

1-1-2007

A Model of DNA Knotting and Linking

Erica Flapan
Pomona College

Dorothy Buck
Imperial College London

Recommended Citation

Buck, Dorothy and Erica Flapan. A Model of DNA Knotting and Linking. *Knot Theory for Scientific Objects*, ed. Akio Kawauchi. Proceedings of the International Workshop on Knot Theory for Scientific Objects: OCAMI Studies Vol 1(1), 2007, 75-83.

This Conference Proceeding is brought to you for free and open access by the Pomona Faculty Scholarship at Scholarship @ Claremont. It has been accepted for inclusion in Pomona Faculty Publications and Research by an authorized administrator of Scholarship @ Claremont. For more information, please contact scholarship@cuc.claremont.edu.

A Model of DNA Knotting and Linking

Dorothy BUCK and Erica FLAPAN

ABSTRACT. We present a model of how DNA knots and links are formed as a result of a single recombination event, or multiple rounds of (processive) recombination events, starting with an unknotted, unlinked, or a $(2, m)$ -torus knot or link substrate. Given these substrates, according to our model all DNA products of a single recombination event or processive recombination fall into a single family of knots and links.

1. INTRODUCTION

Since their discovery in the late 1960s, DNA knots and links have been implicated in a number of cellular processes [6, 23, 28–31, 40, 41]. The variety of DNA knots and links observed has made biologically separating and distinguishing these molecules a critical issue. While DNA knots and links can be visualized via electron microscopy [15, 27, 33, 34, 45], this process can be both difficult and time-consuming. So topological methods of characterizing knotted and linked DNA can be helpful.

Topological techniques have already played a role in identifying DNA knots and links. Notable uses include the node number for knots [8], the Jones polynomial for catenanes [1], and the work of Schubert for 4-plats [43]. Furthermore, Ernst and Sumners have developed the tangle model of recombination [18] to describe the action of a particular class of proteins – the site-specific recombinases – in terms of tangle sums. Building on the experimental work of Wasserman and Cozzarelli [41, 42] and Conway’s theory of tangles [7], Ernst and Sumners used their tangle model to make predictions—later experimentally verified—about how the recombinase Tn3 resolvase interacts with DNA [18]. The tangle model has since been used to determine various features of protein-DNA interactions for a number of specific proteins (see for example [2–4, 9–14, 16–21, 24–26, 32, 35–39, 44]).

Rather than focusing on a specific recombinase as many earlier studies have done, we present a topological model that predicts which knots and links can be the products of site-specific recombination in general. We do this by describing the topology of how DNA knots and links are formed as a result of a single recombination event, or multiple rounds of (processive) recombination, starting with a substrate consisting of an unknot, an unlink, or $T(2, m)$ (i.e., a $(2, m)$ -torus knot or link). Our model relies on only three assumptions. We give biological evidence for each of these assumptions in the longer

Date: December 22, 2006.

This work was partially supported by an AWM Michler Collaborative Research Grant.

version of this paper [5]. In that paper, we also present a topological argument that all knotted or linked products brought about in this way, fall into a single family of knots and links which we characterize.

Our work complements earlier work [35] which used the tangle model [18] and several biologically reasonable assumptions to solve tangle equations, and subsequently determine which 4-plat knots and links arise as a result of site-specific recombination on the unknot. Our model goes further in that, in addition to the unknot, we allow substrates that are the unlink and $T(2, m)$. Furthermore, our assumptions concern only the biology of the recombination process – in particular, we do not assume the tangle model holds or that all products are 4-plats. Finally, we put no limits on the crossing numbers of the products we allow.

1.1. Background and Terminology. During site-specific recombination, two molecules of the site-specific recombinase bind to each of two specific DNA sites of approximately 30-50bp. We refer to these sites as the *crossover sites*. We use the term *recombinase complex* to describe the convex hull of the four bound recombinase molecules and the two crossover sites. If the recombinase complex meets the substrate precisely in the two crossover sites (as the examples in Figure 1 do) then we refer to the recombinase complex as a *productive synapse*.

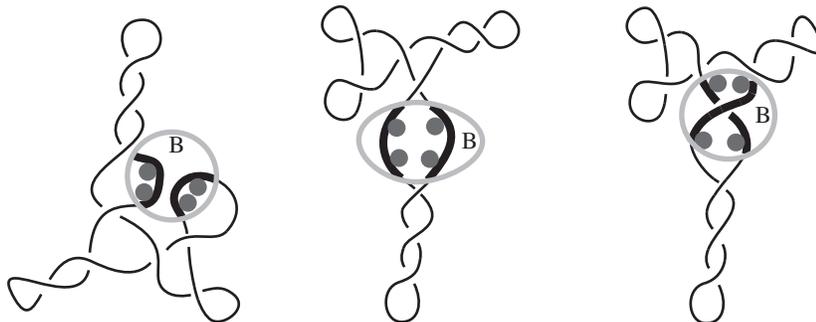


FIGURE 1. In these examples the recombinase complex B meets the substrate in the two crossover sites (highlighted in black).

Note that using our terminology, an enhancer sequence is neither a crossover site nor a part of a crossover site. If a recombinase has an enhancer sequence which is not sequestered from the crossover sites, then there is no productive synapse (for example see Figure 2). Transposases such as Tn5 and Tn10 have no productive synapse because their enhancer sequences are intertwined with the active transposition sites [22].

2. OUR ASSUMPTIONS

We start with a recombinase together with a substrate, which is an unknot, unlink, or $T(2, m)$ (see Figure 3). Let J denote the substrate(s) after synapsis.

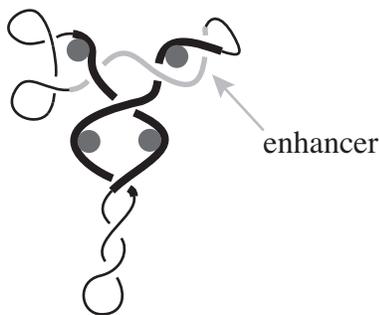


FIGURE 2. In this example a productive synapse does not exist.

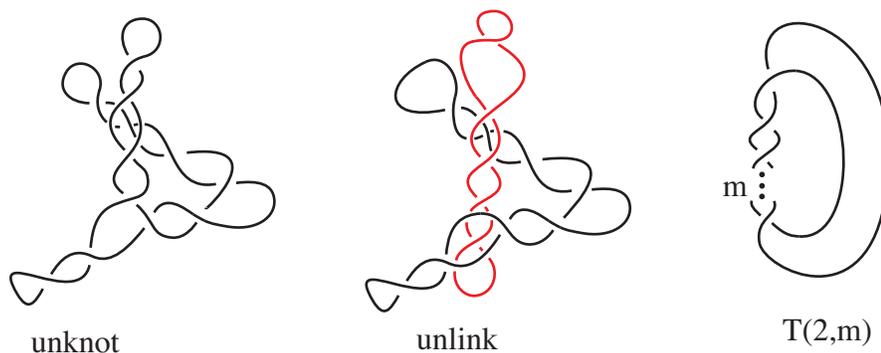


FIGURE 3. We consider these types of substrates.

We make three assumptions about the recombinase-DNA complex, which we state below both biologically and mathematically. Biological evidence for these assumptions is provided in [5].

(Biological) Assumption 1: The recombinase-DNA complex has a productive synapse, and there is a projection of the crossover sites which has at most one crossing between the sites and no crossings within a single site.

(Mathematical) Assumption 1: There is a ball B , containing the convex hull of the four recombinase molecules, which meets J in two arcs, and there is a projection of $B \cap J$ which has at most one crossing between the two arcs, and no crossings within a single arc.

A *spanning surface* for a substrate is a surface which is bounded by the substrate. If J is an unknot, unlink, or $J = T(2, m)$, then we require our spanning surface for J to be one disk, two disjoint disks, or a twisted annulus respectively. Figure 4 gives some examples of spanning surfaces.

(Biological) Assumption 2: There is a spanning surface for the substrate(s) whose interior is not pierced by the productive synapse. Also, no knots are trapped in the branches of the DNA on the outside of the productive synapse.

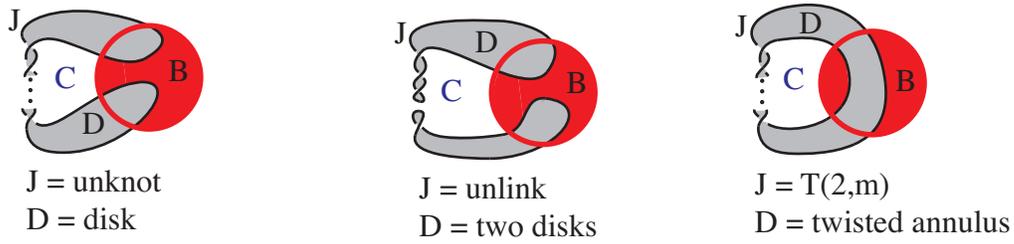


FIGURE 4. Examples of spanning surfaces D for a substrate J .

It follows from Assumption 2 that examples like those in Figure 5 cannot occur.

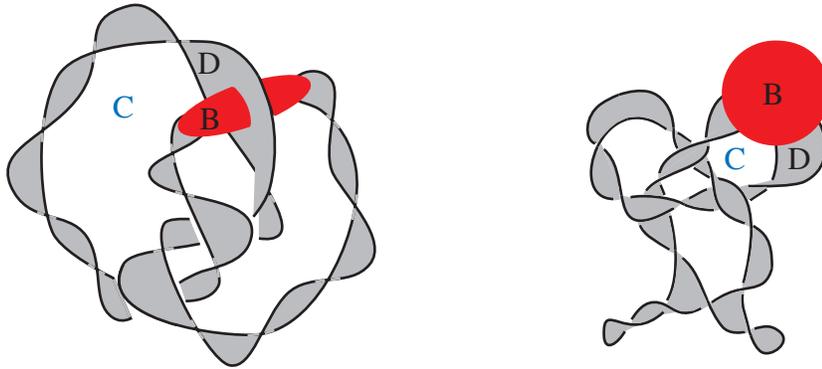


FIGURE 5. On the left every spanning surface for J is pierced by B , and on the right there is a knot in the branches outside of B .

In order to restate Assumption 2 mathematically, we introduce some terminology. Consider a planar surface together with a finite number of arcs whose endpoints are on the boundary of the surface. We can obtain a new surface by replacing each arc in the surface by a half-twisted band within a regular neighborhood of the arc in \mathbb{R}^3 . Figure 6 illustrates how such a surface can be obtained from an annulus together with a collection of arcs. Any surface that can be obtained from a planar surface in this way is said to be a *planar surface with twists*. The surfaces in Figure 4 are all planar surfaces with twists.

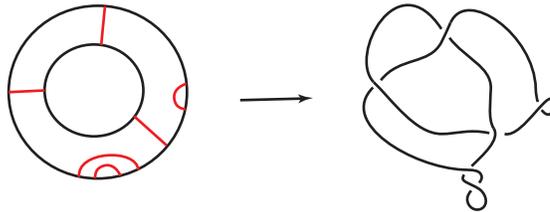


FIGURE 6. We obtain a *planar surface with twists* by replacing each arc by a half-twisted band within a regular neighborhood of the arc in \mathbb{R}^3 .

Let $C = \text{cl}(\mathbb{R}^3 - B)$, where cl denotes a region together with its boundary. Suppose that D is a spanning surface for J . We say that $D \cap C$ is *unknotted*, if there is an isotopy fixing B which takes $D \cap C$ to a planar surface with twists. For example, $D \cap C$ is unknotted for each of examples in Figure 4. By contrast, $D \cap C$ is knotted in the examples in Figure 5. We shall use the notation ∂ to denote the boundary of a region. Observe that in Figure 4, $D \cap \partial B$ consists of two arcs in each of the illustrations. This is not the case for the illustration on the left in Figure 5. Using the above language we restate Assumption 2 as follows.

(Mathematical) Assumption 2: There is a spanning surface D for J such that $D \cap \partial B$ is two arcs and $D \cap C$ is unknotted.

Site-specific recombinases fall into two families, the serine and tyrosine recombinases. Assumption 3 addresses the mechanism of these two recombinases separately.

(Biological) Assumption 3: Serine recombinase performs recombination via the “subunit exchange mechanism.” This mechanism involves making two simultaneous (double-stranded) breaks in the sites, rotating opposites sites together by 180° within the productive synapse and resealing opposite partners. In processive recombination, each recombination event is identical. After recombination mediated by a tyrosine recombinase, there is a projection of the crossover sites which has at most one crossing between the sites and no crossings within a single site.

(Mathematical) Assumption 3: Serine recombinase cuts each of the sites, adds a crossing within B between the cut arcs on different sites, then reconnects. In processive recombination, each recombination event is identical. After recombination with a tyrosine recombinase, there is a projection of the crossover sites which has at most one crossing between the sites and no crossings within a single site.

3. ANALYSIS OF THE POSSIBLE DNA KNOTTED AND LINKED PRODUCTS

Our analysis in [5] proceeds in the following way. We suppose that all three assumptions hold for a particular recombinase-DNA complex starting with one of the given substrates. We use Assumptions 1 and 2 to show that $C \cap J$ must have one of the forms illustrated in Figure 7. Then we determine the post-recombinant forms of $B \cap J$ for tyrosine by using Assumption 3, and for serine by using Assumption 3 together with the forms of $B \cap J$ given by Assumption 1. Finally, we glue each post-recombinant form of $B \cap J$ to each form of $C \cap J$ in Figure 7 to obtain all possible products. In this way, we show that all knotted and linked products of recombination are contained in the family of knots and links illustrated in Figure 8.

By letting p , q , r , and/or s be 0 or 1 in Figure 8 as appropriate, we obtain the five subfamilies illustrated in Figure 9. The knots and links in these subfamilies are possible products of recombination as specified in the theorem. We use the notation $C(r, s)$ for a knot or link consisting of one row of r crossings and a non-adjacent row

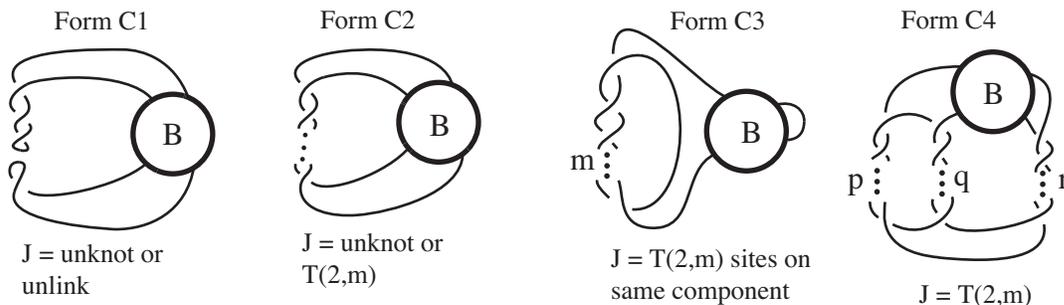
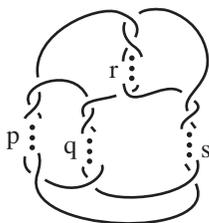
FIGURE 7. The possible forms for $C \cap J$.

FIGURE 8. We show that all knotted and linked products of recombination are in this family.

of s crossings, and the notation $K(p, q, r)$ for a pretzel knot or link with three rows containing p crossings, q crossings, and r crossings.

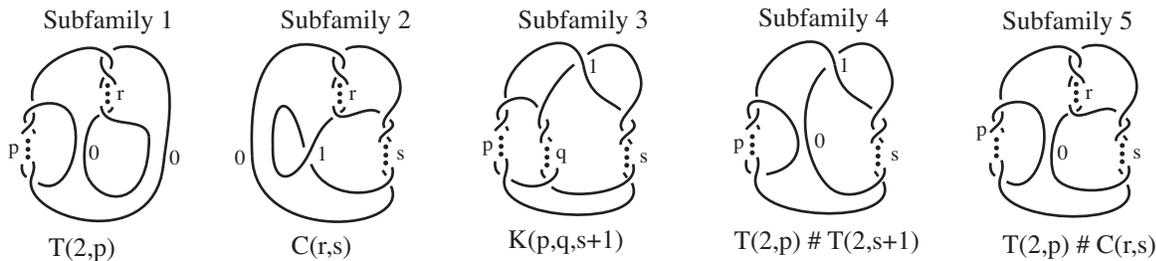


FIGURE 9. These subfamilies are contained in the family illustrated in Figure 8.

In particular in [5], we prove the Theorem below.

Theorem. *Suppose that Assumptions 1, 2, and 3 hold for a particular recombinase-DNA complex with substrate an unknot, unlink, or $T(2, m)$. Then all non-trivial products are contained in the family of knots and links illustrated in Figure 8. Furthermore if the substrate is an unknot or unlink then all non-trivial products are in Subfamilies 1 and 2 in Figure 9. In particular:*

Tyrosine recombinases: For an unknotted substrate, the only non-trivial products are $T(2, n)$ and $C(2, s)$. For an unlinked substrate, the only non-trivial product is a Hopf link.

Serine recombinases: For an unknotted substrate, the only non-trivial products are $T(2, n)$ and $C(r, s)$. For an unlinked substrate, the only non-trivial product is $T(2, n)$.

REFERENCES

1. T.C. Boles, J.H. White, N. Cozzarelli, N.R. *Structure of plectonemically supercoiled DNA*, J. Mol. Biol. **213** (1990) 931–951.
2. D. Buck, *The Topology of DNA-Protein Interactions*, PhD. Dissertation, University of Texas at Austin, 2001.
3. D. Buck and C. Verjovsky Marcotte, *Tangle solutions for a family of DNA-rearranging proteins*, Math. Proc. Camb. Phil. Soc. **139** no. 1, (2005) 59–80.
4. D. Buck and C. Verjovsky Marcotte, *Classification of Tangle Solutions for Integrases, A Protein Family that Changes DNA Topology*, J. Knot Theory Ramifications, to appear.
5. D. Buck and E. Flapan, *A Model of DNA Knotting and Linking in Site-Specific Recombination*, preprint.
6. D.A. Clayton, J. Vinograd, *Circular Dimer and Catenate Forms of Mitochondrial DNA in Human Leukaemic Leucocytes*, Nature **216** (1967) 652–657.
7. J. Conway, *On enumeration of knots and links and some of their related properties*, Computational Problems in Abstract Algebra. Pergamon Press (1970), 329–358.
8. N.R. Cozzarelli, 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.
9. N.J. Crisona, R.L. Weinberg, B.J. Peter, D.W. Sumners and N.R. Cozzarelli, *The topological mechanism of phage lambda integrase*, J. Mol. Biol. **289** no. 4 (1999), 747–775.
10. I.K. Darcy, *Biological distances on DNA knots and links: applications to Xer recombination*, Knots in Hellas '98, J. Knot Theory Ramifications, (2) **10** (2001), 269–294.
11. I.K. Darcy, *Solving unoriented tangle equations involving 4-plats* J. Knot Theory Ramifications, Vol. 14, No. 8 (2005) 993–1005.
12. I.K. Darcy *Applications of Topology to DNA Knot Theory*, Banach Center Publications **42** (1998)
13. I.K. Darcy, *Solving oriented tangle equations involving 4-plats*, J. Knot Theory Ramifications, Vol. 14, (2005) 1007–1027.
14. I.K. Darcy, J. Luecke and M. Vázquez, *Tangle Analysis of the Mu transpososome*, in preparation.
15. F. Dean, A. Stasiak, T. Koller, N. R. Cozzarelli, *Duplex DNA knots produced by Escherichia coli topoisomerase I. Structure and requirements for formation*, J. Biol. Chem. **260** (1985) 4975–5983.
16. C. Ernst, *Tangle equations*, J. Knot Theory Ramifications **5** (1996), 145–159.
17. C. Ernst, *Tangle equations II*, J. Knot Theory Ramifications **6** (1997), 1–11.
18. C. Ernst and D.W. Sumners, *A calculus for rational tangles: Applications to DNA recombination*, Math. Proc. Camb. Phil. Soc. **108** (1990), 489–515.
19. C. Ernst and D.W. Sumners, *Solving tangle equations arising in a DNA recombination model*, Math. Proc. Camb. Phil. Soc. **126** (1999), 23–36.
20. J.R. Goldman and L.H. Kauffman, *Rational Tangles*, Adv. Appl. Math. **18** (1997), 300–332.
21. I. Grainge, D. Buck and M. Jayaram, *Geometry of site alignment during Int family recombination: Antiparallel synapsis by the Flp recombinase*, J. Mol. Biol. **298** no. 5 (2000), 749–764.
22. Gueguen E, Rousseau P, Duval-Valentin G, Chandler M. The transpososome: control of transposition at the level of catalysis. *Trends Microbiol* **13** (2005) 543–9.
23. B. Hudson, J. Vinograd, *Catenated Circular DNA Molecules in HeLa Cell Mitochondria*, Nature **216** (1967) 647–652.

24. L.H. Kauffman and S. Lambropoulou, *On the classification of rational tangles*, *Advances in Applied Mathematics* **33** no. 2 (2004) 199–237.
25. L.H. Kauffman and S. Lambropoulou, *On the classification of rational knots*, *L'Enseignement Mathematiques* **49** (2003) 357–410.
26. L.H. Kauffman and S. Lambropoulou, *Classifying and applying rational knots and rational tangles*, *Contemporary Mathematics AMS Series*, 304 (2002), 223–259.
27. M. A. Krasnow, A. Stasiak, S. J. Spengler, F. Dean, T. Koller and N. R. Cozzarelli, *Determination of the absolute handedness of knots and catenanes of DNA*, *Nature* **304** (1983) 559–560.
28. L. F. Liu, J. L. Davis, *Novel topologically knotted DNA from bacteriophage P4 capsids: studies with DNA topoisomerases*, *Nucleic Acids Research* **9** 16 (1981) 3979–3989.
29. L. F. Liu, L. Perkocha, R. Calendar, J. C. Wang, *Knotted DNA for Bacteriophage Capsids*, *Proc. Natl Acad. Sci.* **78** no. 9 (1981) 5498–5502.
30. J. Menissier, G. de Murcia, G. Lebeurier, L. Hirth, *Electron microscopic studies of the different topological forms of the cauliflower mosaic virus DNA: knotted encapsidated DNA and nuclear minichromosome*, *EMBO J.* **2**, (1983) 1067–1071.
31. K. A. Ryan, T. A. Shapiro, C. A. Rauch, J. D. Griffith, P. T. Englund, *A knotted free minicircle in kinetoplast DNA*, *Proc. Natl Acad. Sci.* **85**, (1988) 5844–5848.
32. Y. Saka and M. Vazquez, *TangleSolve: topological analysis of site-specific recombination*, *Bioinformatics* **18** (2002) 1011–1012.
33. S. J. Spengler, 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.
34. A. Stasiak, E. DiCapua, T. Koller *Unwinding of duplex DNA in complexes with recA protein*, *Cold Spring Harb Symp Quant Biol.* (1983) 47 Part 2:811–20.
35. D. W. Sumners, C. Ernst, S. J. Spengler and N. R. Cozzarelli. *Analysis of the mechanism of DNA recombination using tangles*, *Quarterly Review of Biophysics* **28** no. 3 (1995) 253–313.
36. M. Vázquez and D.W. Sumners, *Tangle analysis of Gin recombination*, *Math. Proc. Camb. Phil. Soc.*, **136**, (2004), 565–582.
37. M. Vázquez, *Tangle Analysis of Site-specific Recombination: Gin and Xer systems*, PhD Dissertation, Mathematics Florida State University (2000).
38. M. Vazquez, S.D. Colloms and D.W. Sumners, *Tangle analysis of Xer recombination reveals only three solutions, all consistent with a single 3-dimensional topological pathway*, *J. Mol. Biol.* **346** no. 2, (2005) 493–504.
39. A. A. Vetcher, A. Y. Lushnikov, J. Navarra-Madsen, R. G. Scharein, Y. L. Lyubchenko, I. K. Darcy, S. D. Levene, *DNA topology and geometry in Flp and Cre recombination*, *J. Mol. Biol.* 2006 Apr 7;357(4):1089–104.
40. J. C. Wang, H. Schwartz, *Noncomplementarity in Base Sequences Between the Cohesive Ends in Coliphages 186 and λ and the Formation of Interlocked Rings between the Two DNAs*, *Biopolymers* **5** (1967), 953–966.
41. S.A. Wasserman and N. Cozzarelli, *Biochemical topology: Applications to DNA recombination and replication*, *Science* **232** (1986), 951–960.
42. S. A. Wasserman, J. M. Dungan, N. R. Cozzarelli, *Discovery of a predicted DNA knot substantiates a model for site-specific recombination* *Science* **229**, (1985), 171–174.
43. J. H. White, N. R. Cozzarelli, *A Simple Topological Method for Describing Stereoisomers of DNA Catenanes and Knots*, *Proc. Natl. Acad. Sci.* **81** (1984) 3322–3326.
44. J.H. White, K.C. Millett, N.R. Cozzarelli, *Description of the topological entanglement of DNA catenanes and knots by a powerful method involving strand passage and recombination*, *J. Mol. Biol.* **197** (1987) 585–603.

45. E. L. Zechiedrich, N. J. Crisona, *Coating DNA with RecA protein to distinguish DNA path by electron microscopy*, In Bjornsti, M. and Osheroff, N. (eds), *Methods in Molecular Biology: DNA Topoisomerase Protocols*. Humana Press, Totowa, NJ, **1** (1989) 99–108.

DEPARTMENT OF MATHEMATICS, IMPERIAL COLLEGE OF LONDON, UK

E-mail address: `d.buck@imperial.ac.uk`

DEPARTMENT OF MATHEMATICS, POMONA COLLEGE, CLAREMONT, CA 91711, USA

E-mail address: `eflapan@pomona.edu`