mathematics which has specifically been set up to handle site-specific recombination problems. We will first introduce the concept of tangles, and then review some results in the field of topological enzymology.

Definition A **2-string tangle** is a pair (B, t), where B is a 3-ball and t is a pair of arcs (with or without orientation) properly embedded in B.

We let the **trivial tangle** be the pair $(D^2 \times I, \{x, y\} \times I)$, where $I = [0, 1]$, and $\{x, y\}$ are points interior to D^2. We can distinguish three types of tangles (Lickorish 1981):

1. A tangle (B, t) is **rational** if there exists a homeomorphism of pairs from (B, t) to the trivial tangle.
2. A tangle (B, t) is **locally knotted** if there exists a local knot in one of its strands. To be more precise, we can find a 2-sphere in B which meets t transversely in 2 points, such that the 3-ball bounded by the 2-sphere contains a knot.
3. A tangle is **prime** if it's neither rational nor locally knotted.

As in every part of topology we have the classification problem. In order to distinguish different tangles, one first has to define some equivalence relation on the set of tangles. Let first B be the 2-disk D^2, and take as the four points which meet the 2-sphere in which D^2 sits to be four equatorial points $P = \{NE, SE, NW, SW\}$. To get the equivalence relation, we demand that there exists a homeomorphism $\Phi\colon\ (\partial B, \partial t) \longrightarrow (S^2, P)$. So we now think of tangles as being triples (B, t, Φ). Two tangles (B, t, Φ) and (B', t', Φ') are

Figure 17: *Tangle addition $A{+}B$ and construction of denominator and numerator of a tangle A*

called equivalent iff there exists a homeomorphism $H\colon\ (B,t) \longrightarrow (B',t')$ such that $\Phi = \Phi' \circ H$ on ∂B.
To be able to visualize tangles, we would like to be able to draw pictures in the plane. Mathematically one then does the following: Let p be the projection of B onto the equatorial plane D^2, and choose a homeomorphism $\Psi\colon\ B \longrightarrow D^2$, such that Ψ extends to Φ and such that the image of the arcs t under $p \circ \Psi$ is a regular projection in the interior of D^2. A **tangle diagram** is the image of (B,t) under $p \circ \Psi$.
We can define a **tangle addition** in the following manner: Let A and B be two tangles. Define $A{+}B$ to be the object displayed in the figure below. Notice that $A{+}B$ may contain simple closed curves which do not intersect $\partial(A{+}B)$, and therefore $A + B$ doesn't have to be a 2-string tangle. Furthermore for some tangle A we can define the **denominator** $D(A)$ and the **numerator** $N(A)$ as in the Figure 17.

From now on we will investigate rational tangles for reasons which will be made clear a bit later on. We first deal with the classification of these tangles.
We can classify the rational tangles in much the same way as we have classified 2-bridge knots. To do this we have to introduce the same kind of notation. To be more specific we introduce a vector notation and a continued fraction which yield an (extended) rational number (where extended means that ∞ is included in the rational numbers) (Conway 1970, Ernst & Sumners 1987).
Recall that a rational tangle t can be made into the trivial tangle by applying a suitable homeomorphism. Equivalent to this is the fact that the tangle can be deformed by moving the arcs inside the interior of the 3-ball, while keeping the endpoints of the arcs attached to the 2-sphere. Or to put it the other way around, any rational tangle can be made from the trivial tangle by performing the inverse moves of the same procedure.We make the precise statement in the following proposition.

Proposition 2.28 *A rational tangle can be obtained by performing a finite sequence of alternating vertical and horizontal twists to the trivial tangles* (0) *or* $(0,0)$*, where these two trivial tangles are the ones displayed in the Figure 18.* $\square$

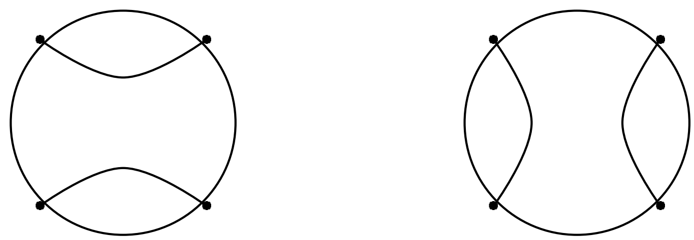


Figure 18: *The two tangles* (0) *and* $(0,0)$

It follows from this proposition that we can completely determine the rational tangles by giving a vector which describes the vertical and horizontal twists which have to be performed from the trivial tangle. We have to distinguish two cases:
If the length of the vector $(x_1, \ldots, x_n)$ is even, we first have to perform x_1 vertical twists to $(0,0)$, then x_2 horizontal ones to (x_1), and in the end we finish with x_n horizontal twists to $(x_1, \ldots, x_{n-1})$.
If the length of $(x_1, \ldots, x_n)$ is odd, we start with x_1 horizontal twists performed on (0), then x_2 vertical twists on (x_1), etc., until we finish with x_n horizontal twists to get the relevant rational tangle.
If the entries $x_1, \ldots, x_n$ all have the same sign, then the regular diagram of the end result is alternating. This is not the case when both signs occur in the vector $(x_1, \ldots, x_n)$. Furthermore we may assume that $x_i \neq 0 \; \forall i \in \{1, \ldots, n\}$, since if we can find an i_0 such that $x_{i_0} \neq 0$ we might shorten the vector to the appropriate size by just removing this x_{i_0}.
Having introduced this vector notation for rational tangles, we can make rational numbers from these vectors in a similar way as in the case of the 2-bridge knot, i.e. by using continued fractions. We recall, that we then define a fraction by setting

$$\frac{\alpha}{\beta} = [x_1, \ldots, x_n] := x_1 + \cfrac{1}{x_2 + \cfrac{1}{\ddots x_{n-1} + \frac{1}{x_n}}}$$

With these preliminaries out of the way we are able to give a full classification of rational tangles:

Theorem 2.29 *There exists a 1-1 correspondence between equivalence classes of rational tangles and the extended rational numbers* $\frac{\alpha}{\beta} \in \mathbb{Q} \cup \{\frac{1}{0} = \infty\}$*, where* $gcd(\alpha, \beta) = 1$. $\square$

A proof can be found in (Burde & Zieschang 1985). Another proof, more elementary than the first, can be found in (Goldman & Kaufmann 1997).

We now investigate the link between rational tangles and 2-bridge knots. There is a nice relation between the two sets, as described by the following theorem.

Theorem 2.30 *We have the following two correspondences between* 2*-bridge knots and rational tangles:*

1. *A* 2*-bridge knot (or link) is the denominator of some rational tangle.*
2. *The denominator of some rational tangle is a* 2*-bridge knot.*

$\square$

The same holds for the numerator of a rational tangle, which will be needed to handle the topological enzymology problems. A nice proof of this theorem can be found in (Murasugi 1996). It involves the same argument with which one proves that a 2-bridge knot can always be projected in such a way that, regarded as a 4-braid, one of its strings doesn't contain any crossings.

3 DNA recombination

In this section we discuss the relevant biochemical background needed for this paper. We discuss some facts about the structure of DNA first.

3.1 The structure of DNA molecules

The DNA molecule is composed of four bases, Adenine (A), Cytosine (C), Guanine (G) and Thymine (T), which are attached to one or two backbones of alternating sugars and phosphorus. In the case of one backbone we call the DNA single-stranded, and in the other case double stranded. We will focus on the latter case. At each site on the two backbones, bases are attached, such that an A in the first backbone fits to a T in the other, and there is a similar pairing of C's with G's. The bases A,T,G and C are covalently bonded by hydrogen-bridges. This makes a double-stranded DNA molecule an immense long word written in A, T, G and C, which can be read in the inverse on the other strand. This ladder of base-pairs is according to the classical Crick-Watson model a right-handed helix, and the two backbones are intertwined millions of times, with approximately 10.5 basepairs per full twist. To be able to store such huge molecules in the nucleus of a cell, it has to be made more compact. The helical form of double-stranded DNA solves this problem: in order to minimise the energy of the intertwined strings, the molecule will supercoil, much the same as a telephone wire will supercoil if you twist the wire every time you have used the phone. The millions of twists in DNA molecules will yield four to five levels of supercoiling. The molecule has to be coiled up in a cell much like 200 km of fishing line would have to fit in a basketball!

To see whether this neat double stranded helical structure has any implications on its functionality we have to review the four most important processes needed to sustain life: replication, transcription, repair and recombination. We explain these terms very briefly.

Replication is the process where the two strands of a piece of DNA are split at the *replication fork*, and new strands are made at each of the two strands with two complete identical DNA pieces as result. The process in which the genetic information on the DNA is being 'read' and used to make proteins we refer to it as *transcription*. DNA *repair* has to be done since both replication and transcription tend to be quite 'sloppy processes', i.e. many mistakes are made. *Recombination* is the interchanging of pieces of DNA, be it within one piece or between several. Here one generally makes a distinction between interchanging of almost entire chromosomes, referred to by *homologous recombination*, and relocating small bits of genetic information which is called

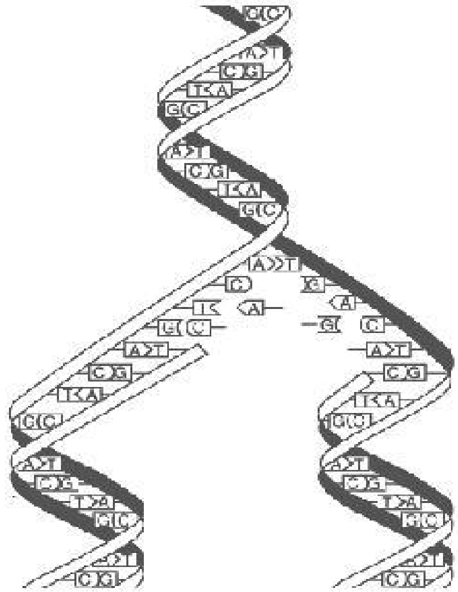


Figure 19: *Replication of double stranded DNA*

site-specific recombination.
Being heavily supercoiled does not make these tasks much easier. In order to deal with the problem several enzymes have evolved to mediate these important processes of living matter. For instance, to replicate DNA which occurs in circular form, as all DNA, it has to be locally unwinded. But since the twists cannot really be removed by unwinding due to the circular shape of the molecule this gives additional stress to the rest of the molecule. To relieve this stress a group of enzymes, called **topoisomerases**, break the strands and let them unwind as much as needed. For the types of DNA we will discuss, this will not solve the problem completely. To discuss this matter appropriately is beyond the cope of this paper. The whole story can be found in (Mathews & Van Holde 1996).
In site-specific recombination events we also find difficulties arising from the geometrical properties of DNA. Here either two blocks of DNA are interchanged or a piece of DNA is integrated in a host genome. We will see concrete examples of both processes, in the form of phage λ integrase (Int) and Tn3-resolvase (Tn3).[3]
Since site-specific recombination is from the topological point of view the

[3]The recombination enzymes which have been discussed in the literature are Hin, Gin, Cre, Tn3 and Int. For the first three, which will not be dealt with in this paper see on the Cre enzyme: Abremski *et al.* 1986; on the Hin enzyme: Heichman & Johnson 1990; on the Gin enzyme: Kanaar *et al.* 1988, 1990.

most interesting process we will give a more detailed account of its properties and in the subsequent sections.

3.2 Circular DNA and site-specific recombination

Although in most textbooks DNA is visualised in linear fashion, a substantial part of DNA in living cells is found in circular form. They are formed from the linear ones by gluing together both ends of the double-stranded string. Topologically the result is a strip which is winded a great number of times. A priori one would expect a mixture of orientable and non-orientable bands in samples of DNA-circles. In nature only the orientable bands are found. To explain this phenomenon, we have to look a bit better at the precise structure of DNA-backbones.

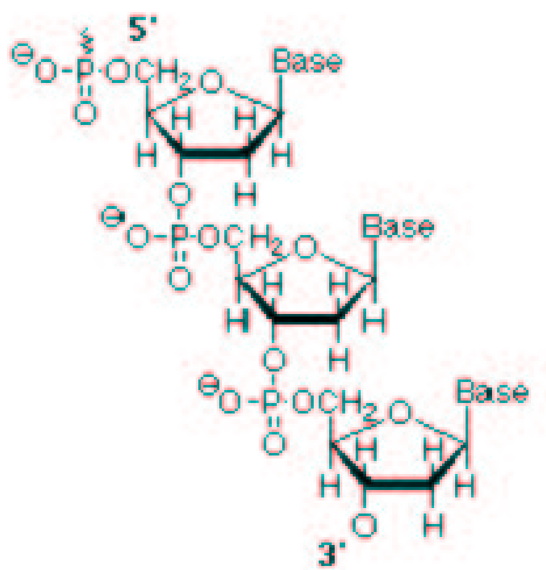


Figure 20: *3'-ends and 5'ends of the backbone of DNA*

Recall that the backbone of DNA strands are alternating sequences of sugars and phosphorus. The sugars are rings of four carbon atoms and one oxygen atom. If we number the ring positions from one to five, by convention the first position is a C-atom with a base attached to it. Then we find three more C's and the fifth position is occupied by the O-atom. At position four, a CH_2-group, and the phosphate-group are attached in linear fashion. This phosphate-group then is attached to the sugar-phosphorus-group below it,

to the 3'-carbon in the ring. See Figure 20. In effect, the 3'-carbon and the fifth C from the CH_2 (denoted by 5'-carbon, although it's not in the ring) are the links to the sugar-phosphorus-groups below and above.
They are the key to the orientability problem of the circular DNA: if we look at a straight segment of double-stranded DNA, we will find that at the ends we have an anti-symmetrical situation: the two strands both terminate with 3'-C's or with 5'-C's, but by the explanation above a 3'-C will only fit to a 5'-C with a phosphate-group inbetween them. Now the problem is solved by the observation that one string ends with a 3'-C on one side and a 5'-C on the other, and both strands' ends are inverse to each other. So if we find a 3'-C of one strand at one side we will find a 5'-C of the other strand at the same side. Now the two ends will only fit together to form a circle if the number of twists of the two strands is even, since we can only bind 3'-C's to 5'-C's. This will induce the orientability needed to explain the phenomenon.

Having settled this little problem we can discuss the mechanisms of site-specific recombination of circular DNA. Characteristic of these processes is the local nature of the events when it comes to gene-exchange, whereas the global geometry of the DNA has to be altered before this can take place. We will first explain the type of recombination involved in Tn3-resolvase.
In this situation we have two recombination sites on the circular DNA, which are identical pieces of DNA. The action performed by the enzyme, is breaking the DNA at the ends of these sites (after having given the DNA an orientation), and crossing the two sites to exchange them. Two situations can occur: the recombination sites have the same orientation or the opposite on the oriented DNA. The first case is called **direct repeats** and the latter **inverted repeats**. We now look at the recombination event in a bit more detail.
To exchange the two recombination sites on the DNA-circle, the enzymes has to align them in order to perform the action. The enzyme has to perform a global move on the DNA to achieve this, which results in bringing the sites close together to be able to exchange them. The end of this move is called **synapsis**, and this intermediate stage of enzyme-DNA-complex is called the **synaptic complex**. The part of the DNA which is really bound to the enzyme, including the sites, together with the enzyme is called the **synaptosome**. Having formed this synaptic complex, the enzyme performs the recombination and lets the DNA go. In nature we find that enzymes may perform recombination more than once. Either the action is performed several times before releasing the DNA, an event which is called **processive recombination**, or the enzyme may act in multiple encounters, which is called **distributive recombination**. Both forms may be found for the one enzyme.

It is evident that the recombination events induce topological changes to the DNA-substrates (i.e. the DNA which will be recombinated). One way of trying to understand the actions which are performed by an enzyme is just looking at the event under a microscope (or take pictures of the stages). This is possible to some extent, as is illustrated by the photograph below, but the crucial part where the actual exchange of sites takes place can not be revealed with this method. The synaptosome is only visible as a small blob, and no further analysis of it can be made. A rather recent development to attack the problem is the **topological approach to enzymology** introduced by Wasserman & Cozzarelli (Wasserman & Cozzarelli 1986). In this protocol data from biochemical analyses of DNA recombination events are translated into topological facts, and the machinery from topology is then applied to analyse them. In this indirect manner one can learn more about the actual moves the enzyme has to perform to achieve the recombination. In the next section we will discuss the first case of site-specific recombination: that of Tn3-resolvase.

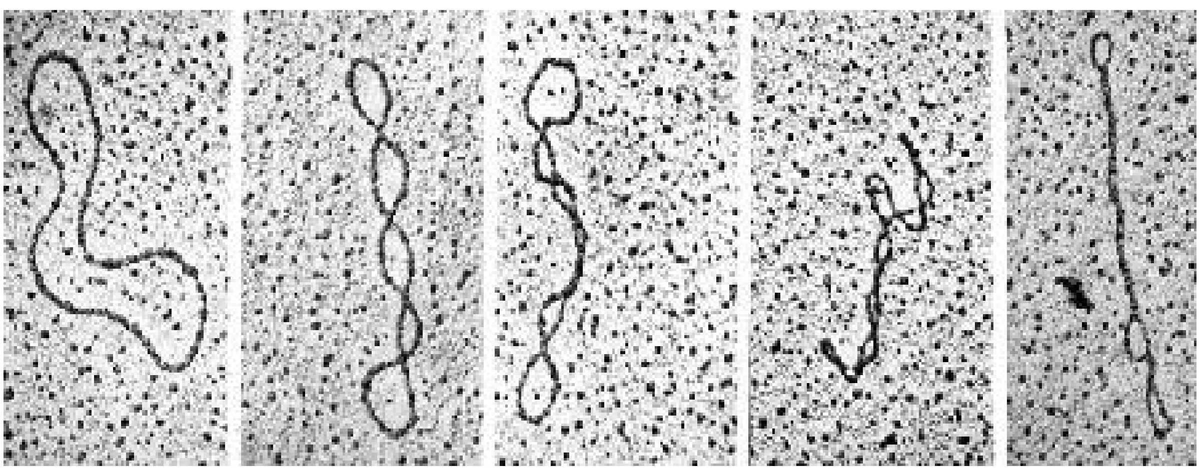


Figure 21: *Supercoiled circular DNA*

3.3 Tn3-resolvase and its topological behaviour

In the early 1990's C. Ernst and D.W. Sumners have made a calculus for rational tangles to investigate the problem of enzyme activity (Ernst & Sumners 1990, Ernst 1996, Ernst 1997). The idea is as follows: to investigate the topological performance of a particular enzyme a circular DNA substrate is incubated with such an enzyme. The reaction products are then analyzed by a method called gel electrophoresis. During gel electrophoresis the reaction products can be analyzed by size and shape. An electric current pulls the negatively charged DNA slowly through through agarose or polyacrylamide (the gel), which contains a microscopic network of pores. The speed with which the molecules are being transported through the gel depends on the structure of the DNA. More compact (supercoiled) DNA, will experience less friction, and will go faster than voluminous DNA. In a similar manner products with high catenation are faster than trivial knot DNA molecules. After some time the solution will be divided into neat strips in which the molecules of the same form are accumulated. One then coats the individual strips with a protein called *recA*, which relaxes the DNA, in order to make single molecules visible under the microscope. See Figure 22 for examples. This yields the biological data needed to get the mathematical model started.

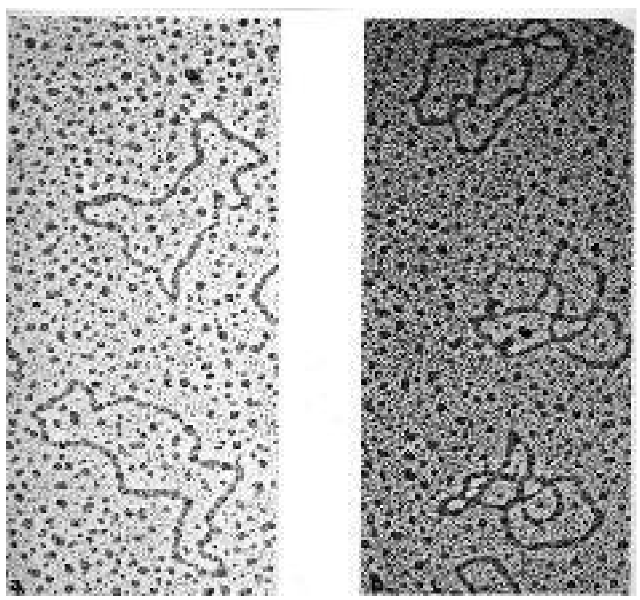


Figure 22: *DNA knots in vitro*

In several studies biochemists have analyzed the structure of reaction products made by Tn3-resolvase.

Tn3-resolvase is an enzyme which reacts circular duplex DNA substrates with directly repeated recombination sites (Wasserman & Cozzarelli 1985a, Wasserman *et al.* 1985). In most cases resolvase will mediate one round of recombination and releases the product knot. This is the principle product of the reaction and is known as the Hopf link $< 2 >$. One in twenty encounters will produce more complex DNA links and knots: additional rounds of processive recombination are mediated in these cases and subsequent products are found after each round. The sequence of knots is

$$< 1 > \implies < 2 > \implies < 2,1,1 > \implies < 1,1,1,1,1 > \implies < 1,1,1,2,1 > .$$

Observe that all product knots are 2-bridge knots. Therefore, by Theorem 2.30 we know that there are rational tangles such that their numerator or denominator are the known 2-bridge knots. We now review a model for site-specific recombination from (Ernst & Sumners 1990) which enables us to set up the required set of tangle-equations.
By the local nature of site-specific recombination events, we may assume that we may subdivide the substrate knot in two parts. More specifically, we can find rational tangles O and T such $N(O+T)$ equals the unknot. The O tangle is the *outside* tangle, and the T tangle refers to the recombination sites on the substrate unknot. We now suppose that the enzyme performs a **tangle surgery**, i.e. it removes T and substitutes a new tangle, R say. Then the product knot is equal to $N(O+R)$ and we already know this to be the Hopf link. Multiple rounds of processive recombination are modeled by tangle additions of the identical reaction tangle R. Thus the second round will yield a product knot $N(O+R+R)$ and so on. We may now set up a set of equations with O, T and R as unknowns:

$$\begin{aligned} N(O+P) &= < 1 > \\ N(O+R) &= T_1 \\ N(O+R+R) &= T_2 \\ &\vdots \\ N(O+mR) &= T_m \end{aligned} \tag{2}$$

where the T_i are reaction products known from experiments. Ernst and Sumners have devised a calculus to solve such equations in more general context than Tn3-resolvase and have established the following theorem, proved in (Ernst & Sumners 1990):

Theorem 3.1 *Suppose that tangles O, T and R satisfy the following equations: (i) $N(O+T) = < 1 >$; (ii) $N(O+R) = < 2 >$; (iii) $N(O+R+R) = < 2,1,1 >$; (iv) $N(O+R+R+R) = < 1,1,1,1,1 >$. Then $\{O,R\}$ is $\{(-3,0),(1)\}$ and $N(O+R+R+R+R) = < 1,1,1,2,1 >$.* □

One might ask how many rounds of processive recombination are needed to uniquely determine the rational tangles O, T and R. This has been answered by Ernst to be maximally four (Ernst 1997).

With the Theorem the analysis of the Tn3-resolvase enzyme is complete. We precisely know what the topological moves are performed by the enzyme to perform the recombination: first it makes a triple twists in the unknot to align the recombination sites, and performs a recombination. In further steps Tn3 either introduces 1 twists before recombination (from $< 2 >$ tp $< 2, 1, 1 >$) or merely introduces one extra crossing (all other steps).

3.4 Torus knots and DNA recombination

In this section the applications of torus knots in DNA recombination are discussed. We will first review results which have been obtained in the phage λ integrase recombination experiments.

Recall that site-specific recombination has mainly been studied in two classical examples, Tn3-resolvase and phage λ integrase (Int) recombinations. Int-mediated reactions are far more versatile than Tn3-recombinations. In concert with other proteins, Int can recombine direct and inverse repeats in circular or linear substrates, supercoiled or not. We will simplify the situation a bit to make it more accessible. For those who want to read a more detailed account, see (Crisona *et al.* 1999).

Rather than performing an intramolecular exchange of genes, Int recombinates the DNA between two molecules, thereby integrating the DNA of the bacteriophage λ, a virus, into the genome of the bacteria *Escherichia coli.* The genes which are identified by Int on *E. coli* and on phage λ are called $attB$ and $attP$ respectively. $attB$ is some 240 base pairs (bp) long, while $attP$ measures only 30 bp. On $attB$ we can find the precise sites where Int will break the DNA to perform recombination. They will be denoted by C and C'. Analogously, the sites on phage λ will be called B and B', which contrary to *E. coli*'s $attB$ happen to be the very ends of the $attP$. The regions $attB$ and $attP$ are both parts of circular DNA-sequences. We denote $attB$ by $---C-O-C'---$ and $attP$ by $B-O-B'$. The actual crossover occurs between the homologous core regions, denoted by O. Now the first step involves making two new gene fragments called $attL$ and $attR$. $attL$ consists of $B-O-C'---$ and $attR$ of $---C-O-B'$. In a diagram we have

$$\left.\begin{array}{ll} attB & ---C\text{–}O\text{–}C'--- \\ attP & \qquad B\text{–}O\text{–}B' \end{array}\right\} \Longrightarrow \left\{\begin{array}{ll} \qquad B\text{–}O\text{–}C'--- & attL \\ ---C\text{–}O\text{–}B' & attR \end{array}\right.$$

An illustration of the recombination is given in Figure 23.

The integration of the phage λ is called the **Int PB** reaction. Contrary to the integration we can also distinguish the reverse reaction, which excises the phage λ genome out of the bacteria. This reaction is denoted by the **Int LR** reaction. We will focus on the Int PB reaction.

The purpose of the integration is solely benefiting the virus: it cannot replicate its own DNA. When the bacteria with integrated virus DNA replicates, automatically the virus DNA will be copied as well, and these will then form new viruses, making the circle of life complete.

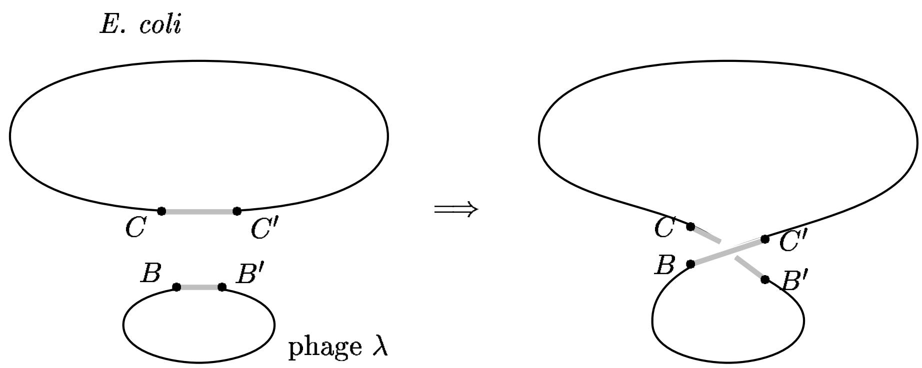


Figure 23: *The Int mediated recombination*

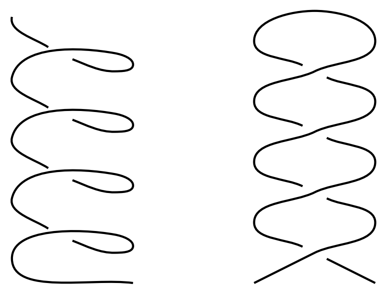


Figure 24: *Helical vs. plectonemic supercoiling*

The genome of the bacteria is circular, so the tangle calculus may be applied to study the topological characteristics of the Int enzyme. Due to the great variety of possible reactions, it has been difficult to unravel the precise mechanisms responsible for the experimental observations. A lot of work has been done on the subject, culminating in a paper which has just been published this year (Crisona *et al.* 1999). We will give a short tour through the experimental results. These can be reviewed in (Spengler *et al.* 1985, Cozzarelli *et al.* 1984, Crisona *et al.* 1999).

To be able to examine the topological activity induced by the Int PB reaction, one doesn't look at the original bacteria-phage complex, but instead one takes ordinary circular substrate, with the appropriate binding sites. The actions then are directly translated into knotting of the substrates. We will now review the results.
The substrates have been plectonemically supercoiled, which is showed in the figure below.

The reaction products where then run through a gel, and the separated knots and links where coated with *recA* protein to relax the superhelical structure and allow electron microscopy. This way a list was made with all knots and links occurring in the product solution. The list with all the knots is given below.

Knots made by Int-mediated recombination

$c(k)$	number of molecules	number of knot types	observed type
3	1	2	Torus
5	2	4	Torus
7	5	16	Torus
9	5	122	Torus
11	4	1,527	Torus
13	7	20,992	Torus
15	7	300,000	Torus
17	2	3,000,000	Torus
19	2	50,000,000	Torus

From: Spengler *et al.* 1985

A similar kind of table can be set up for the links, but no precise data can be found in the literature. It is known, however, that a similar distribution of torus links is found in recombination products (Spengler *et al.* 1985, Crisona *et al.* 1999). We make some remarks on these distributions.

In the second column of the table we have placed the number of molecules observed in the reaction product. We can directly see that the distribution of the number of molecules over the crossing number of the knots is quite different than with Tn3-experiments. Recall that in that case, we saw a great decrease in numbers when crossing number increased, which reflected the fact that there were less DNA-products resulting from multiple processive recombination. But in the Int-case we see, that even for higher crossing numbers there are still quite a few product molecules with that crossing number. Another observation is the consistent appearance of torus knots, which is most remarkable in the case of 19 crossings, since there are some 50 million knot types with 19 as their crossing number, of which only one is a torus knot.[4] We can conclude from these observations, that we have a different mechanism for Int-recombination compared with Tn3-mediated processes: It is shown in (Spengler *et al.* 1986) that the plectonemic supercoiling of the substrate contributes directly to the product knot. If a substrate molecule is supercoiled a

[4]That there is only one torus knot among the 50 million is seen using Prop. 2.22, and the fact that 19 is prime

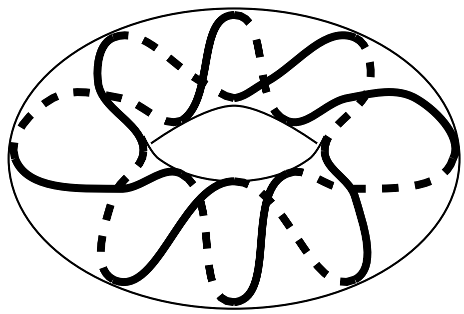


Figure 25: *A torus knot of type (2,8)*

number of times, and the 'ends' of the molecules are recombined (see the figure for explanation), then the result is a torus knot of type $(2, 2n+1)$, where n runs over the positive integers (why they have to be positive is explained by looking at the orientations of the crossings, see below).
Similar results were obtained by (Crisona *et al.* 1999).

In the literature nothing can be found on processive recombination in Int-mediated reactions. This is probably due to the complexity of the reactions, which will probably not reveal any side products between the array of different 'first order' products that can be found, through gel electrophoresis. Another reason might be, that since Int is certainly not the only enzyme at work in the reactions, multiple recombinations will not be very probable to happen. A third reason for the lack of processive recombination might be the fact that supercoiling is needed for Int to recombinate. Once products like torus knots have formed the supercoiling can not occur any more, at least not on the same level. If one would like to coil a torus knot, it's the torus in which the knot can be embedded that will be coiled, whereas in the case of supercoiled substrate, the strands will be coiled. This will probably hinder another round of recombination.

Some mathematical remarks on the Int-reactions.

There are two classes of torus knots which are found in the Int PB reaction. For knots we find $t(2, 2n+1), n \in \mathbb{N}$, and for links we find the torus links with an even number of crossings.[5] The crossings do not have equal signs in both cases: recall that Int PB needs negatively supercoiled substrate in order to perform its action. When the number of supercoils is $-2n, n \in \mathbb{N}$,

[5]In (Crisona *et al.* 1999) the list of links found consists with 2-links with ≥ 4 crossings. For clarity of mathematical exposition we include the 2-link $< 2 >$.

the recombination product will be a torus knot $t(2, 2n-1)$ which has *positive* crossings instead of negative. On the other hand, from a $-2n+1$ times supercoiled substrate is recombinated, a $-2n$ link with *negative* supercoiling is produced. This is explained by following orientations during recombination.

The bridge number of the torus knots $t(2, 2n+1), n \in \mathbb{N}$ is by Prop. 2.22 equal to 2 whenever $n > 1$, which is always the case. For the links we can also see this to be true. We can give a general description of all the product knots and links that have been found.
If we look at the set of 2-bridge knots of the form $\{< n > | n \in \mathbb{Z}\}$, then it's easy to see that the set $\{< 2i+1 > | i \in \mathbb{N}\}$ represents the subset of the torus knots found in the Int PB reaction. The subset of positively even-crossed 2-links $\{< -2i > | i \in \mathbb{N}\}$, equals the subset of links. In other words, the 2-links found in the reaction products are also 2-bridge knots. This shows that as in the case of Tn3-resolvase, all the products are a subset of the 2-bridge knots.

Although claimed in (Spengler *et al.* 1985) that torus knots always have an odd crossing number, this is not the case. Prop. 2.22 lets us construct a torus knot $t(4, 3)$ with crossing number $\min\{3 \cdot 3, 4 \cdot 2\} = 8$. The torus knots that are found in Int-mediated recombination products do have odd crossing number.

Proposition 2.14 lets us compute the genus of the different knot and link products. For the torus knots $\{< 2n+1 > | n \in \mathbb{N}\}$ we have genus

$$\frac{1}{2}(2n+1-1) = n,$$

and for the torus 2-links $\{< -2n > | n \in \mathbb{N}\}$ we have genus

$$\frac{1}{2}(|-2n| - 2) = n - 1.$$

Unfortunately Int does not perform multiple rounds of recombination, unlike Tn3. In the next chapter we try to learn more about site-specific recombination. Because of the rather simple nature of Int reactions, we will focus on the Tn3-resolvase mechanism.

4 The mathematical structure of DNA knots in site-specific recombination events

In this section we try to expand our knowledge of site-specific recombination by Tn3-resolvase by investigating the mathematical structure of these processes. In the previous section we have seen that the models set up to investigate site-specific recombination are based on the assumption that recombination is performed processively. That means multiple rounds of recombination are performed on a substrate before it is released by the enzyme. We will discuss both distributive and processive recombination events.

If we look at a single step in the process of site-specific recombination we know from the biological and mathematical results reviewed in previous sections that both substrate and product knots are 2-bridge knots. This provides motivation to consider the questions: given some 2-bridge knot k, what are the possible forms of the product knot made by recombinating k? How many possibilities are there, and does the form give us any new information on where the enzyme may perform its actions? These matters are considered in section 4.2.

We will also see some results in which we perform multiple rounds of site-specific recombination. If we regard all recombination steps in a certain sequence to be the same, we are discussing distributive recombination events, since this is the situation in which the enzyme makes one recombination and lets the product knot go every single step. If we on the other hand regard the first step as a 'preliminary and global' move and all further events as 'local', we are in the field of processive recombination. As we will see, the latter is a simplified version of the first situation. this means that processive recombination events may be modelled as 'trivial' distributive recombination events.

In sections 4.3 and 4.4 we try to answer how the knot invariants change if we perform a site-specific recombination. Do these changes tell us anything about the mechanism? This leads to discussions on four well-known invariants and study their application on site-specific recombination events. First the genus and the graph of a knot will be conidered. After that we look into the application of the Alexander and Jones polynomials.

First we look at our model of site-specific recombination more accurately.

4.1 A more general model for site-specific recombination

As we have seen in the previous chapter the topological characterisation of Tn3-resolvase has been determined satisfactorily. Theorem 3.1 provides us the with specific information required to make a generalisation. The Theorem states that the unique pair of tangles $\{O, R\}$ such that the equations are satisfied is $\{(-3,0),(1)\}$. The first tangle is a negative twist of length 3 of two strings, and the second tangle is the recombination of two strings. The twists are made in the 'outside' tangle and are a global move of the substrate. They are needed to align the two recombination sites, which are then recombinated by the enzyme. This last step is seen as the substitution of a tangle T by a new one $R = (1)$. This also implies that T is the trivial tangle (0). The complete action can thus be divided in two steps: a global move of 3 twists and a local recombination of the sites inside the enzyme. More generally we model one step in the site-specific recombination event by n twists and one recombination.

Note that this model for site-specific recombination based on the Tn3 enzyme also covers the topological moves performed by the Int enzyme. Therefore, apart from the fact that Tn3 performs processive recombination while Int does not, their first recombination step is the same. Of course, Tn3 really fixes the number of twists before it performes the recombination, and this aligning of sites is done actively by the enzyme. With Int on the other hand we find that the number of twists may vary. This is due to that fact that aligning of the sites is done passively: by random collision. Different substrate knots just have a different amount of supercoiling. We proceed our discussion on processive recombination.

Processive recombination is seen as addition of multiple R tangles as discussed before. This can also be regarded a bit differently. If we perform no twists and only make one recombination, we have actually done nothing else than one step in a processive recombination event (save the initial step). This means that processive recombination can be considered as a special case of distributive recombination, i.e. such that the first step may contain non-trivial twists but all consecutive steps do not.

We now briefly investigate if there are any constraints to this model. More specifically, is it always possible to add a crossing from a recombination in a processive sequence of recombinations and write this in a form $N(A+B)$? If the new crossings is added on the 'outsides' of the knot (i.e. left of A_1 and right of A_{2m+1}) this is seen to be easy. But if we add such a crossing in the interior of the knot, there doesn't seem to be enough space to draw closed curves as the boundaries of the two tangles A and B to form the $N(A+B)$

complex. But if we regard the tangles as three-dimensional objects with S^2's as boundaries one may see that the recombination tangle can be lifted above the rest of the 2-bridge knot, and we are able with this construction to write the entire 2-bridge knot as the numerator of the sum of A and B.
As we have seen processive recombination is quite the same as distributive recombination: they differ only in the fact that in each step after the initial one the latter may have non-trtivial twist whereas the first may not. In this perspective processive recombination is a more restricted form of distributive recombination.
With these preliminaries out of the way we start with the actual discussion on classification of DNA knots in site-specific recombination events.

4.2 A classification of DNA knots in site-specific recombination events

We investigate the question: given a 2-bridge knot $k =< a_1, \dots , a_{2m+1} >$, what will be the product of the enzyme action of the type of Tn3-resolvase? To deal with the question, we introduce some additional notation.

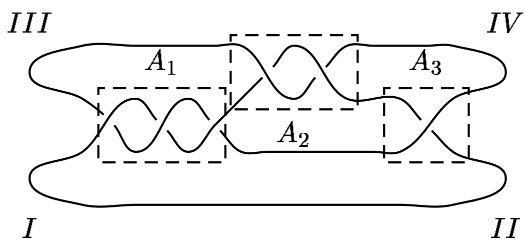


Figure 26: *The different regions of the 2-bridge knot* $< 3, 2, 1 >$

The region of k which contain the crossing of length a_i, will be denoted by A_i. Regarding k as a 4-plat coming from a suitable 4-braid, we have four semi-circles: one above and one below A_1, joining the two left pairs of strings, and one above and one below A_{2m+1}, joining the other two pairs (cf. Fig. 8). We denote the lower two by I and II resp. and the upper two by III and IV. See Figure 26 for an example. The set of 2-bridge knots will be denoted by $\mathcal{K}$. The enzyme action p_n is modelled by n twists and 1 recombination. We have the following theorem.

Theorem 4.1 *Let* $k = < a_1, \dots , a_{2m+1} >$ *be a 2-bridge knot, and* $p_n \colon \mathcal{K} \longrightarrow \mathcal{K}$ *be a map which performs n twists and* 1 *recombination on k. Then the possible product knots $p_n(k)$ are:*

$$< a_1, \dots, a_{2m+1}, n, 1 >,$$
$$< a_1, \dots, a_{2m+1}, n-1, 1 >,$$
$$< 1, n, a_1, \dots, a_{2m+1} >,$$
$$< 1, n-1, a_1, \dots, a_{2m+1} >,$$
$$< a_1, \dots, a_{i-1}, j, \pm 1, a_i - j, \dots, a_{2m+1} >,$$
$$< a_1, \dots, a_i \pm 1, \dots, a_{2m+1} >$$

We prove the Theorem by asserting that these are product knots under p_n-action on k and furthermore that there are no other possibilities. We make two distinctions: either the action of n twists, with $n \geq 1$, is performed somewhere between the regions A_i, $1 \leq i \leq 2m+1$, and end-regions included, or outside the crossing regions (so at the far ends of the knot). Since the substrate is the unknot and since the enzyme will always perform either positive or negative twists, and not both during distributive recombination, we can fix the orientation of the twist after we know whether they are left-twists or right-twists (where right and left are determined in the usual way). We therefore may assume that twists are always positive, and the signs of the crossings and of the twists are the same. We don't have any information however on the sign of the recombination event and have to consider both negative and positive cases.

We begin with a lemma.

Lemma 4.2 *If the enzyme performs a non-trivial twist and a recombination between A_1 and A_{2m+1}, then the product is not a 2-bridge knot.*

The proof of this Lemma will be postponed to section 4.3.

PROOF OF PROPOSITION We first consider a trivial twist action and one recombination between A_1 and A_{2m+1}, end-regions included. We make a distinction between positively and negatively oriented recombination w.r.t. the orientation of the crossings. We denote this action by p_0^+ and p_0^- resp. where 0 denotes the trivial twist.
We have two different situations when recombination occurs between the first and second string as opposed to recombination between the other two pairs. We can divide these situations in recombination above A_i (for i odd) and below A_i (for i even), and between two A_i and A_{i+1}, again between the middle two strings or between the upper pair. We deal with actions 'inside' the regions A_i for i odd and positive orientation of the recombination to illustrate the idea:

The recombination site right above a region A_i divides A_i into two pieces of j and $a_i - j$ crossings for some $0 < j < a_i$. So we have

$$p_0^+(< a_1, \dots, a_{2m+1} >) = < a_1, \dots, a_{i-1}, j, 1, a_i - j, \dots, a_n > .$$

or more generally

$$p_0^\pm(< a_1, \dots, a_{2m+1} >) = < a_1, \dots, a_{i-1}, j, \pm 1, a_i - j, \dots, a_{2m+1} > .$$

For recombination next to some region A_i we have in an analogous fashion:

$$\begin{aligned} p_0^+(< a_1, \dots, a_{2m+1} >) &= < a_1, \dots, a_i + 1, \dots, a_{2m+1} >, \\ p_0^-(< a_1, \dots, a_{2m+1} >) &= < a_1, \dots, a_i - 1, \dots, a_{2m+1} >. \end{aligned}$$

For recombination with n twists outside A_1 or A_{2m+1}, we can immediately state the result for enzyme actions at III and IV:

$$\begin{aligned} p_n^{III,+}(< a_1, \dots, a_{2m+1} >) &= < 1, n, a_1, \dots, a_{2m+1} >, \\ p_n^{IV,+}(< a_1, \dots, a_{2m+1} >) &= < a_1, \dots, a_{2m+1}, n, 1 >. \end{aligned}$$

and for negatively oriented recombination

$$\begin{aligned} p_n^{III,-}(< a_1, \dots, a_{2m+1} >) &= < 1, n-1, a_1, \dots, a_{2m+1} >, \\ p_n^{IV,-}(< a_1, \dots, a_{2m+1} >) &= < a_1, \dots, a_{2m+1}, n-1, 1 >. \end{aligned}$$

For actions at I and II, we turn to a different notation for k. For non-canonical 4-plats, we can have crossings between the first and second strings, between the second and third strings, and between the third and fourth strings. We denote these with a_i, b_i and c_i respectively, and write

$$k = (a_1, \dots, a_k | b_1, \dots, , b_{k+1} | c_1, \dots, c_k).$$

By Murasugi, p.183 - 186, we can transform D into

$$(0, 0, \dots, 0 | b_1, \dots, b_{k+1} | a_1 + c_1, \dots, a_k + c_k).$$

With our previous notation we can identify

$$< a_1, \dots, a_{2m+1} >= \ (0, \dots, 0 | a_1, a_3, \dots, a_{2m+1} | a_2, a_4, \dots, a_{2m}).$$

We have

$$p_n^{I,+}(< a_1, \dots, a_{2m+1} >) = (n, 0, \dots, 0 | 1, a_1, a_3, \dots, a_{2m+1} | 0, a_2, a_4, \dots, a_{2m})$$

which is equivalent to

$$(0, 0, \dots, 0 | 1, a_1, a_3, \dots, a_{2m+1} | n, a_2, a_4, \dots, a_{2m}) = < 1, n, a_1, a_2, \dots, a_{2m+1} >$$

which equals $p_n^{III,+}(< a_1, \dots, a_{2m+1} >)$.
For negatively oriented recombination we get

$$p_n^{I,III,-}(< a_1, \dots, a_{2m+1} >) =< 1, n-1, a_1, a_2, \dots, a_{2m+1} > .$$

Similarly we find

$$p_n^{II,+}(a_1, \dots, a_{2m+1} >) = \ (0, \dots, 0, n|a_1, a_3, \dots, a_{2m+1}, 1|a_2, a_4, \dots, a_{2m}, 0),$$

which is equivalent to

$$< a_1, \dots, a_{2m+1}, \pm n, 1 > = \ p_n^{IV,+}(< a_1, \dots, a_{2m+1} >).$$

and an analogous result for $p_n^{II,IV,-}$. In conclusion we have

$$\begin{aligned} p_n^{I,III,+} &= \ < 1, n, a_1, a_2, \dots, a_{2m+1} >, \\ p_n^{II,IV,+} &= \ < a_1, a_2, \dots, a_{2m+1}, n, 1 >, \\ p_n^{I,III,-} &= \ < 1, n-1, a_1, a_2, \dots, a_{2m+1} >, \\ p_n^{II,IV,-} &= \ < a_1, a_2, \dots, a_{2m+1}, n-1, 1 >. \end{aligned}$$

□

Theorem 4.1 allows us to gain insight in the complexity of the product knots in Tn3-resolvase mediated reactions. Here complexity is defined to be the total number of crossings in the canonical representation of the 2-bridge knot, or crossing number of the knot.[6] We distinguish the two cases of processive vs. distributive recombination. In processive recombination we see that in the first step up to $n+1$ crossings are added (depending on signs of both topological moves), and in each next step only one crossing is added. In the other case we add up to $n+1$ crossings in each step. Therefore we have proved

Proposition 4.3 *Let $k =< a_1, \dots, a_n >$ be a 2-bridge knot, and $p_n \colon \mathcal{K} \longrightarrow \mathcal{K}$ be a processive resp. distributive enzyme action of n twists and one recombination on k. Then for the complexity of the product knot after q rounds of recombination we have*

$$\begin{array}{ccccc} 0 & \leq & c(p_0^{q-1}(p_n(k))) & \leq & n+q, \\ 0 & \leq & c(p_n^q(k)) & \leq & q(n+1), \end{array}$$

respectively. □

[6]According to (Soteros *et al.* 1992) this measure is a 'good' measure to study knottedness. In their article the authors have introduced some properties which any real-valued function on the set of knots should have in order to measure complexity. One of the functions which has these properties is the total crossing number.

The theorem invites more investigation in the question which transformations preserve 2-bridge knots? Since we have a canonical representation of 2-bridge knots, we can try to solve the question using the vector notation of the 2-bridge knots. Given a knot $< a_1, \ldots, a_n >$ the transformed knot must also be of the form $< x_1, \ldots, x_m >$. In site-specific recombination processes we find three different reactions. The first gives extra complexity by either making n twists and performing one strand exchange, as we have seen in Tn3-resolvase mediated reactions. A second reaction recombinates ends for integration of a piece of DNA which was discussed in the Int PB action. The last reaction does the opposite of the previous one, i.e. it excises DNA from the (already integrated) DNA-knot. An example is the Int LR action. So for our discussion we restrict ourselves to transformations which preserve the greater part of the molecule, and only perform local operations. To put this mathematically, only few of the entries of the vector will change and most will stay the same. All transformations then either yield additional entries, in a way that they preserve the canonical representation, or they delete entries from the original knot. By Theorem 4.1 we know that new entries can only be put on the outsides of the knot (w.r.t. the canonical representation) in blocks of two successive regions with crossings A_i and A_{i+1}, to preserve 2-bridge structure, while excision can only be done in blocks of two in the interior of the knot (between A_1 and A_n).[7] Therefore we've proved the following result:

Proposition 4.4 *Let $k =< a_1, \ldots, a_n >$ be a 2-bridge knot, and $p\colon \mathcal{K} \longrightarrow \mathcal{K}$ be a map which performs local recombination actions as described above. Then $p(k)$ is made from k by excision or addition of pairs of blocks, where addition is done on the outside of the regular diagram of k, and excision in the interior.* □

4.3 The genus of the handlebody

Recall that any knot can be embedded into a handlebody of genus g, for suitable g, by Proposition 2.17. We try to find relations between the successive knot types found in processive site-specific recombination reactions and the genus of the handlebodies in which these knots can be embedded. Remark that this definition of genus of the handlebody is the same as the original 'Seifert' genus of a knot. We will denote this genus by $\mathfrak{g}(k)$.
We would like to prove a proposition which states that is in some way the

[7] For addition of blocks, we can of course immediately say that on the left of the regular diagram blocks are added such that the first is between the middle two strings. If addition occurs on the right side, then the far right block has to be in the middle.

genus $\mathfrak{g}(k)$ of some knot is a lower bound of the number times distributive recombination has been performed. The original definition of the genus of a knot using handlebodies can unfortunately not be used to do this. Explaining why this is so, we will give a slightly different one, in a sense the dual definition of the genus $\mathfrak{g}(k)$.

To construct the graph of a knot one has to choose which area enclosed by the curve one first draws black (or white for that matter). After having chosen the first area, the rest of the colour scheme is fixed. If by convention you place the vertices of the graph in the white regions, you can make two graphs. One with a vertex in the outer region (meaning, the outer region is white), or no vertex in the outer region. We have stated at the point where we've defined the graph of a knot (cf. section 2.5), that by convention the outer region will be chosen white. This has consequences for our further discussion.
The genus of the knot, which measures the number of inner areas in which the region enclosed by the graph is divided, changes when changing from one graph to the other. The relations are as follows:
Consider a graph $\Gamma_1 = (V_1, E_1)$, where V_1 is the set of vertices and E_1 the set of edges of Γ_1. The graph encloses a region, which is divided into f_1 subregions. If we denote $|V_1|$ (the number of vertices of Γ_1) by v_1 and $|E_1|$ by e_1, then by the well-known Euler characteristic for 1-simplices, we have

$$v_1 - e_1 + f_1 = 2.$$

Now we can look at the dual graph of Γ_1, i.e. we put vertices in the regions of Γ_1 and connect them if there is an edge from Γ_1 between them. For this dual graph Γ_2, we have the same relation

$$v_2 - e_2 + f_2 = 2.$$

Furthermore we observe that the number of regions of Γ_1, f_1, equals the number of vertices of Γ_2, v_2, and vice versa.
For 2-bridge knots we can thus have the following two definitions of the genus $\mathfrak{g}(k)$. Either the genus belonging to the graph with vertex in the outer region, and the one belonging to the dual graph of this graph. We first examine the situation when we choose the outer region to be black, since that is the one suitable for our proposition.
We write $k = < a_1, \ldots, a_{2m+1} >$ as usual, and after having projected it onto the plane in the canonical way, we make the graph $\Gamma(k)$. By Lemma 4.6 $\Gamma(k)$ has the following form:

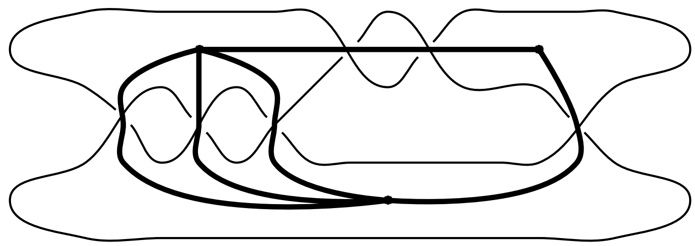


Figure 27: *The graph of* $< 3, 2, 1 >$

The genus $\mathfrak{g}(k)$ is equal to the number of areas inside the graph (or the total number of areas in which the plane is divided by $\Gamma(k)$ minus 1). By construction this is equal to

$$\mathfrak{g}(< a_1, \ldots, a_{2m+1} >) = \sum_{i=0}^{m} a_{2i+1} - 1. \tag{3}$$

The example in the figure above gives an illustration of this fact.
The genus belonging to the dual graph, is computed by

$$\mathfrak{g}(< a_1, \ldots, a_{2m+1} >) = \sum_{i=1}^{m} a_{2i} + 1.$$

We have the following proposition, stressing the fact that the genus used in the proposition is the one having *no* vertex in the outer region.

Proposition 4.5 *Let k be a knot, and $p_n^l(k)$ be the l times distributively recombinated knot with n twists per recombination, then the genus $\mathfrak{g}(k)$ of $p_n^l(k)$ is a lower bound for the number of times k has been recombinated, i.e. $l \geq \mathfrak{g}(k)$.*

Before we are able to prove the proposition, we introduce a lemma from (Murasugi 1996). As we have already remarked the lemma holds in both definitions of the genus.

Lemma 4.6 *Let k be a knot. Then the graph of k, $\Gamma(k)$, is the graph of a 2-bridge knot if and only if there exists a vertex $v \in V_{\Gamma(k)}$ such that $\Gamma - v$ and all edges incident to v consists of a simple line segment.* $\square$

The proof is straightforward and left to the reader. More on these subjects can be found in (Murasugi 1996).

PROOF OF PROPOSITION By Theorem 4.1 we can give a list of all possible product knots $p_n(k)$ of a given knot k:

(I) $< a_1, \dots , a_{2m+1}, n, 1 >$,
(II) $< a_1, \dots , a_{2m+1}, n-1, 1 >$,
(III) $< 1, n, a_1, \dots , a_{2m+1} >$,
(IV) $< 1, n-1, a_1, \dots , a_{2m+1} >$,
(V) $< a_1, \dots , a_{i-1}, j, \pm 1, a_i - j, \dots , a_{2m+1} >$,
(VI) $< a_1, \dots , a_i \pm 1, \dots , a_{2m+1} >$.

We use identity (3) to compute the genera of the various knots. We can see eg. from type (I), that

$$\mathfrak{g}(< a_1, \dots , a_{2m+1}, n, 1 >) = \mathfrak{g}(k) + 1,$$

since the n twists are not counted when the genus $\mathfrak{g}(p_n(k))$ is determined, since it's an even spot a_{2i}, and $\mathfrak{g}$ only sums the odd entries. Similarly we get:

$$\begin{aligned}
\mathfrak{g}(< a_1, \dots , a_{2m+1}, n-1, 1 >) &= \mathfrak{g}(k) + 1, \\
\mathfrak{g}(< 1, n, a_1, \dots , a_{2m+1} >) &= \mathfrak{g}(k) + 1, \\
\mathfrak{g}(< 1, n-1, a_1, \dots , a_{2m+1} >) &= \mathfrak{g}(k) + 1, \\
\mathfrak{g}(< a_1, \dots , a_{i-1}, j, \pm 1, a_i - j, \dots , a_{2m+1} >) &= \mathfrak{g}(k)(\pm 1)^*, \\
\mathfrak{g}(< a_1, \dots , a_i \pm 1, \dots , a_{2m+1} >) &= \mathfrak{g}(k) \pm a^\dagger.
\end{aligned}$$

Remarks:

* If i is odd then $\mathfrak{g}(p_n(k)) = \mathfrak{g}(k)$, and if i is even then $\mathfrak{g}(p_n(k)) = \mathfrak{g}(k) \pm 1$.

† When i is odd, a is equal ± 1, depending on the sign of the recombination. When i is even, a is equal to zero.

This proves the proposition, since with every recombination, the maximum contribution to the genus $\mathfrak{g}(k)$ is 1. □

Remark Since the enzymes which we have discussed (Tn3, Int) both require a non-trivial twist before recombination it's important to remark that for non-trivial twist action of enzymes the proposition gives equality between genus and number of distributive recombination events. This may be a practical way of obtaining extra information directly after having analysed the biological data with gel electrophoresis.

We are now in a good position to give a proof of Lemma 4.2. We simply introduce non-trivial twist inside the standard representation of a 2-bridge knot, and recombinate to find a product. After that we compute the genus and use Lemma 4.6 to decide whether the new graph belongs to a 2-bridge knot or not. Since this idea is conceptually easy, we give one worked out example and for the others we just give the graphs from which we can immediately decide if there exists a vertex such that the graph minus the vertex and all edges incident to it is a simple line segment.

PROOF OF LEMMA 4.2 Let k be a 2-bridge knot, denoted by

$$< a_1, \ldots, a_{2m+1} > .$$

We will perform non-trivial twist between the lower two strands, since this is the easiest case to describe. Let A_i be the region in which k has its a_i crossings. Suppose the twist occurs beneath A_i and A_{i+1}, where we choose i odd (i.e. A_i is a region of crossings between the middle strands). Note that the twists have to be vertical, which was also important in the proof idea given for Lemma 4.2. This means that the lower regions which did not contain any crossings, is now divided between a left and right region, with the vertical crossings in the middle. Since we have defined our graph in such a way that there is *no* vertex in the unbounded region outside the knot (regarding the knot to be projected onto a plane), we will now have two vertices between the lower strands: one for the left region and one for the right. Completing the graph in the usual fashion we find that the line segment is still there with its families of curves attached to it. Formerly these families all came together in one vertex, the one from the lower region. After vertical twists and recombination in this lower region the families are split into two subset. The first subset will have all its families of edges incident with the vertex from the left region, and the other with the right one. On top of that the left and right vertex are joined by a family of edges corresponding with the amount of twisting one has performed. For a picture of this situation, see Figure 28.

It's clear that this new graph does not contain a vertex with the desired property needed to satisfy Lemma 4.6, which proves that the knot with non-trivial twisting in the interior cannot be a 2-bridge knot.
In a completely similar way the other places where twisting and recombination might take place have to be considered. The resulting graphs are both of a different form than as stated in Lemma 4.6 and are illustrated in Figure 29. This completes the proof of Lemma 4.2. □

Lemma 4.6 might be extended to n-bridge knots.

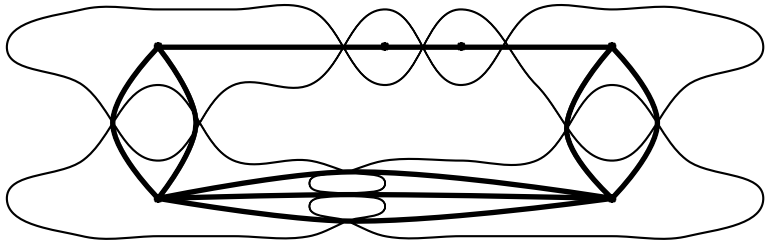


Figure 28: *A different proof of Lemma 4.2 using graphs*

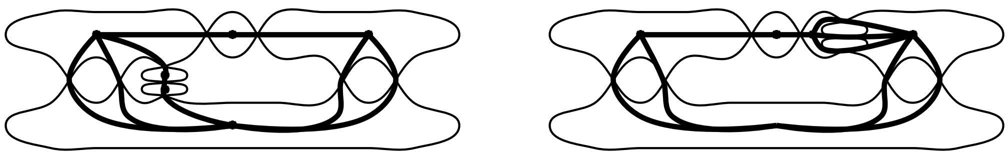


Figure 29: *The other two graphs of knots with vertical twists*

Proposition 4.7 *Let $\Gamma(k)$ be the graph of some knot k. k is an n-bridge knot if and only if there exists a $v \in V_{\Gamma(K)}$ such that $\Gamma(K) - v$ and all edges incident to it has a structure as described below.*
The structure can be defined as follows: we have n parallel lines $l_1, \ldots, l_n$, and on every line m points which fill a matrix

$$\begin{pmatrix} P_1^1 & P_2^1 & \ldots & P_n^1 \\ P_1^2 & P_2^2 & \ldots & P_n^2 \\ \vdots & \vdots & \ldots & \vdots \\ P_1^m & P_2^m & \ldots & P_n^m \end{pmatrix}.$$

Two adjacent point P_i^j, P_{i+1}^j are joined by a family of curves to form a structure as seen in Fig. 30, and the number of curves corresponds with the number of crossings equal to a_{i+1}^j.

Proof One direction of this statement is readily seen to be true: for a given n-bridge knot we can define the same kind of notation as we have for 2-bridge knots. More specifically we regard k as an $2n$-plat, and specify the crossings between adjacent parallel strands. As with 2-bridge knots, this can be done in such a way, that crossings in adjacent *pairs* of strings can be found not directly above or below each other (but slightly shifted horizontally). Then, if we number the n pairs of strands from top to bottom, we can write

$$k = (a_2^1, a_4^1, \ldots, a_{2m}^1 | a_1^2, \ldots, a_{2m+1}^2 | \ldots | a_2^n, \ldots, a_{2m}^n)$$

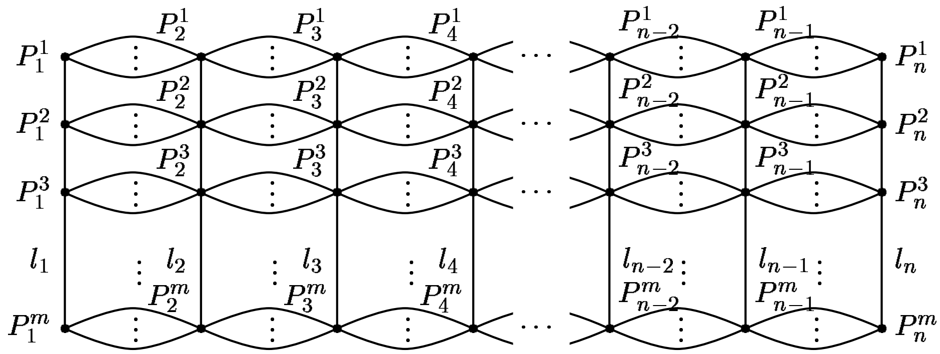


Figure 30: *The graph of an n-bridge knot*

If we make a graph of such a knot, we see that Γ has the structure as defined in the Proposition.

To prove the other direction we use the genus $\mathfrak{g}$ of a knot, and try to get back the knot from the handlebody of genus $\mathfrak{g}(k)$. We recall that by Proposition 2.17 a knot can always be embedded in a handlebody of genus $\mathfrak{g}(k)$. As we have already stated in the proof of Proposition 2.17 it is possible to give the knot which belongs to a particular graph or its tubular neighbourhood. For every vertex of the graph, we can put a pair of crossing arcs on the tubular neighbourhood precisely at the locus of the vertex, as seen in the figure in the proof of Proposition 2.17. It should be noted that non-equivalent graphs (w.r.t. the general equivalence relations from graph theory) may yield equivalent knots, but non-equivalent knots will yield non-equivalent graphs (Murasugi 1996). Therefore the knot which will be constructed from the graph with has a structure as defined in the proposition will be an n-bridge knot by construction. This completes the proof. $\square$.

When we consider n-bridge knots, with $n \geq 3$ we should pay attention to our definition of the graph of these knots. In the case of 2-bridge knots, after having projected the knot onto a plane and dividing this plane into the regions inclosed by the projected knot, we had the choice whether or not to include the outer unbounded region of the plane as a vertex of the graph or not. In either case Lemma 4.6 was seen to hold, and we could use the choice of graph suitable to prove that the genus of the knot w.r.t. this choice of graph is a lower bound for the number of times an enzyme had performed distributive recombination on the unknot. If the number of bridges of the knot is greater than two we can't generally project the knot in such a way that two adjacent strands in the $2n$-braid don't contain any crossings. This can be seen by adjusting the diagrams in the detailed proof of Theorem 2.30 in (Murasugi 1996) to the n-bridge case, with $n \geq 3$. Therefore there is only one graph of the two possible ones for which Proposition 4.7 holds, and this is the one most widely used in knot theory.

4.4 DNA recombination and knot polynomials

In section 2.6.2 we have introduced the Alexander and Jones polynomial of a knot or link. Here we try to use these invariants to study the enzyme action as discussed in previous sections.

The Alexander polynomial

Before we make an attempt to characterise site-specific recombination events using the Alexander polynomial we recall the fact that this polynomial is not well-defined for links, but only for true knots. We may thus forget the ambition to give a detailed account on the site-specific recombination mechanism as a whole, since there will obviously be many links involved. As a simple example, we consider a substrate knot $< 1, 2, 2 >$ and introduce one extra crossing in the left region, thus giving $< 2, 2, 2 >$ as product knot. We might easily compute the Alexander polynomial of the substrate knot, but since the product is a link (as are all $< 2, n, 2 >$ knots) the analysis stops. The number of links among 2-bridge knots is substantial, as seen in the following proposition, from (Murasugi 1996):

Proposition 4.8 *Let k be a 2-bridge knot of type (α, β). Then k is a link if and only if α is even.* □

None the less, good insight in the composition of the Alexander polynomial might give more grip on the biological processes.

As a first step towards a characterisation of site-specific recombination processes using the Alexander polynomial we would like to know the Alexander polynomial for a general 2-bridge (true) knot. To do this we use the first of the two definitions introduced in the first chapter to construct a presentation matrix for the knot by setting up the Seifert matrix of the knot. We have the following proposition.

Proposition 4.9 *Let $k = < a_1, \dots, a_{2m+1} >$ be a 2-bridge (true) knot. Then* *a Seifert matrix of k is of the form*

$$\begin{pmatrix}
1 & 1 & & & & & & & & & & & & & & \\
 & 1 & 1 & & & & & & & & & & & & & \\
 & & \ddots & & & & & & & & & & & & & \\
 & & & 1 & 1 & & & & & & & 0 & & & & \\
 & & & & -a_2 & 1 & & & & & & & & & & \\
 & & & & & 1 & 1 & & & & & & & & & \\
 & & & & & & 1 & 1 & & & & & & & & \\
 & & & & & & & \ddots & & & & & & & & \\
 & & & & & & & & 1 & 1 & & & & & & \\
 & & & & & & & & & -a_4 & 1 & & & & & \\
 & & & & & & & & & & \ddots & & & & & \\
 & & & & & & & & & & & -a_{2m} & 1 & & & \\
 & & & & 0 & & & & & & & & 1 & 1 & & \\
 & & & & & & & & & & & & & 1 & 1 & \\
 & & & & & & & & & & & & & & \ddots & \\
 & & & & & & & & & & & & & & & 1
\end{pmatrix}$$

where the $m+1$ blocks of consecutive 1's on the diagonal are of length $a_{2i+1}-1$ respectively, for $i = 1, \ldots, m$.

PROOF First we project the knot onto a plane in the usual way and create the 'standard' Seifert surface which is formed in the following way: for every crossing make cuts in the projection as in Figure 31. We then end up with a collection of disjoint discs in the plane. We now construct a surface by gluing back small twisted strips with a rotation corresponding with the original crossing. The result is a 'soap film' surface, and is clearly a Seifert surface for the 2-bridge knot k.

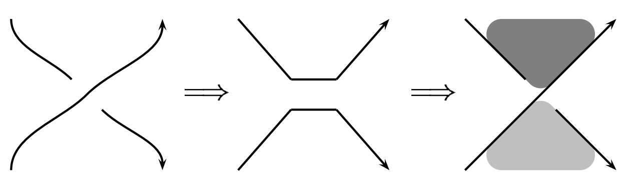


Figure 31: *The 'standard' Seifert surface of a knot*

As an example we consider the Seifert surface of the trefoil. The result can be found in Figure 32.

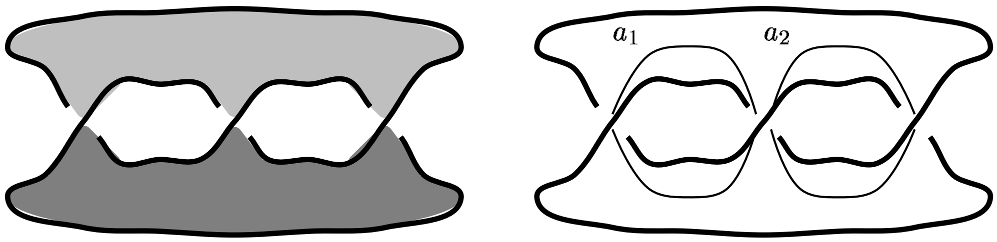


Figure 32: *The 'standard' Seifert surface M of the trefoil and the two generating closed curves of $H_1(\dot{M})$.*

For general 2-bridge knots we may proceed in an analogous fashion as in the example of the trefoil: we find one large disc at the bottom and the upper disc of the trefoil will in general be subdivided in smaller ones corresponding with the crossings in the regions $A_{2i},\ i = 1, \ldots, m$. We find a horizontally aligned array of closed curves which generate the first homology group of the Seifert surface. They will be denoted by

$$\{\alpha_1^1, \ldots, \alpha_{a_1-1}^1, \alpha^2, \alpha_1^3, \ldots, \alpha_{a_3-1}^3, \ldots, \alpha^{2m}, \alpha_1^{2m+1}, \ldots, \alpha_{a_{2m+1}-1}^{2m+1}\}.$$

Consequently we will only find a non-trivial linking number of two of these curves if they are next to each other on the surface. It's an easy exercise to verify the following relations:

$$\begin{aligned} lk(\alpha_j^i, (\alpha_j^i)^+) &= 1 \text{ if } i \text{ is odd} \\ lk(\alpha_j^i, (\alpha_l^k)^+) &= 1 \text{ if the two curves are adjacent} \\ lk(\alpha^{2i}, (\alpha^{2i})^+) &= -a_{2i} \end{aligned}$$

This completes the proof. □

With this general matrix one may easily compute the Alexander polynomial for given 2-bridge knots. To give general polynomials in the variables $a_2, a_4, \ldots, a_{2m}$ is a much harder task. Even for quite small knots with only a_2 non-zero, the determinant can become unintelligable. In the event that all odd entries $a_{2i+1},\ i = 0, \ldots, m$ are equal to 1, the Seifert matrix becomes very simple indeed: we only find the $-a_{2i},\ i = 1, \ldots, m$ on the diagonal and all entries zero. Since this matrix is obviously symmetric the Alexander matrix $V - tV^T$ is still diagonal with entries $-a_{2i}(1-t),\ i = 1, \ldots, m$, and therefore we find

$$\Delta(t) = (-1)^m (1-t)^m \prod_{i=1}^{m} a_{2i}$$

Recall that we have defined a second presentation matrix which may be calculated by a simple algorithm (cf. p.26). To calculate this matrix we have to choose a regular projection, which will of course be the standard projection for 2-bridge knots. Having oriented it we choose the overpasses for each of the n crossing points, where n equals the sum of the a_i, $i = 1, \dots, 2m+1$. The dimension of the matrix as a presentation matrix for $H_1(\tilde{X}) \oplus \Lambda$ is thus simply this sum. We now fill all non-zero entries as defined on page 26. We give, as a simple example, a presentation matrix for $< 2, 2, 1 >$.

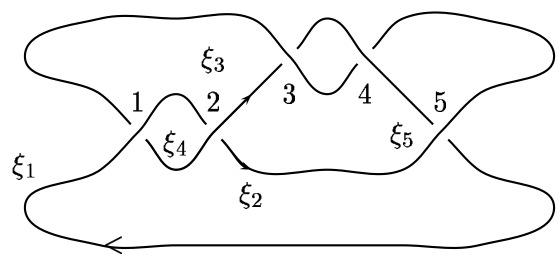


Figure 33: *Construction of a presentation matrix for* $k = < 2, 2, 1 >$

The generators of the knot group, x_i, correspond to the ξ_i in the knot projection. We see that the matrix for this projection is

$$\begin{pmatrix} 1-t & t & 0 & 0 & -1 \\ 0 & -1 & 0 & t & 1-t \\ t & 0 & 1-t & -1 & 0 \\ -1 & 1-t & -t & 0 & 0 \\ 0 & 0 & -1 & 1-t & t \end{pmatrix}.$$

By computing a 4×4 minor of this matrix and considering the fact that the coefficients of the polynomial have to be symmetric w.r.t. the powers of t we find that the Alexander polynomial of $< 2, 2, 1 >$ equals $-2t^{-1} + 3 - 2t$. This example shows the following matter, which will occur in the next section on Jones polynomials as well: the numbers of the crossings may be chosen at will since this will only produce a permutation of the columns, which give equivalent matrices according to our previous discussion in section 2.6.1. The labelling of the overpasses though may not be done at random, but has to be according to the orientation! (See Fig. 33 for an example.) Therefore, after having chosen one of the two orientations of a given knot and beginning at some crossing, we can *a priori* not determine which underpass lies left or right of the overpass at that particular crossing. This is due to the orientation and can only be seen after having physically drawn the knot on a piece of paper. A general form for the Alexander matrix for a 2-bridge knot

$< a_1, \ldots, a_{2m+1} >$ can not be given.
We will encounter this orientation problem in a different context in the section on Jones polynomials.

We have already noted that the dimensions of the two presentation matrices are generally different. In the case of the Seifert matrix, the dimension is equal to the number of closed loops which generate the first homology group of the open Seifert surface. This number is twice the genus of the knot (Rolfsen 1976). As we have seen in the proof of Proposition 4.9, the number of closed loops is equal to

$$\sum_{i=0}^{m} a_{2i+1}.$$

Working out the determinant of the matrix, we observe that all powers of the Alexander polynomial $\Delta(t)$ are non-negative, and the dimension is an upper bound on the the difference between the smallest and largest power of the polynomial, which is called the degree of $\Delta(t)$. A similar fact may be seen for the second presentation matrix.
Here the dimension of the matrix is equal to the number of crossings in the regular projection of the knot, which in the case of 2-bridge knots is equal to

$$\sum_{i=1}^{2m+1} a_i.$$

This number is of course greater than the dimension of the Seifert matrix the same knot. Therefore, if we would like to estimate the degree of $\Delta(t)$ for a given knot, the Seifert matrix gives a better estimate.
In general, knot complexity (measured by the minimal number of crossings for any regular projection) and top degree of $\Delta(t)$ admit a fine relation:

Proposition 4.10 *Let $deg(\Delta(k))$ be the difference between the highest and lowest degree of the (symmetric) Alexander polynomial for a knot k, and let $c(k)$ be the crossing number of k. Then we have*

$$c(k) \geq deg(\Delta(k)) + 1.$$

This means that to estimate knot complexity, using the second presentation matrix gives the optimal result by construction. On the other hand, estimating knot complexity will usually not be done using Alexander polynomials since the knowledge required to give this polynomial using the second construction requires gives you the total number of crossings anyway, before you have computed the Alexander matrix!

We now turn to more specific considerations: how does the Alexander polynomial change under site-specific recombination events? We have already noted that the Alexander polynomial is not well-defined for links. It's easy to see that with the second definition using under- and overpasses no structure due to the recombination will be picked up: introducing a single crossing changes the whole array of overpasses. The matrix may look completely different after only a small alteration of the knot. For the method using Seifert matrices things seem more promising.
Using the Classification Theorem 4.1 and Proposition 4.9 we may immediatly write down the different Seifert matrices for the original knot and its various possible products. The structure of the determinant however is too loose in general to recover the recombination event from the polynomials. Even in the simplest case, when we perform n twists and 1 recombination, giving a product $< 1, n, a_1, \ldots, a_{2m+1} >$, the Alexander polynomials of the substrate and product knots have little in common: neither may have any factors in common, nor may there be any any relation between the difference of the two.
In conclusion we can surely say that the Alexander polynomial, though easily computable, is not a good invariant to study site-specific recombination events, due to the fact that it's not well defined for links and the little structure the determinant bears w.r.t. its matrix.
We will now discuss a more recently developed polynomial, the Jones polynomial, and it's behaviour under recombination events.

The Jones polynomial

The motivation to study the Jones polynomial in relation to the biological processes is based on the following observation: in the procedure of calculating the Jones polynomial we simplify the knot by making overcrossings into undercrossings (w.r.t. some fixed projection), and by recombinating the ends as illustrated in Figure 14 in section 2.6.2. The local moves performed by enzymes such as topoisomerases on the one hand and Int or Tn3 on the other are examples of actions which are similar to the moves performed in the calculation of the Jones polynomial. On top of that the action of an enzyme as modelled in our previous discussion resembles these types of moves.

Recall that for the computation of the Jones polynomial we first choose a suitable crossing which is replaced to get knots k_- and k_0 needed for the skein relation (cf. p. 27). There are two possible ways of looking at this subject:

given a knot k we may study what effect k has on k_0 or k_-. On the other hand, we might be given a knot k_0 or k_- and try to find relations between these knots and a 'product knot' k_+. Both ways are useful for illustrating difficulties or creating insights in our model of recombination events. Although the latter way proves to be more profitable we start with the first.

The 'moves' which we have to perform to calculate the Jones polynomial can be seen as 'recombination moves'. Here we mean the following: the construction of k_0 can be viewed as an 'inverse recombination' in the sense that k_+ can be made from k_0 by a move similar to a site-specific recombination.[8] Similarly the move performed to construct k_- is to first do an 'inverse recombination' on k_+ to get k_0 and then perform a negative recombination to get k_-. In the following discussion we emphasise that when we use the word recombination it does not refer to an actual action performed by an enzyme but rather to the types of moves necessary to calculate the Jones polynomial.

In order for the recombination move to be similar to one made by an enzyme such as Tn3 we have to make sure that k_0 is still a 2-bridge knot. This is not always the case: e.g. if we take the knots $< 2,3,1 >$ and $< 2,2,1 >$ as examples, with some fixed orientations, then we see that an inverse recombination at the first crossings at the left of the knots produces a 2-bridge knot for $k =< 2,2,1 >$ but a non-2-bridge knot when $k =< 2,3,1 >$. Figure 34 illustrates the two outcomes. This is caused by the difference of the induced orientation of the knot at the site of the crossing where the inverse recombination is performed.
For 2-bridge links we have a similar problem. Having fixed an orientation of one of the two links, we may choose an orientation of the other. Only one of the two possible orientations (or two out of four if both orientations are chosen simultaneously) will yield 2-bridge knots for k_0. On the other hand it's always possible for 2-bridge links to choose a suitable pair of orientations such that k_0 and k_- are 2-bridge knots or links.

From this side of the problem we don't get much information other than to be careful with inverse recombinations of 2-bridge knots. Starting with a 'substrate knot' k_0 and producing k_- and k_+ from it turns out to be more fruitful.
We seem to be able to deal with the subtleties explained above by looking closely to the biological situation of enzyme mediated recombination: the two

[8] This move is not *equal* to a site-specific recombination move in the sense as we've discussed. This will be clarified when we discuss the other way of approaching the problem.

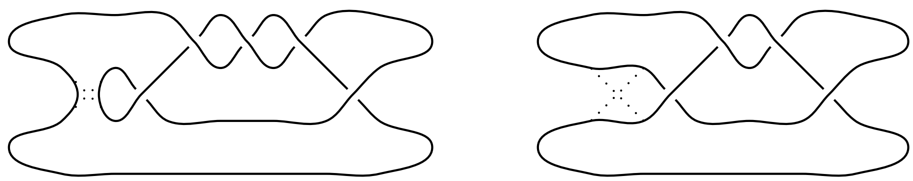


Figure 34: *The two possible inverse recombinations of a crossing*

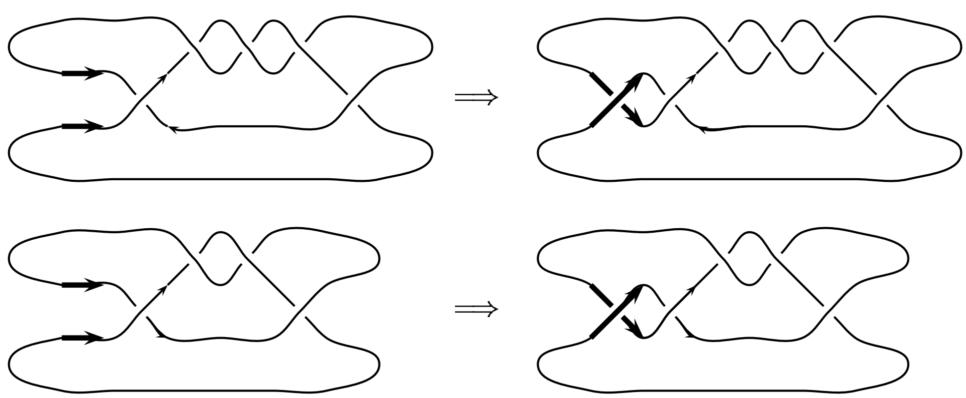


Figure 35: *Recombination of inverted vs. direct repeats*

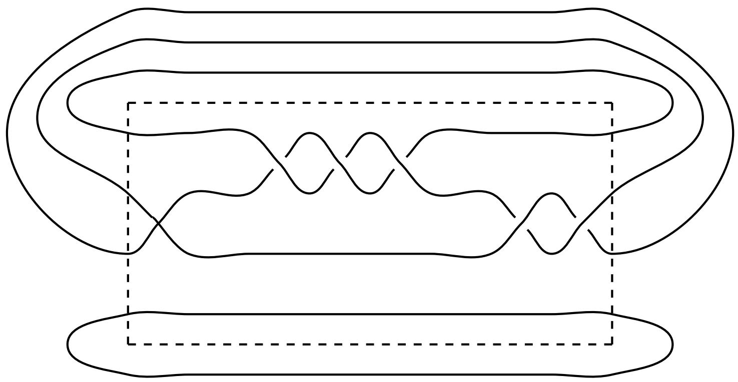


Figure 36: *The closed braid associated to 2-bridge knot* $< 1, 3, 2 >$

recombination sites involved in the site-specific recombination event have the same or opposite orientations (i.e. they are direct or inverted repeats respectively). But they consist of two homologous pieces of DNA, which have to aligned such that the 'orientations of the recombination sites', regardless of the induced orientation of the DNA knot, have to be equal. That simply means that they have to be aligned such that they have the same beginning and end. This ensures that the recombinated knot is still a 2-bridge knot. Or put differently, that the inverse recombination of a product knot is always a 2-bridge knot, if we take into account the role of the orientations of the recombination sites. Figure 35 illustrates this.
This approach is tempting since it's easy to write down skein relations for all possible recombination products using Theorem 4.1, but there is a flaw in this reasoning: the calculation of the Jones polynomial insists on choosing a fixed orientation with which positivity of the crossing points is determined. Therefore we may not regard the 'orientations' of the recombination sites to be the ones determining this. But the idea of forming 'product knots' from 'substrates' can be viewed alternatively, and we will be able to ensure that the crossing which is made by the recombination of the substrate is a positive crossing. To do this we have to consider our knots a bit differently.

Let $k = < a_1, \dots, a_{2m+1} >$ be a 2-bridge knot in canonical projection. Recall that we may regard k to be a 4-braid with the ends of the four strings tied together to form a 4-plat. But we may also make a new knot or link by gluing the ends differently, actually in the standard way in which braids are glued: the four points on the two opposite sides of the 4-braid are identified in pairs such that the sides collide. Figure 36 illustrates this. The braid made from k (enclosed in the dashed rectangle in Fig. 36) will be denoted by $b(k)$, and its closure $\bar{b}(k)$. The reader may also recall that the generators of the braid group are defined to be $\sigma_1, \sigma_2, \sigma_3$ (cf. Fig. 5, p.11). Since the fourth strand doesn't contain any crossings by the result in (Bankwitz & Schumann 1934), σ_1 and σ_2 will suffice. Then the braid $b(k)$ can be written as

$$\sigma_1^{a_1}(\sigma_2^{-1})^{a_2}\sigma_1^{a_3}\cdots\sigma_1^{a_{2m+1}}, a_i > 0.$$

It's easy to see that if we choose an orientation for $\bar{b}(k)$ we are always in the position that the crossings in the $\sigma_1^{a_1}$ region are positive. This enables us to study the effect of site-specific recombination events on 2-bridge knots using Jones polynomials. But our troubles have not been evaporated by this construction. In order to have maximal information about the recombination events using braids, it would be convenient if equivalence of knots implied equivalence of braids of these knots, or vice versa. Neither of these situations is the case however. We may construct the following counterexamples:
Let k be $< 1,2,3 >$ and $b(k)$ hence be $\sigma_1^1\sigma_2^{-2}\sigma_1^3$. Then we may perform a cyclic permutation on the $\sigma_{i_j}^j$ without changing the type of $\bar{b}(k)$. For instance let b' be $\sigma_2^{-2}\sigma_1^3\sigma_1^1 = \sigma_2^{-2}\sigma_1^4$. The knot corresponding to $\bar{b}'$ is $< 4 >$ (the crossings from the σ_2's are trivial, which is seen by drawing a picture). But it is evident that $< 1,2,3 >$ and $< 4 >$ are not equivalent knots, by the Classification Theorem of 2-bridge knots.
Similarly we may choose two equivalent knots, $k_1 =< 1,3,1 >$ and $k_2 =< -5 >$, say. Then $\bar{b}(k_1)$ is a 2-link, and $\bar{b}(k_2)$ a 3-link. So these closed braids cannot be equivalent.
From the classification point of view things don't seem to be much better either. From the time the concept of braids has been introduced in the 1920's by E. Artin, people have been trying to classify n-braids. A great deal of progress has been made during the years but the question remains unsettled to a great extend. Regardless of this fact we would like to know if classification of closed alternating 3-braids admits a solution, since that's all we need for our discussion. Murasugi has written a monograph on the topic of closed 3-braids in general, and solves the problem for various subclasses of the set of 3-braids, but dealing with the alternating closed 3-braids seems to be the hardest part (Murasugi 1974). It has been known for a long time that conjugate classes of the braid group $\mathfrak{B}_n$ give a great deal of insight in the link types of closed braids, and the important question which people have tried to answer is under what conditions do non-conjugate braids define different knots or links (i.e. closed braids regarded as knots or links). It is known that conjugate classes of braids classify all but the alternating closed 3-braids (Murasugi

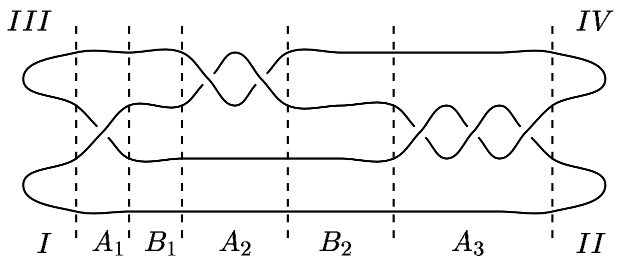


Figure 37: *The regions of $\mathcal{O}$, $\mathcal{A}$ and $\mathcal{B}$ of a 2-bridge knot.*

1974). Hartley gives some more recent results on this subject (Hartley 1980). Jones has tried to classify braids with the Jones polynomial in the paper in which he has introduced his invariant (Jones 1985). Since the classification of alternating closed 3-braids is still an unsolved problem we have to confine ourselves to giving skein relations and try a different approach.

Before we state these skein relations we introduce a bit of notation. For a 2-bridge knot $k = < a_1, \dots, a_{2m+1} >$ we distinguish three types of areas where recombination may take place: the regions on the 'outside' of the knot, which have been denoted by I, II, III and IV in Theorem 4.1. Here we set

$$\mathcal{O} = \{I, II, III, IV\}.$$

Recall that the region where a crossing of magnitude a_i occurs is denoted by A_i. Here we enlarge A_i by including the regions above and below the region with crossing a_i. Taken together they form a set denoted by $\mathcal{A}$. We are left with the areas of the knot between the A_i's, which we call B_i, and there are taken together to form a set $\mathcal{B}$. Figure 37 illustrates these regions. Recombination events on l at areas in $\mathcal{A}$ will now be denoted by

$$p_0^{\pm}(\mathcal{A}; l).$$

With these preliminaries out of the way we can state the skein relations more efficiently in the following proposition:

Proposition 4.11 *Let k be a 2-bridge knot on which we perform a site-specific recombination modelled by n twists and 1 recombination, denoted by $p_n(k)$. Then we have the following skein relations:*

$$V_{\bar{b}(p_n^+(\mathcal{O};k))}(t) = t^2 V_{\bar{b}(p_n^-(\mathcal{O};k))}(t) + t(\sqrt{t} - \frac{1}{\sqrt{t}})V_k(t),$$
$$V_{\bar{b}(p_0^+(\mathcal{A};k))}(t) = t^2 V_{\bar{b}(p_0^-(\mathcal{A};k))}(t) + t(\sqrt{t} - \frac{1}{\sqrt{t}})V_k(t),$$
$$V_{\bar{b}(p_0^+(\mathcal{B};k))}(t) = t^2 V_{\bar{b}(p_0^-(\mathcal{B};k))}(t) + t(\sqrt{t} - \frac{1}{\sqrt{t}})V_k(t),$$

PROOF The claim is proved by merely changing the recombinated crossing (which is always positive by previous arguments, since the recombination in question is positive) into its negative counterpart on the one hand, and in the inversely recombinated crossing on the other. By doing this we find the two required closed braids $\bar{b}(p_n^-(k))$ and $\bar{b}(k)$. This argument is possible in all three cases of recombination at $\mathcal{O}$, $\mathcal{A}$ and $\mathcal{B}$. □

Apart from writing down new skein relations we gain little information using braids since the braids $\bar{b}(p_0^+(k))$ and $\bar{b}(p_0^-(k)$ are not related when it comes to the knots they are made from.

It seems that simply actually computing the Jones polynomial from scratch, working your way up to more complicated 2-bridge knots seems the only alternative. If we limit ourselves to the case that the substrate unknots have recombination sites in direct repeats we can make sure that the crossing made by a recombination is always positive and can thus be used for reduction to knots which have to be used for the skein relation.

We can set up a general scheme to compute the Jones polynomial of any knot formed in distributive recombination processes with the method of recurrence relations from the field of combinatorics. We stress the fact that this is only valid for recombinations of direct repeats which is the case in Tn3-resolvase (Wasserman & Cozzarelli 1986). In this case we can always assume that a positive recombination yields a positive crossing.
To address the problem of calculating the Jones polynomial for any 2-bridge knot made by distributive recombination processes we choose a substrate knot $k = < a_1, \ldots, a_{2m+1} >$. Then k can be recombinated by actions on the regions $\mathcal{O}$, $\mathcal{A}$ and $\mathcal{B}$, and products will be conform Theorem 4.1. The setup of the scheme is to reduce the knot using skein relations such that whole regions A_{2i+1}, $0 \leq i \leq m$ are deleted in every step. This is possible since we may assume that the crossings in these regions are positive and hence may be replaced by there negative counterpart to form k_- on the one hand and by two

parallel lines to obtain k_0. We will see that in the reduction of every region A_{2i+1} we get a recurrence relation. We will now make these ideas more precise.

We start with an action on $\mathcal{A}$. We first set up new skein relations for the product knots. Set $V_{[q]}(t) = V_{<a_1-q,a_2,\dots,a_{i-1},j,\pm 1,a_i-j,a_{i+1},\dots,a_{2m+1}>}(t)$.

Proposition 4.12 *Let $k = < a_1, \dots, a_{2m+1} >$ be the substrate knot with direct repeats for the recombinated modelled by 0 twists and 1 recombination at region $\mathcal{A}$. Then the skein relations for the knots $p_0^+(\mathcal{A}; k)$ and $p_0^-(\mathcal{A}; k)$ are*

$$\begin{aligned} V_{p_0^+(\mathcal{A};k)}(t) &= t^2 V_{p_0^-(\mathcal{A};k)}(t) + t(\sqrt{t} - \frac{1}{\sqrt{t}}) V_k(t) \\ V_{p_0^-(\mathcal{A};k)}(t) &= t^2 V_{[2]}(t) + \\ &\qquad t(\sqrt{t} - \frac{1}{\sqrt{t}}) V_{[1]}(t). \end{aligned}$$

PROOF By Theorem 4.1 we know that the knots in question are of the form

$$p_0^\pm(\mathcal{A}; k) = < a_1, \dots, a_{i-1}, j, \pm 1, a_i - j, a_{i+1}, \dots, a_{2m+1} > .$$

Starting with the case the recombination is positive we choose the recombinated crossing as the one used for the skein relation. By setting $k_+ = p_0^+(\mathcal{A}; k)$ we get

$$\begin{aligned} k_- &= p_0^-(\mathcal{A}; k), \\ k_0 &= k. \end{aligned}$$

If we want to set up a new skein relation for $p_0^-(\mathcal{A}; k)$ we have to consider a different crossing than the one used for $p_0^+(\mathcal{A}; k)$ since this is not positive anymore. But there are plenty others left: we start systematically with the positive crossings in A_1, beginning with the outer left one. Then by setting $k_+ = p_0^-(\mathcal{A}; k)$ we get

$$\begin{aligned} k_- &= < a_1 - 2, a_2, \dots, a_{i-1}, j, \pm 1, a_i - j, a_{i+1}, \dots, a_{2m+1} >, \\ k_0 &= < a_1 - 1, a_2, \dots, a_{i-1}, j, \pm 1, a_i - j, a_{i+1}, \dots, a_{2m+1} > . \end{aligned}$$

This completes the proof. $\square$

These skein relations lets us give a proposition on the Jones polynomial of knots $p_0^+(\mathcal{A}; k)$ and $p_0^-(\mathcal{A}; k)$.

Proposition 4.13 *Let k be a 2-bridge knot which is recombinated by 0 twists and 1 recombination at $\mathcal{A}$. Then we have*

$$
\begin{aligned}
V_{p_0^+(\mathcal{A};k)}(t) &= t^2 V_{p_0^-(\mathcal{A};k)}(t) + t(\sqrt{t} - \frac{1}{\sqrt{t}})V_k(t) \\
V_{p_0^-(\mathcal{A};k)}(t) &= k_1\Big(\frac{t}{2}(\sqrt{t} - \frac{1}{\sqrt{t}}) + \sqrt{t(\sqrt{t} - \frac{1}{\sqrt{t}}) + 4t^2}\Big)^{a_1} + \\
& \quad k_2\Big(\frac{t}{2}(\sqrt{t} - \frac{1}{\sqrt{t}}) - \sqrt{t(\sqrt{t} - \frac{1}{\sqrt{t}}) + 4t^2}\Big)^{a_1}
\end{aligned}
$$

where k_1 and k_2 are uniquely determined by $V_{[0]}$ and $V_{[1]}$.

Proof With the definition of $V_{[q]}$ above we may set up a relation

$$V_{[q]}(t) = t^2 V_{[q-2]}(t) + t(\sqrt{t} - \frac{1}{\sqrt{t}})V_{[q-1]}(t), \quad 0 \le q \le a_1 - 1.$$

This is a second order recurrence relation with coefficients t^2 and $t(\sqrt{t} - \frac{1}{\sqrt{t}})$ which admits a concrete solution found in any textbook on combinatorics:

$$V_{[q]}(t) = k_1\alpha^q + k_2\beta^q$$

where α and β are solutions of the equation $x^2 - t(\sqrt{t} - \frac{1}{\sqrt{t}})x - t^2 = 0$. In this case these roots are equal to

$$
\begin{aligned}
\alpha &= \frac{t}{2}(\sqrt{t} - \frac{1}{\sqrt{t}}) + \sqrt{t(\sqrt{t} - \frac{1}{\sqrt{t}}) + 4t^2}, \\
\beta &= \frac{t}{2}(\sqrt{t} - \frac{1}{\sqrt{t}}) - \sqrt{t(\sqrt{t} - \frac{1}{\sqrt{t}}) + 4t^2}.
\end{aligned}
$$

The k_1 and k_2 are uniquely determined by the initial conditions, i.e. the first polynomials of the sequence, by solving the set of equations

$$
\begin{aligned}
V_{[0]} &= k_1 + k_2 \\
V_{[1]} &= k_1\alpha + k_2\beta
\end{aligned}
$$

This completes the proof. $\square$

We have now reduced the computation of $V_{[q]}$ to that of $V_{[0]}$ and $V_{[1]}$. These knots are of the form $< 1, n, b_1, \dots, b_k >$ and me be seen as reaction products of a substrate knot $< b_1, \dots, b_k >$ with action performed at $\mathcal{O}$.

We thus proceed discussing actions at $\mathcal{O}$. Analogously to the approach above we state the skein relations for the knots made by this action in the following proposition.

Proposition 4.14 *Let $k = < a_1, \dots, a_{2m+1} >$ be the substrate knot with direct repeats for the recombinated modelled by n twists and 1 recombination at region $\mathcal{O}$. Then the skein relations for $p_n^+(k)$ and $p_n^-(k)$ are:*

$$V_{p_n^+(k)}(t) = t^2 V_{p_n^-(k)}(t) + t(\sqrt{t} - \frac{1}{\sqrt{t}})V_k(t)$$
$$V_{p_n^-(k)}(t) = t^2 V_{p_{n-1}^-(k)}(t) + t(\sqrt{t} - \frac{1}{\sqrt{t}})V_k(t)$$

PROOF By Theorem 4.1 we know that the product knots are

$$< 1, n, a_1, \dots, a_{2m+1} >$$

or

$$< 1, n-1, a_1, \dots, a_{2m+1} >,$$

depending on the sign of the recombination.[9] Then if we take

$$k_+ = p_n^+(k) =< 1, n, a_1, \dots, a_{2m+1} >$$

we see immediately that $k_- = p_n^-(k)$ and $k_0 = k$. For the second skein relation we may draw a picture of $p_n^-(k)$ and see that it's equivalent to $< 1, n-1, a_1, \dots, a_{2m+1} >$. Now the proof is completed by regarding $p_{n-1}^+(k)$ as k_+ and repeating the first argument. □

The Jones polynomial of a knot which has been recombinated at $\mathcal{O}$ may now be computed by multiple reductions of the n crossings in A_2. As an example we compute the Jones polynomial of the first reaction product of the unknot. The result is found in the next Proposition.

[9]Of course we also should consider $< a_1, \dots, a_{2m+1}, n, 1 >$ and $< a_1, \dots, a_{2m+1}, n-1, 1 >$ but these are done completely analogous by regarding them as $< 1, n, a_{2m+1}, \dots, a_1 >$ and $< 1, n-1, a_{2m+1}, \dots, a_1 >$ respectively.

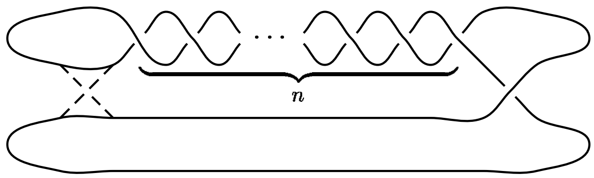


Figure 38: k_0 *of* $< 1, n, 1 >$ *is equivalent to the unknot.*

Proposition 4.15 *Let k be the unknot with its recombination sites as direct repeats. Then the Jones polynomial of the n-twisted positively recombinated unknot is equal to*

$$V_{p_n^+(k)}(t) = t^{2n} + t(\sqrt{t} - \frac{1}{\sqrt{t}}) \sum_{i=0}^{n-1} t^{2i}. \tag{4}$$

Proof By Theorem 4.1 we know that the product knot described in the proposition is equal to $< 1, n, 1 >$. The only two positive crossings we see are in A_1 and A_3 and we take the first one. Then by setting $k_+ =< 1, n, 1 >$ we see that $k_- =< -1, n, 1 >$ which is equivalent to $< 1, n-1, 1 >$. Furthermore we observe that k_0 which is drawn in Fig. 38 is equivalent to the unknot.

By doing multiple reductions using skein relations such as sketched above we obtain a list of knots $< 1, n, 1 >$, $< 1, n-1, 1 >$, $\ldots$, $< 1, 1, 1 >$, $< 1 >$. Now substituting all skein relations picked up during the reduction and recalling that $V_{<1>}(t) = 1$ yields the desired result. $\square$

This example illustrates how $p_n(k)$ may be reduced in n steps to

$$< -1, 1, a_1, \ldots, a_{2m+1} >$$

which is equivalent to

$$< -a_2, 1, a_3 - 1, a_4, \ldots, a_{2m+1} >$$

and we may write

$$V_{<1,n,a_1,\ldots,a_{2m+1}>}(t) = t^{2n} V_{<-a_2,1,a_3-1,a_4,\ldots,a_{2m+1}>}(t) + t(\sqrt{t} - \frac{1}{\sqrt{t}}) V_k(t) \sum_{i=0}^{n-1} t^{2i}.$$

Lastly we consider actions at $\mathcal{B}$. By Theorem 4.1 we know that we do not introduce new regions with positive or negative crossings, but simply enlarge or diminish the existing ones. Therefore we do not get in further difficulties if we want to calculate the Jones polynomial of a knot which has been formed by an action at $\mathcal{B}$.

By performing two reductions, one at $\mathcal{A}$ and one at $\mathcal{O}$, using skein relations, we have done one step in the scheme. All next steps consist of choosing the positive crossings at the far left in some A_{2i+1}, and reducing these and the crossings in A_{2i+2} by the above two steps. By systematically reducing all positive crossings in this way we end up with relatively small knots.

Having reduced all positive crossing in this systematic way we have to make certain that these can always be computed explicitely, thereby solving the huge implicit equation which has been built during all reduction stages of the original large recombination knot. The methods to deal with the first stages have been discussed before and are completely analogous to the proof of Propositions 4.15 or 4.13. We state some results. Any other needed to compute the Jones polynomial for some large knot may be left to the reader. In the following p, q, m, n are positive integers, α and β are the roots of $x^2 - p(t)x - t^2$ where $p(t) = t(\sqrt{t} - \frac{1}{\sqrt{t}})$.

The set of second order recombination knots contains knots of type $<1, p, 1, q, 1>$. Their polynomials are

$$V_{<1,p,1,q,1>} = t^{2(q-1)} + V_{<-p-q-2>} + p(t)V_{<-p>}\sum_{i=0}^{q-1} t^{2i}. \tag{5}$$

Here we make the following remark: The knots $<-p-q-2>$ and $<-p>$ have implicitly been discussed before in Proposition 4.15. More precisely, $<1, n, 1>$ can be seen to be equivalent to $<-n-2>$ by drawing a picture. One may also try to prove this by doing a small exercise in continued fractions.

Among these second order knots we also find $<m, n, 1>$ knots whose Jones polynomials can be seen to be

$$V_{<m,n,1>} = m_1\alpha^m + m_2\beta^m,$$

where m_1 and m_2 are determined by

$$\begin{aligned} m_1 + m_2 &= 1 \\ m_1\alpha + m_2\beta &= V_{<-n-2>}. \end{aligned}$$

In third order processive recombination processes we may encounter $<1, n, m, n, 1>$ knots which have Jones polynomial

$$V_{<1,n,m,n,1>} = n_1\alpha^m + n_2\beta^m,$$

where n_1 and n_2 are determined by

$$\begin{aligned} n_1 + n_2 &= V_{<-2n-2>} \\ n_1\alpha + n_2\beta &= V_{<1,n,1,n,1>}. \end{aligned}$$

Many other polynomials may be computed using the techniques described above. This concludes the exposition on computation of Jones polynomials of recombinated knots.

With this scheme we may try to ask the question what the Jones polynomial might tell us when it comes to enzyme mediated recombination mechanisms. However it soon becomes apparent that little is gained using this invariant: when we perform multiple steps of reduction we find some recurrence relation at every stage or an expression such as (5) on p. 76. Let's assume we only find the first type of relation between original and reduced knots. It's fairly easy to see that we get enormous polynomials in relatively few reduction steps. To illustrate this explosion of complexity let's consider the following simple example. Let V_1 be the Jones polynomial of some knot k. Then we might find

$$V_1 = l_1\alpha^n + l_2\beta^n,$$

where l_1 and l_2 are uniquely determined by the initial conditions, i.e.

$$\begin{aligned} l_1 + l_2 &= V_2 \\ l_1\alpha + l_2\beta &= V_3 \end{aligned}$$

for suitable polynomials V_2, V_3. We may compute l_1 and l_2 explicitly and substitute these in equation (6). We may go a step further and compute V_2 and V_3, say

$$\begin{aligned} V_2 &= k_1\alpha^m + k_2\beta^m, \\ V_3 &= n_1\alpha^p + n_2\beta^p. \end{aligned}$$

where the coefficients are determined by

$$\begin{array}{rclcrcl} k_1 + k_2 &=& V_4 & \text{and} & n_1 + n_2 &=& V_6 \\ k_1\alpha + k_2\beta &=& V_5 & & n_1\alpha + n_2\beta &=& V_7 \end{array}$$

for suitable polynomials $V_4, \ldots, V_7$.

If we write down V_1 in terms of V_2 and V_3 we find

$$\alpha^n(k_1\alpha^m + k_2\beta^m - \frac{n_1\alpha^p + n_2b^p - \alpha(k_1\alpha^m + k_2\beta^m)}{\beta - \alpha}) + \frac{\beta^n(n_1\alpha^p + n_2\beta^p - \alpha(k_1\alpha^m + k_2\beta^m))}{\beta - \alpha}.$$

A third step with suitable definitions for $V_4, \dots, V_7$ will yield

$$\alpha^n \Bigg((p_1\alpha^c + p_2\beta^c - \frac{q_1\alpha^d + q_2\beta^d - \alpha(p_1\alpha^c + p_2\beta^c)}{\beta - \alpha})\alpha^m + k_2\beta^m$$
$$- \frac{\alpha^p}{\beta - \alpha}(h_1\alpha^f + h_2\beta^f - \frac{j1\alpha^g + j2\beta^g - \alpha(h_1\alpha^f + h_2\beta^f)}{\beta - \alpha}) + n_2\beta^p -$$
$$\alpha(\alpha^{\mathrm{m}}(p_1\alpha^c + p_2\beta^c - \frac{q_1\alpha^d + q_2\beta^d - \alpha(p_1\alpha^c + p_2\beta^c)}{\beta - \alpha}) + k_2\beta^m) \Bigg) +$$
$$\alpha^p(h_1\alpha^f + h_2b^f - \frac{j_1\alpha^g + j_2\beta^g - \alpha(h_1\alpha^f + h_2\beta^f)}{\beta - \alpha}) + n_2\beta^p -$$
$$\frac{\alpha\beta^n}{\beta - \alpha}\left((\alpha^m k_2\beta^m(p_1\alpha^c + p_2\beta^c - \frac{q_1\alpha^d + q_2\beta^d - \alpha(p_1\alpha^c + p_2\beta^c)}{\beta - \alpha})) \right).$$

It's clear from these formulae that in general it will be very hard to analyse a Jones polynomial from a given recombination knot.

In conclusion we have seen that in general it's quite straightforward to calculate the Jones polynomial of a 2-bridge knot. Nevertheless if we try to analyze the polynomials of substrate and product knots we find ourselves entangled in a large set of recurrence relations in which all information of the recombination event is lost. Ultimately this is due to that fact that for the computation of the Jones polynomial we have to consider both skein relations rather than only the first one. In this first relation we find the structure which has lead to the investigation of Jones polynomials: a relation between the polynomial of the substrate versus the product knot. But in the second relation this structure is lost. We therefore conclude that, although several authors have suggested a coorespondence between DNA knots and the Jones polynomial, this invariant is not suitable to study the site-specific recombination mechanism as proposed in this model.

5 Further applications of topology in biochemistry

In this section we review some of the other fields in biochemistry in which topology is used to shed light on matters otherwise non-accessible. We give short but up-to-date accounts of the fields of topology of polymers, geometric and topological quantisation of DNA-sequences and the evolution of DNA from a topological point of view, and the use of topology in relation to NMR and x-ray diffraction methods.

5.1 Geometric and topological quantisation

In living cells DNA is used in various ways to sustain life. Processes such as transcription, replication and recombination all contribute to achieve this goal. Since the configuration of the DNA plays the crucial role in a way DNA functions, it's of utmost importance to get a good grip on the stereostructure of DNA in general, but also how these structures changes in the vital processes of DNA manipulation. We have already seen the way in which we try to keep track of the topology of DNA during site-specific recombination, which should give the reader a bit of a flavour for things to come.

By far the most interesting geometric and topological changes occur with circular DNA. This subject is of course completely in the line of thought of previous sections. We will first model such circular DNA-sequences with methods from differential geometry and then come up with a very nice classical theorem which gives us important insight in the global behaviour of DNA circles under local operations. A more thorough account of these matters can be found in (White 1992) and the original ideas have been published in (White, 1969).
We envision the DNA as being a supercoiled and circular strand. The supercoiling can be described in three quantities: **linking**, **twisting** and **writhing**. We give description of them and then investigate how they are related.

Consider a 2-link formed by k and l and regard a regular diagram of them in some plane P. Then we already know that the crossing number of the link is defined as the total number of (oriented) crossings. We now make the following definition:

Definition The **linking number** of a 2-link k, l is half the crossing number of a regular diagram of k, l, and is denoted by $Lk(k, l)$.

Obviously we have made a choice in taking but one regular diagram of the 2-link. It can be shown that the linking number is independent of this choice and doesn't change under continuous deformations of the links. For our applications in DNA, we regard the 2-link as the pair of backbones which are supercoiled around each other. The supercoiling is then measured by the linking number of the backbone strands. We can equivalently define this quantity to be the linking number of one of the strands with the axis of the DNA circle.
The other two quantities can be seen as complementary parts of the linking number. To be more precise, we can subdivide the crossings which add up to the linking number (modulo a factor $1/2$) in 'distant crossings' and 'local crossings'. Distant crossings are found when the axis of the DNA circle seems to cross itself when projected on a plane P. See the figure below for an example. The writhe of a 2-link measures this distant crossing, and can be defined as the number of times the axis of the two strands crosses itself (which is of course dependent on the projection), and denoted by $Wr(k)$. The local crossings are the ones made from the coiling of one of the backbone strands about the axis.

To get a better idea of the twist of a 2-link, we let $\mathbf{T}$ be the unit tangent vector field of k. Let furthermore $\mathbf{v}$ be a unit differentiable normal vector field along k. Then we have the following definition:

Definition The **twist of l about k** is defined by the line integral

$$Tw(k,l) = \frac{1}{2\pi}\int_k \mathrm{d}\mathbf{v}\cdot\mathbf{T}\times\mathbf{v}$$

Now we can state the remarkable fact: given to quantities which change under deformations of the strands, we can add them up to a third which doesn't:

Theorem 5.1 *For a 2-link of two supercoiled backbones forming a DNA circle have the following property:*

$$Lk(k,l) = Tw(k,l) + Wr(k,l).$$

□

This Theorem has direct applications for understanding certain facts about DNA-geometry. For instance, it gives an explanation for the supercoiling found in the genome: by twisting the backbones millions of times around each other a great deal of extra energy is being put in the DNA. To

counterbalance this, the DNA will writhe in the opposite direction to sustain the amount of linking. The result is a compacted form of DNA which can be neatly stored inside the nucleus. Another useful application of the Theorem is the fact that gel electrophoresis, which is used to analyse DNA-recombinant products, 'measures' the writhing of the knots, since the speed with which the DNA is being transported under electrical current through the gel is a measure for supercoiling, and therefore for writhing. By determination of writhing and twisting (which is known before the DNA substrate is being put into the solution with the enzyme), one can find information about the change of linking number by simply adding the two known numbers to get the third. It should be noted that the precise relation between gel speed and writhing are not clear mathematically. This is regarded as one of the important questions of contemporary topology in DNA research (Sumners 1995).
The Theorem was first proved by White in (White 1969), as an application of Gauß integrals.
For completeness we state another result from (White 1992), in which DNA-knots are being situated on **protein surface**. The best example of such a surface is the nucleosome core, a 'ball' of proteins called histones around which the DNA is wounded to form part of the complex structure of the compacted DNA chromosomes. The nucleosome is shaped as a cylinder. In this situation the DNA is wrapped around the surface nearly twice in a helical fashion. More generally the DNA is said to lie on a *solvent accessible surface*, which is the surface generated by moving a water-sized 3-ball around the atomic surface of the protein at the Van der Waals distance of the outer atoms, and is the continuous surface made from the loci of the centres of the 3-ball. In this case, it has proved to be less informative to consider writhing and twisting, but one can define two different quantities, called **winding number** and **surface linking number**. Giving real definitions would be beyond the scope of this paper, but if we denote them by $\Phi(k)$ and $Slk(k)$ respectively, one can prove a similar result as the Theorem stated above.

Theorem 5.2 *For a DNA-knot k which lies on a protein surface, we have*

$$Lk(k) = Slk(k) + \Phi(k).$$

□

5.2 Evolution of DNA from a topological point of view

This section gives an overview of the evolutionary side of DNA-research: how have structures found in DNA complexes evolved in history and can we explain some of the benefits or constraints that we can find in the geometrical or topological configuration of DNA? We summarise some of the recent developments, which can be found in (Cozzarelli 1992). We must warn the reader that the reasonings are teleological, and are not proved in any manner.

As we have seen the topology of DNA is rather complex. Two backbones are linked many times around each other, but the linking number of the strands is less then needed for a planar configuration of the DNA. This gives rise to the supercoiling which in interaction with proteins (mainly histones) compacts the DNA in such an efficient way that the huge molecules can be put into a relatively small space, c.q. the nucleus. Furthermore, we can find various links and knots in DNA complexes *in vivo* as well as *in vitro*. Although most DNA found in living cells is linear, the molecules are often subdivided into smaller loops of a few hundred kilobase pairs (kb), which can supercoil, catenate and knot.
As life exists nowadays, it is heavily dependent on this topological structure of DNA. An important example which indicates that small deviations from the usual topological configuration of DNA is lethal is illustrated by the topoisomerases which have been discussed in section 3. We recall that these enzymes are responsible for relieving DNA from additional stress which is a side-product of replication. By doing so, it maintains a constant topological configuration. Some cancer cells inhibit the effect of topoisomerases, and it has been shown that this causes rapid cell death.
The fact that DNA has such a complex structure is even more startling when we observe that it is the only molecule which is found with such a complex topology. To answer the questions why such an aberrant form is found only in DNA, we have to consider the first beginnings of DNA, when the language became fixed in which the information was being stored. The four base pairs can be found in all DNA (in RNA we Thymine is replaced by the homologous base Uracil). The fact that all organisms still use the same code to store genetic information reflects the fact that this choice of code really has been made at the origin if the occurrence of DNA. During the earliest times of the use of the one-dimensional language of base pairs, the length of DNA must have been limited and organisms fairly simple. As evolution progressed[10] more complex organisms evolved which required longer DNA.

[10]Of course we do not mean that evolution is a progressive process (which is still one of the greatest misconceptions of neo-Darwinism) in the sense, that evolution can be seen as a ladder, in which the most recently evolved species are the most advanced and best adapted for life on earth. Rather, we stress the fact that evolution is always tree-like, with inevitable gain of complexity during time, and look at evolution as a process which has progressed through time, until the present day.

With humans as examples of one of the species with very long DNA, some 10 million kb, the need for storage is evident. To cope with very long stretches of one-dimensional DNA, a double-stranded structure is well-suitable. The two backbones complement each other, and some errors which might take place on one strand can be fixed using the other. There are exceptions to prove the rule: some simple organisms still have single-stranded DNA, as do some simple viruses. As we have seen in the phage λ integration system, some viruses depend on a host to sustain their own survival.
The duplex structure not only solves the problem of securing the information stored in the huge molecules, it takes care of the storing of DNA as well: the helical winding which can also be found in many large molecules such as RNA and complex proteins causes supercoiling and compacts the molecule in an efficient way. To put it in an evolutionary manner, the linking number of the twists has been fixed at a quantity less then needed for ordinary relaxed DNA. This has resulted in a very efficient way of coiling the DNA.
This supercoiling is indirectly also responsible for the occurrence of knots and links. As we have seen in the Int recombinations in section 3.4, torus knots and links can be made from ordinary coiled substrate by just breaking and gluing the circle once. Topoisomerases also introduce knotting *in vivo*, and are also responsible for unknotting them.
The topology which has evolved to deal with the problems of storage, both overcoming errors introduced in the code during all kinds of manipulations and compacting, gives difficulties when it comes to other important functions such as replication. To illustrate one such problem, we consider replication of a circular piece of DNA. At the replication fork the DNA is locally unwinded and DNA is being duplicated. The unwinding gives additional linking, which is dealt with by the topoisomerases. So far so good. But when the replication fork nears its beginning point, the region where the topoisomerase has to lock onto the DNA becomes too small, and unlinking is impossible. To solve this situation, the DNA has to denaturate the last bit, which gives the complex of intertwined DNA-strands the possibility to migrate away from each other. Having achieved this, the ends which had yet to be duplicated can now be made at the separate DNA-strands. The result is a link of two double-stranded DNA circles. These still need to be broken and glued together to get two separate DNA circles. See the figure for an illustration of this process.
Another relevant and still unanswered question is how the cell deals with the fact that the nucleus is crowded with DNA in all kinds of formats, linear and circular, long and short, and the abundance of topoisomerases which act by random collision on the 'nearest DNA'. How does it sustain constant linking number, and how is total unknotting achieved? The answer to this question is not yet clear.

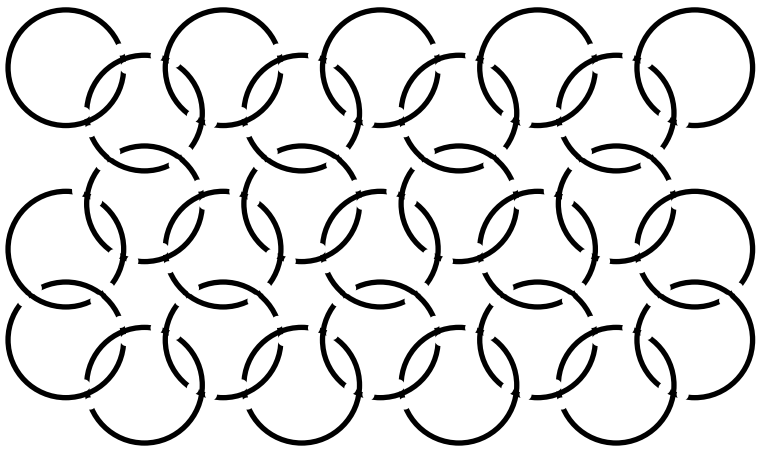


Figure 39: *A section of a kinetoplast DNA network*

As a last look at the evolution of DNA structures we discuss a strange family of clinically important unicellular parasites which have DNA structures which do not conform to all rules mentioned above about supercoiling and knotting. These so called **kinetoplasts** (kDNA in short) are networks of some 10^4 DNA rings, which are interlocked as in the figure below. Being a very stable catenated network it can sustain through all processes of life without using supercoiling or compacting as seen in regular cells. It is assumed (Cozzarelli 1992) that these structures are an early alternative in the evolution of DNA molecules. The way these rings stick together is a more efficient way of compacting large quantities of genetic code then is the conventional way of one long coiled string. Recently a number of people have been trying to solve questions about the topology of kDNA (Rauch *et al.* 1995; Ryan *et al.* 1988; somewhat earlier: Marini *et al.* 1980; Englund *et al* 1982). One of the topological questions which still stands is whether we can give a classification of 'periodic linked networks' (Sumners 1995), of which the kDNA network is one example. The periodicity of the network can be viewed as a mapping from the network to some compact 2-dimensional manifold such as a torus, and taking the universal cover of this manifold to get the entire network in the plane. Another question in which topology plays its part is to explain the observation that trefoils (the simplest knotted objects, with three crossing points, hence its name) are found in an intermediate stage of the replication process of kDNA (Sumners 1995, Ryan *et al.* 1988). How do the actions of enzyme responsible for replication use the structure of kDNA to produce these intermediate trefoils? Can we make generalisations of network topology for the replication process to yield these knots?

5.3 Limitations of biochemical knowledge

In this section we look at some future possible interactions between topology and DNA research. For instance one might look at the problem how the functionality of DNA is effected by knotting and linking. One might suggest, as has been proved already, that knotting may have serious constraints on the functions it might be involved in. The most obvious situation might be when DNA is never unknotted. Then knotting becomes so complex that no enzyme may be able to lock on to it, which will have lethal consequences: the major processes of replication, recombination and transcription are completely lost. A bit less trivial but not less important constraint on DNA functionality has been reviewed in the previous section on replication of DNA circles. Here the knotting which occurred already asked for some ingenuity of the topoisomerases to deliver two disjoint daughter DNA circles. The general question what kind of effects DNA knotting has on its functioning *in vivo* is still poorly understood. Research has rather been focused on revealing the *in vitro* conditions under which knotting occurs, and people have used this information to discover the mechanisms underlying enzyme activity as discussed in previous chapters (Cozzarelli, *pers. comm.*).
There are more situations in which topology might shed light on molecular issues.

Since DNA (and many other molecules found in living cells) are being manipulated by enzymes, the configuration of one may give information about the other. The reason this approach may work, is that molecules have to have special structure before they can be handled by some particular enzyme. These enzymes have *active sites* at which the relevant parts of the molecules are being brought close to each other, and after these global moves the enzyme does the action it's designed to perform. At a first glance one might try to use the topological and geometric information of the enzyme and in particular its active site inside the molecule to subtract information about the configuration of the molecule. Unfortunately the methods to study enzymes and/or DNA strings have not become sophisticated enough to do this. The best techniques which are available at present are *Nuclear Magnetic Resonance* (NMR), a spectroscopic method which measures the spin unique among only few isotopes (e.g. ^1H, ^{13}C, ^{31}P) by changing the orientation of the nuclear spin with microwave radiation and *x-ray diffraction*, which measures the deviation which scattered radiation will make when passing through a structure (Mathews & Van Holde 1996, for a more rigorous treatment: Van

Holde 1985). The first method creates images of enzymes in solutions. This means the enzyme and all its parts will move constantly, and it has not yet been proved to be possible to get good information about the real underlying structure of the enzyme. The great problem with NMR techniques is that it takes a relatively long amount of time to make a scan. The resulting image will thus give an average configuration of the components during reaction.[11]

Figure 40: *A detailed picture of a topoisomerase enzyme*

On the other hand we can use x-ray diffraction to get images of enzymes as ideal crystals: water is 'squashed' out of the enzyme, and during this process the various parts of the protein will move a bit to get into a stable configuration, which is therefore not the same as the original enzyme in solution. One can even attach some very tightly bonded substrate to the enzyme (which is usually a slightly altered form of the original substrate, since the bonds made by the original substrate tend to be too weak to withstand the transformation of the enzyme into its stable formation) which gives additional insight

[11] For the more biologically interested reader we give some additional references on NMR techniques on protein structure: Glore & Gronenborn 1989, Wagner *et al.* 1987.

how the groups found in the active site are bonded to a reactant. All relevant structures can be seen by nice computer graphics. See Figure 40 for the state of the art in enzyme graphics.

So although these methods can give some information, they are not yet detailed enough to give a good picture of what actually goes on during enzyme reactions. Things get even more complicated by the following observation: knowledge of the precise structure of the active site, where all the topological changes happen and which is therefore the most important region of the enzyme from the mathematical point of view, does not yield exact knowledge of the action of the enzyme. Contrary to the fact that the molecule has to be of a very specific form before it might be manipulated by the enzyme, the actual performance does in general not depend on the internal structure of the active site too much! By making some mutation of the enzyme on the outer periphery, at some place which does not seem to have any function at all, the actions performed by the enzyme might turn out quite different. At present, the best way of studying enzyme reactions in relation to there stereostructure is the following: by making random mutations in the DNA which codes for the enzyme, and by using large numbers of different substrates, one tries to unravel the mysteries behind enzyme function. So far progress has been slow and no real insight has been obtained. Only for the smallest enzymes made in the laboratory it has been proved possible to get some information on the relation between the various groups of one enzyme and their function. For all physiological enzymes this is still far from possible. It does not seem likely with these considerations that topology will shed light on the interaction between DNA and enzymes from the geometrical point of view for the years to come.

6 References

Abremski, K., B. Frommer, R.H. Hoess, Linking number changes in the DNA substrate during Cre-mediated loxP site-specific recombination, *J. Mol. Biol.* **192**, 17-26.

Alexander, J.W., Topological invariants of knots and links, *Trans. Am. Math. Soc.* **30**, 1928, 275-306.

Artin, E., Theorie der Zöpfe, *Abh. Math. Sem. Univ. Hamburg* **4**, 1925, 47-72.

Bankwitz, C. and H.G. Schumann, Über Viergeflechte, *Abh. Math. Sem. Univ. Hamburg* **10**, 1934, 263-284.

Burde, G. and H. Zieschang, *Knots*, De Gruyter studies in mathematics 5, de Gruyter, New York, 1985.

Clore, G.M. and A.M. Gronenborn, Determination of three-dimensional structures of proteins and nucleic acids in solution by nuclear magnetic resonance spectroscopy, *CRC Crit. Rev. Biochem.* **24**, 1989, 479-564.

Conway, J., On enumeration of knots and links and some of their related properties, *Computational problems in abstract algebra; Proc. Conf. Oxford 1967*, Pergamon Press, 1970, 329-385.

Cozzarelli, N.R., Evolution of DNA-topology: implications for its biological roles, *Proc. Symp. Appl. Math.* **45**, New scientific applications of geometry and topology, Sumners, D.W., ed., 1992, 1-16.

Cozzarelli, N.R., M.A. Krasnow, S.P. Gerrard, J.H. White, A topological treatment of recombination and topoisomerases, *Cold Spring Harbor Symp. Quant. Biol.* **49**, 1984, 383-400.

Crisona, N.J., R.L. Weinberg, B.J. Peter, D.W. Sumners, N.R. Cozzarelli, The topological mechanism of phage λ integrase, *J. Mol. Biol.* **289**, 1999, 747-775.

Englund, P.T., S.L. Hajduk, J.C. Marini, The molecular biology of trypanosomes, *Annu. Rev. Biochem.* **51**, 1982, 695-726.

Ernst, C., Tangle equations, *J. Knot Th. and its Ramifications* **5**, 1996, no.2, 145-159.

Ernst, C., Tangle equations II, *J. Knot Th. and its Ramifications* **6**, 1997, no.1, 1-11.

Ernst, C. and D.W. Sumners, the growth of the number of prime knots, *Math. Proc. Camb. Phil. Soc.* **102**, 1987, 303-315.

ERNST, C. AND D.W. SUMNERS, A calculus for rational tangles: applications to DNA recombination, *Math. Proc. Camb. Phil. Soc.* **108**, 1990, 489-515.

FOX, R.H., *A quick trip through knot theory, Topology of 3-manifolds*, Prentice-Hall, 1962.

GOLDMAN, J.K. AND L.H. KAUFFMAN, Rational tangles, *Adv. Appl. Math.* **18**, 1997, 300-332.

GOODRICK, R.E., Numerical invariants of knots, *Illinois J. Math.* **14**, 1970, 414-418.

GORDON, C.MCA. AND J. LUECKE, Knots are determined by their complements, *J. Amer. Math. Soc.* **2**, 1989, 371-415.

HARTLEY, R., On the classification of three-braid links, *Abh. Math. Sem. Univ. Hamburg* **50**, 1980, 108-117.

HEATH, D.J. AND T. KOBAYASHI, Essential tangle decomposition from thin position of a link, *Pac. J. Math.* **179**, 1997, no.1, 101-117.

JONES, V.F.R., A polynomial invariant for knots via von Neumann algebras, *Bull. Am. Math. Soc.* **12**, 1985, 103-111.

JONES, V.F.R., Hecke algebra representations of braid groups and link polynomials, *Ann. Math.* **126**, 1987, 335-388.

JONES, V.F.R., Milnor's work and knot polynomials, *Topological methods in modern mathematics* L.R. Goldberg & A.V. Phillips , eds., Publish or Perish, Inc., Houston, 1993, 195-202.

KANAAR, R., P. VAN DE PUTTE, N.R. COZZARELLI, Gin-mediated DNA inversion: product structure and the mechanism of strand exchange, *Proc. Nat. Ac. Sc. USA* **85**, 752-756.

KANAAR, R., A. KLIPPEL, E. SHEKSTMAN, J.M. DUNGAN, R. KAHMANN, N.R. COZZARELLI, Processive recombination by the phage μ gin system: implications for mechanisms of DNA exchange, DNA site alignment, and enhancer action, *Cell* **62**, 353-366.

KANENOBU, T. AND TOSHIO S., Polynomial invariants of 2-bridge links through 20 crossings, in *Advanced Studies in Pure Math.* **20**, *Aspects of low dimensional topology*, 1992, 125-145.

KRONHEIMER, P. AND T. MROWKA, Gauge theory for embedded surfaces I, *Topology* **32**, 1993, 773-826; II, *ibid.* **34**, 1995, 37-97.

LICKORISH, W.B.R., Prime knots and tangles, *Trans. Amer. Math. Soc.* **267**, 321-332.

LICKORISH, W.B.R., Polynomials for links, *Bull. London Math. Soc.* **20**, 1988, 558-588.

MARINI, J.C., K.G. MILLER, P.T. ENGLUND, Decatenation of kinetoplast DNA by topoisomerases, *J. Mol. Biol.* **255**, 1980, 4976-4979.

MATHEWS, C.K. AND K.E. VAN HOLDE, *Biochemistry,* 2^{nd} *ed.*, The Benjamins/Cummings Publishing Company, Inc. Menlo Park, CA, 1996.

MILNOR, J., On the total curvature of knots, *Annals of Math.* **52**, no.2, 1950, 248-257.

MURASUGI, K., On closed 3-braids, Memoirs of the A.M.S. no. 151, *American Mathematical Society, Providence, R.I.*, 1974.

MURASUGI, K., *Knot theory and its applications*, Birkhäuser, Boston, 1996.

RAUCH, C.A., P.T. ENGLUND, J. CHEN, N.R. COZZARELLI, J.H. WHITE, The topology of the kinetoplast DNA network, *Cell* **80:1**, 1995, 61-69.

REIDEMEISTER, K., Knotentheorie, *Ergebn. Math. Grenzgeb.* **1**, Springer-Verlag, Berlin, 1932.

ROLFSEN, D., *Knots and Links*, Math. Lecture Series **7**, Publish or Perish, Inc., Berkeley, 1976.

RYAN, K.A., T.A. SHAPIRO, C.A. RAUCH, J.D. GRIFFITH, P.T. ENGLUND, A knotted free minicircle in kinetoplast DNA, *Proc. Nat. Ac. Sc. USA* **85**, 1988, 5844-5848.

SEIFERT, H., Über das Geschlecht von Knoten, *Math. Ann.* **110**, 1934, 571-592.

SCHUBERT, H., Die eindeutige Zerlegbarkeit eines Knoten in Primknoten, *Akad. Wiss. Heidelberg, math.-nat. Kl.*, 1949, 3. Abh., 57-104.

SCHUBERT, H., Über eine numerische Knoteninvariante, *Math. Z.* **61**, 1954, 245-288.

SCHUBERT, H., Knoten mit zwei Brücken, *Math. Z.* **65**, 1956, 133-170.

SOTEROS, C.E., D.W. SUMNERS, S.G. WHITTINGTON, Entanglement complexity of graphs in $\mathbb{Z}^3$, *Math. Proc. Camb. Phil. Soc.* **111**, 1992, 75-91.

SPENGLER, S.J., A. STASIAK, N.R. COZZARELLI, The stereostructure of knots and catenanes produced by phage λ integrative recombination: implications for mechanism and

DNA structure, *Cell* **42**, 1985, 325-334.

SUMNERS, D.W., Knot theory and DNA, *Proc. Symp. Appl. Math.* **45**, New scientific applications of geometry and topology, Sumners, D.W., ed., 1992, 39-72.

SUMNERS, D.W., Lifting the curtain: using topology to probe the hidden action of enzymes, in *Calculating the secrets of life*, E.S. Lander and M.S. Waterman, eds., National Academy Press, Washington, D.C., 1995.

VAN HOLDE, K.E., *Physical Biochemistry,* 2^{nd} *ed.*, Prentice-Hall, Englewood Cliffs, N.J., 1985.

WAGNER, G, W. BRAUN, T.F. HAVEL, T. SCHAUMANN, G. NOBUIRO, K. WÜTHRICH, Protein strucures in solution by nuclear magnetic resonance and distance geometry. The polypeptide fold of the vine pancreatic trypsin inhibitor determined using two different algorithms, DISGEO and DISMAN, *J. Mol. Biol.* **196**, 1987, 611-639.

WALDHAUSEN, F., Heegaard-Zerlegungen der 3-Sphäre, *Topology* **7**, 1968, 195-203.

WASSERMAN, S.A., AND N.R. COZZARELLI, Determination of the stereostructure of the product of Tn3 resolvase by a general method, *Proc. Nat. Acad. Sci. U.S.A.* **82**, 1985, 1079-1083.

WASSERMAN, S.A., J.M. DUNGAN AND N.R. COZZARELLI, Discovery of a predicted DNA knot substantiates a model for site-specific recombination, *Science* **229**, 1985, 171-174.

WASSERMAN, S.A. AND N.R. COZZARELLI, Biochemical topology: Applications to DNA recombination and replication, *Science* **231**, 1986, 951-960.

WHITE, J.H., Geometry and topology of DNA and DNA-protein interactions, *Proc. Symp. Appl. Math.* **45**, New scientific applications of geometry and topology, Sumners, D.W., ed., 1992, 17-38.

WHITTINGTON, S.G., Topology of polymers, *Proc. Symp. Appl. Math.* **45**, New scientific applications of geometry and topology, Sumners, D.W., ed., 1992, 73-96.