Graphs for Modeling DNA Recombination Processes

— Background from Biology —

August 28, 2011

1 Introduction

This document is posted at http://math.usf.edu/∼saito/DNAweb/BGbio.pdf. This is an overview of a series of work [2, 3, 4] on template models of DNA recombination. The purpose of this document is to provide a background material for research projects and their results presented in this web site http://math.usf.edu/∼saito/DNAweb.

2 Gene assembly in ciliates

Several species of ciliates, such as *Oxytricha* and *Stylonychia*, undergo massive genome rearrangement during sexual reproduction. These massively occurring recombination processes make them ideal model organisms to study gene rearrangements. See [47] and references therein for details of the descriptions below.

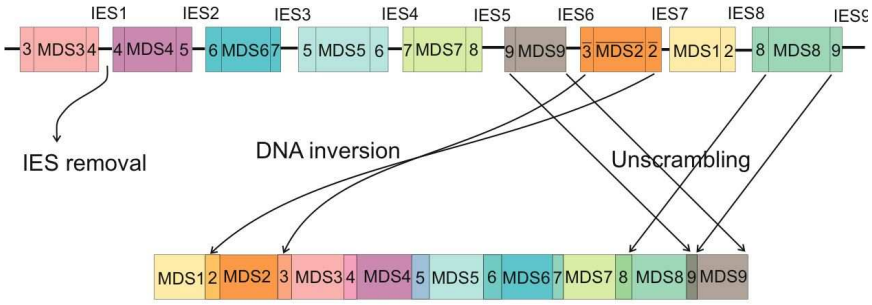

Figure 1: DNA recombinations in cilliates, Actin I of Oxytricha Nova from micronuclear gene to macronuclear gene, not to be scaled

There are two nuclei, a micronucleus and a macronucleus in these ciliates.

Micronuclear genes are reassembled to macronuclear genes during sexual reproduction. These DNA processing events involve global deletion of 95-98% of the germline DNA, effectively eliminating *all* so-called "junk" DNA, intervening DNA segments (internal eliminated sequences, IESs) that interrupt genes. Because IESs interrupt coding regions in the micronucleus, each macronuclear gene may appear as several nonconsecutive segments (macronuclear destined sequences, MDSs) in

the micronucleus. During macronuclear development, the IESs that interrupt MDSs in the micronucleus are all deleted. Moreover, the order of MDS segments for thousands of genes in the micronucleus can be permuted or sequences reversed. Formation of the macronuclear genes in these ciliates thus requires any combination of the following three events: unscrambling of segment order, DNA inversion, and IES removal. Fig. 1 shows an example of a typical scrambled gene requiring all three events.

3 Template models for RNA-guided DNA recombination

The general mechanism that guides this process of assembly, as recently proposed in [2], is guided by maternal RNA template sequence. It has been observed that there exist pointer-like sequences that are repeated at the end of each nth MDS and the beginning of each $(n+1)$st MDS in the micronucleus. Each pointer sequence is retained as exactly one copy in the macronuclear sequence [109, 127]. Such repetition of sequences suggests "pointer guided" homologous recombination. Several models for these processes have been proposed, including the models in [47, 96] which all assume that a correct pair of pointers align and splice.

Using the DNA recombinations in ciliates as a model system to describe DNA rearrangements that may occur more generally [110], Prescott et al. [128] and later [2] proposed an epigenetic model in which an RNA or DNA template derived from the maternal macronucleus guides assembly of the new macronuclear chromosomes. By our model, macronuclear templates could provide a scaffold to organize the layout of segment order and DNA deletion, using strand displacement and branch migration to align pointer pairs for recombination. Recently, our model was supported by several experimental observations that maternal RNA templates guide DNA rearrangement in the early development of the macronucleus [120].

For reviewing the template model, we include (edited) excerpts from [3] for the rest of the subsection. First we establish notations to present the model with a dsRNA template. Our assumption that templates are dsRNA molecules means that the portion of the molecule that plays the role of templates is double-stranded. The whole molecule itself may be part of a secondary structure of a ssRNA, such as a hairpin-like ssRNA. For easier representation, we depict the double-stranded molecules as ladders, ignoring the helical structure.

Let T be the dsRNA molecule that plays the role of a template, and let X and Y be two portions of a DNA molecule(s) that contain the same pointer.

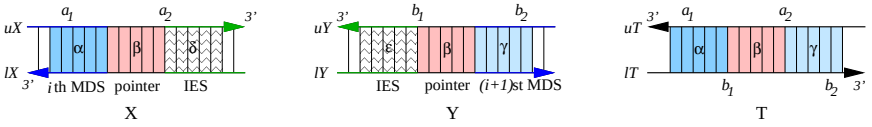

Figure 2: Two segments X, Y to be recombined, and a template T

The figures represent X, Y and T as ribbons oriented $5'$ to $3'$. Base pairs are represented as vertical stripes (Fig. 2). The "upper" strand of X (denoted uX) reading $5'$–$3'$ contains block $\alpha\beta\delta$, where α is a portion of the ith MDS, β is the $(i+1)$st pointer, and δ is a portion of an IES. The "upper" strand of Y (uY), read $3'$–$5'$, contains a block $\bar{\epsilon}\bar{\beta}\bar{\gamma}$, where γ is a portion of the ith MDS and

ϵ a portion of an IES (barred symbols represent Watson-Crick complements of unbarred symbols).

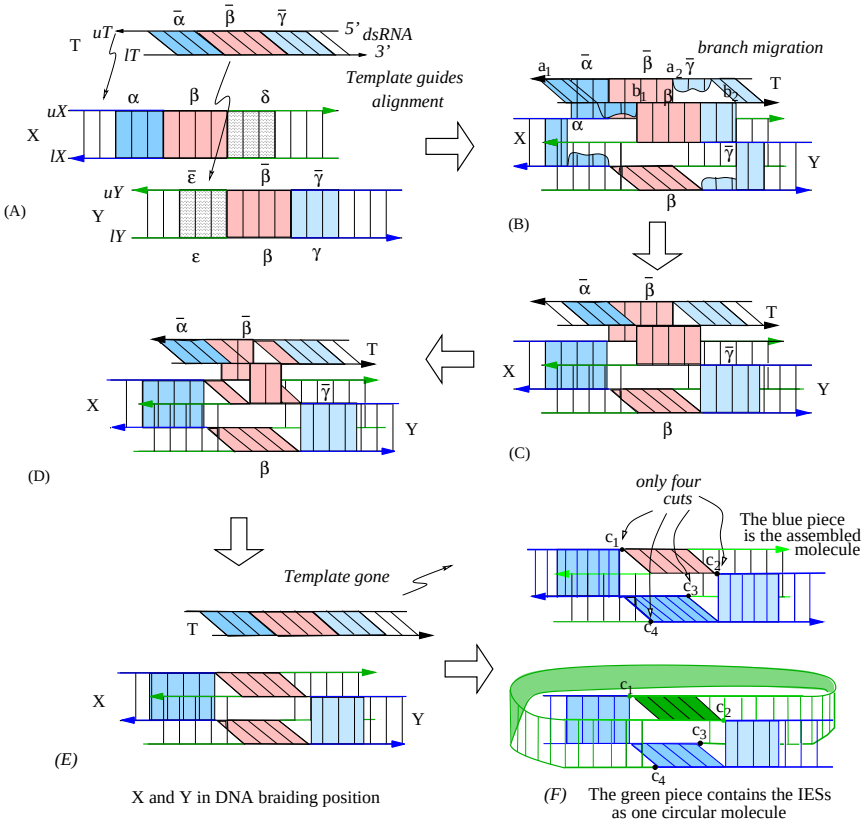

Figure 3: Step by step model of DNA braiding guided by dsRNA template.

We propose a dsRNA template T, such that its "upper" strand in direction $3' - 5'$ (denoted with uT) has a block $\bar{\alpha}\bar{\beta}\bar{\gamma}$ composed of sequences $\bar{\alpha}$, $\bar{\beta}$ and $\bar{\gamma}$. The lower strands of T, X and Y (denoted lT, lX, lY) are complementary to the upper strands. The proposed steps of the recombination are as follows:

[A.] All three molecules X, Y and T are present in the environment at the same time and the template strands find their corresponding complements in molecules X and Y as shown in Fig. 3(A). We postulate that the template is short enough to initiate branch migration.

Even if the pointer sequence β is as short as two nucleotides and occurs more than twice in the DNA sequence, the context of β in T ($\alpha\gamma$), the left context in X (α) and the right context in Y (γ) would be sufficient to lead to the alignment of the correct pointer sequences.

[B.] Through branch migration the ends of the strands of template T, once in a neighborhood of complementary sequences, can easily anneal with their complements. An unzipping of the three double-stranded stripes occurs, from point a_1 to a_2 on X, from b_1 to b_2 on Y and from a_1 to b_2 on T (see Fig. 2). A portion of lX and a portion of lY, containing $\bar{\beta}$ and β, respectively, become single-stranded. Because they are in close proximity of each other and single-stranded, hydrogen bonds form between the complementary regions connecting lX and lY, as shown in Fig. 3(B).

The original pairing and the new pairing are considered probabilistic. At some point during this process, cuts are made at c_1, c_2, c_3 and c_4 on the lower and upper backbones of X and Y as shown

on Fig. 3(F). These cuts may depend on the way in which the pointers align and which portion of the pointer sequence participates in the branch migration process.

[C.] Note that the substrings of α, β, γ, $\bar{\alpha}$, $\bar{\beta}$ and $\bar{\gamma}$ that can not find their complementary strings might remain unpaired. RNA-DNA hybrids are stronger than DNA-DNA, so the process (to be successful) may be thermodynamically driven. In the situation depicted in Fig.3(C), only the single-stranded subsequences $\bar{\beta}$ and β from T hybridize to the corresponding complementary sequences of uX and uY.

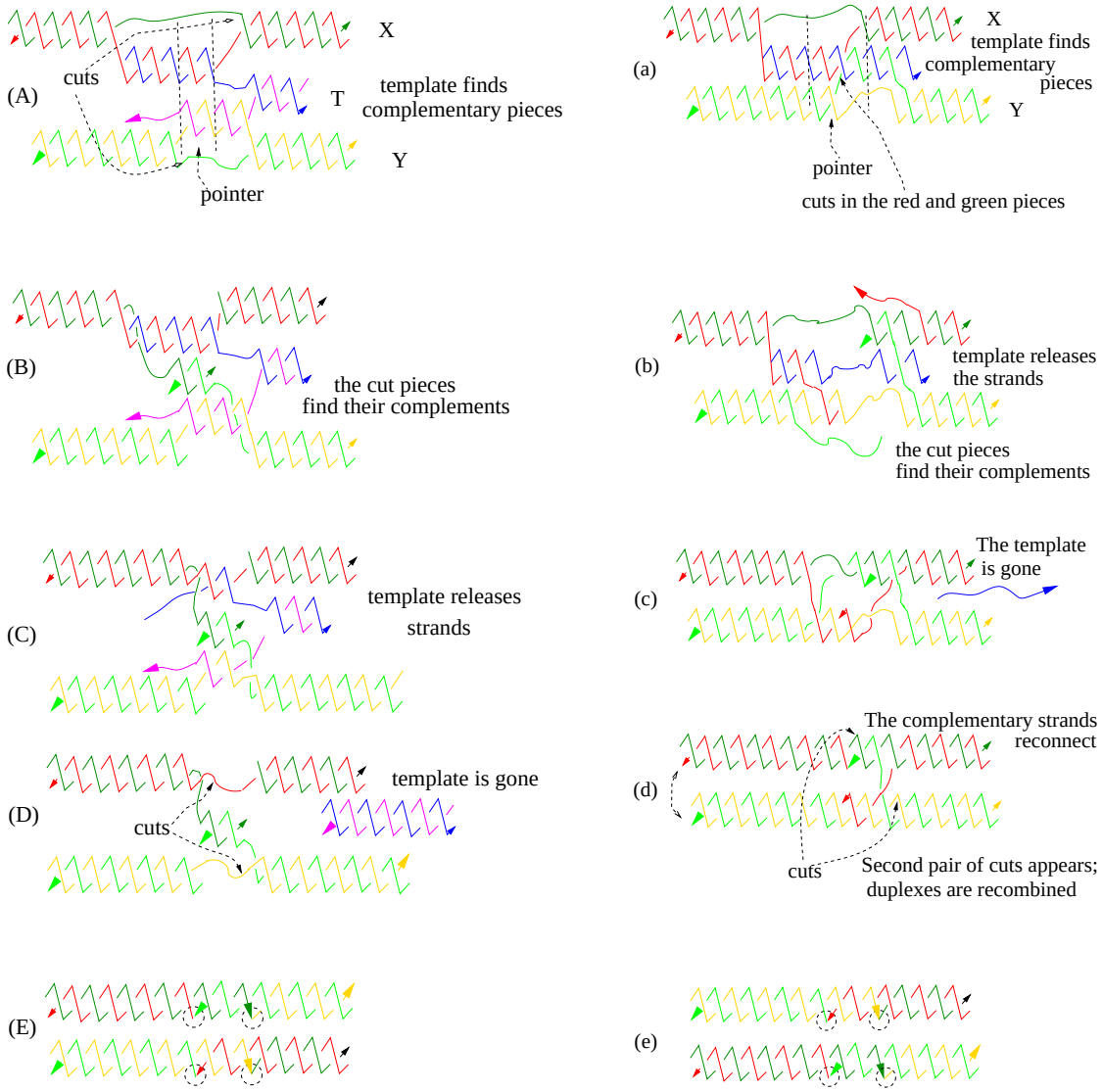

Figure 4: Schematic view into the postulated steps of strand branch migration in the braiding process described in previous sections.

[D.] In the next step, illustrated in Fig. 3(D), the hydrogen bonds, between uX and uT on one hand and lT and uY on the other hand, start to dissociate. Through branch migration, because RNA duplexes are more stable than RNA-DNA duplexes, the template strands release uX and uY. At the same time, enabled by strand complementarity, new hydrogen bonds develop between uX and uY.

[E.] Branch migration permits the complementary regions of uX and uY corresponding to β to hybridize, releasing the template (shown in Fig. 3(E)). Thus the template is available to serve for further recombinations if needed.

We refer to the pairing between molecules X and Y shown in Fig. 3(E) as a *DNA vertex*. This portion shown in Fig. 3(E) is a molecule that has been studied and characterized *in-vitro* before (e.g., [135]) as a type of DX molecule known as 'double parallel cross over molecule'.

[F.] Fig. 3(F) shows the resulting molecules obtained after the cuts are introduced at $c_1, \ldots, c_4$. The blue portion of the braiding molecule indicates the new recombined molecule containing the sequence $\alpha\beta\gamma$. If we view this process schematically, when the cuts are introduced relieving possible strain, the right portion of molecule Y rotates towards molecule X ("falls down") and the left portion of molecule X rotates towards molecule Y (also "falls down"), permitting the nicks to be ligated, forming product strands.

Assuming that the portions that have undergone recombinations, portions X and Y, belong to the same DNA molecule, after recombination, the remaining fragments (containing sequence $\epsilon\beta\delta$) could be released as a circular molecule, indicated with green in Fig. 3(F). Schematically, in this case, the left portion of molecule Y rotates towards molecule X ("goes up") and the right portion of molecule X rotates towards molecule Y (also "goes up") at which point, the nicks are ligated. With the braiding process, the part of the molecule that needs to be extracted easily can become circular when it is removed.

The proposed dsRNA template guided recombination is presented in Fig. 4 on the left and the case of ssRNA template is presented in Fig. 4 on the right. Both cases can be described in a similar way. In these figures, the double helical structures of DNA are taken into considerations, and possible physical positions during the proposed recombination processes are shown.

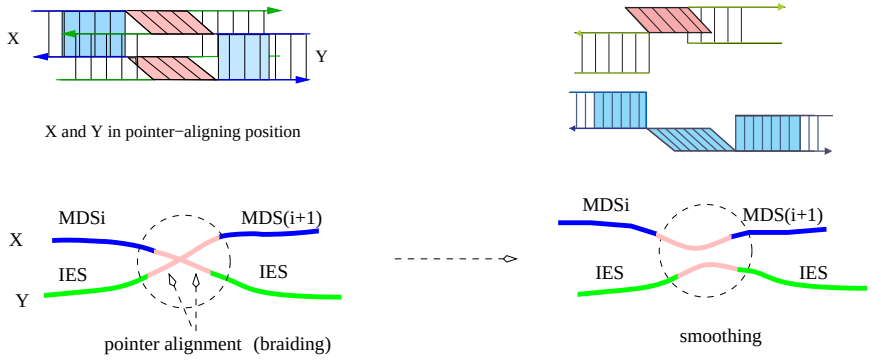

Figure 5: **(Left)** Schematic representation of the pointer alignment shown as a 4-valent vertex. Two micronuclear DNA segments exchange pointer nucleotides through branch migration as in Fig. 3(E) with MDS segments indicated in blue (top). This alignment is represented as a 4-valent vertex in a graph (bottom). **(Right)** Two molecules after the recombination (top), with the MDSs joined on the same molecule. Schematic representation of the finished recombination as smoothing of the vertex (bottom).

4 Geometric models of DNA recombinations

The model in [2] utilizes graphs as a physical representation of the DNA at the time of recombination. Schematically, the moment of homologous recombination, the alignment of the pointers can be represented as a vertex in a graph as depicted in Fig.5 (left). The pointer alignment and recombination can be seen as a 4-valent rigid vertex v made of two DNA segments where each edge e, incident to v, has predefined "predecessor" edge, and a predetermined "successor" edge with respect to v, which constitute the *neighbors* of e. The homologous recombination corresponds to removal of the crossing (vertex) called "smoothing" (see Fig.5 (right)). The RNA or DNA template enables alignment of pointers and, as shown in [2].

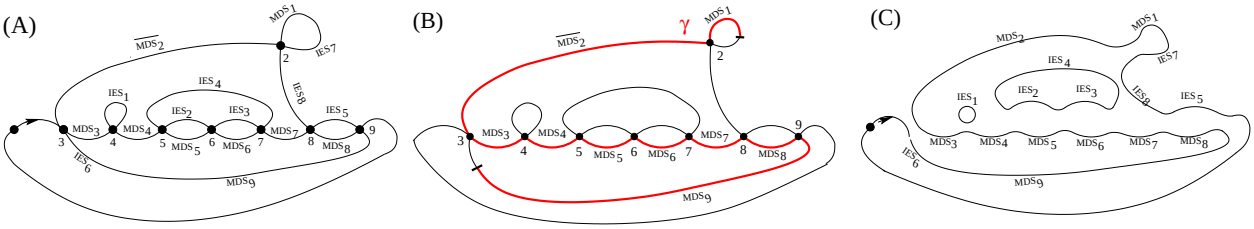

Figure 6: (A) Graph structure of simultaneous recombination for Actin I gene, at the moment of recombination; (B) polygonal path in the graph containing the MDSs for the macronuclear gene; (C) smoothing of the vertices relative the polygonal path in (B) and the resulting molecules after recombination.

The graph depicted in Fig. 6(A) is a schematic planar representation of the MDS-IES structure of the scrambled gene in Fig. 1 at the time of rearrangement. Joining the neighboring MDSs forms a path in the graph which in some sense determines the smoothings of the vertices as depicted in Fig. 6(B). The resulting graph, after smoothing of every vertex, is depicted in Fig. 6(C). As shown, this result is composed of three connected components, although in reality these molecules are likely non-circular. Two of them are labeled only with IESs indicating IES excisions, while one of the components contains $\mathrm{MDS}_1 - \mathrm{MDS}_2 - \cdots - \mathrm{MDS}_9$, i.e., the assembled macronuclear gene in correct MDS order. In [2] and subsequently in [4] we showed that every MDS-IES micronuclear gene structure can be modeled by a spatial graph, and the assembly of a macronuclear gene can be viewed as a smoothing of every vertex in the graph.

In these models, a micronuclear sequence is represented by an *assembly graph* which is a finite connected graph with 4-valent rigid vertices. We often allow assembly graphs to have two end points (1-valent vertices) representing the initial and terminal points of a DNA molecule, instead of circular molecules. A macronuclear gene consisting of the ordered MDS segments are modeled with a *polygonal path*, an open path such that consecutive edges are neighbors with respect to the joint incident vertex. In other words, a polygonal path makes a 90 degree turn at every rigid 4-valent vertex. A *(transverse) component* of an assembly graph is a path of maximal length containing non-repeating edges, such that no two consecutive edges are neighbors. The graph in Fig. 6 (A) consists of a single component made of edges labeled with the micronuclear sequence of MDSs and IESs.

There are two possibilities of cuts when recombination occurs. In Fig. 7 (A), a DNA molecule aligned at a pointer, when the cuts are about to occur, is depicted with half a twist with the double

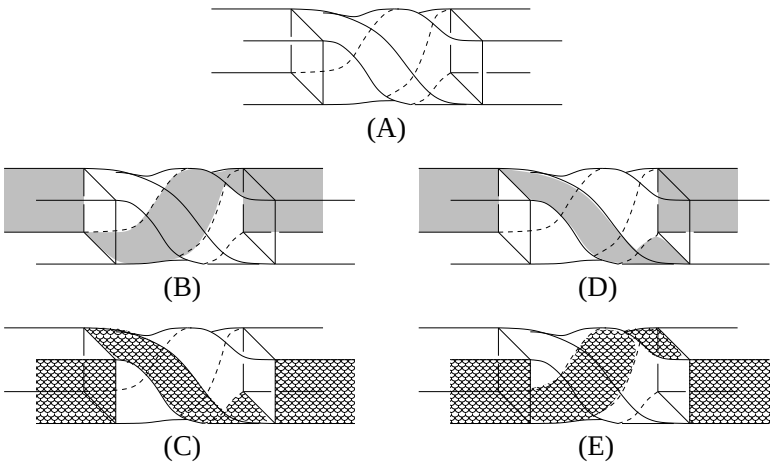

Figure 7: Twisted box in a template model

helical structure. Since it has half a twist, it is assumed that the length of a pointer is about several base pairs. In (B) and (C), two strands are depicted with a particular choice of cuts, after a recombination, In (D) and (E), the other choice for cuts is depicted. The situation depicted in Fig. 7 (B) and (C) resembles the schematic diagram at the bottom right of Fig. 5.

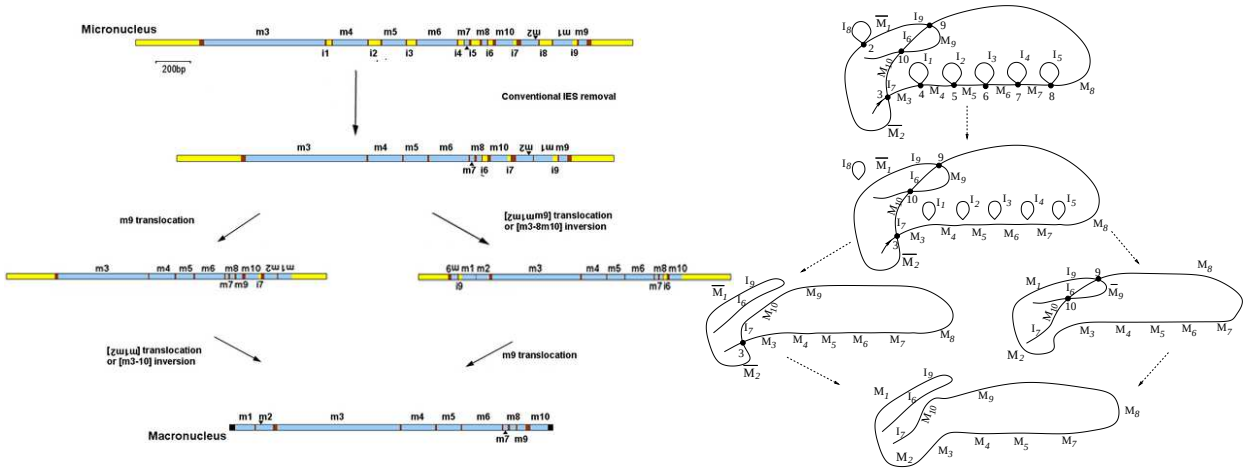

Figure 8: **(Left)** Two possible pathways to assemble the *S. lemnae actin I* gene from its precursor form, experimental data from [118]. **(Right)** Two assembly strategies corresponding to the pathways.

5 Pathways of DNA recombination processes

DNA rearrangements may occur in more than one ways. Different sequences of DNA rearrangement processes that result in the same final molecule are called *pathways*. For the *S. lemnae actin I* gene two different assembly pathways have been detected [118], which theoretically correspond to two distinct assembly strategies, as depicted in Fig. 8.

References

[1] K. Ameur, M. Elhamdadi, T. Rose, C. Smudde, M. Saito, *Tangle Embeddings and Quandle Cocycle Invariants,* to appear in Experimental Mathematics.

[2] A. Angeleska, N. Jonoska, M. Saito, L.F. Landweber, *RNA-Guided DNA Assembly,* J. of Theoretical Biology 248(4):706–720.

[3] Angeleska, A.; Jonoska, N.; Landweber, L.F.; Saito, M., *Strategies for RNA-guided DNA Recombination,* to appear in a book of collected papers "Algorithmic Bioprocesses", Natural Computing Series, Springer Verlag.

[4] A. Angeleska, N. Jonoska, M. Saito, *DNA Recombinations through assembly graphs,* Discrete Applied Mathematics 157 (2009) 3020-3037

[5] A. Angeleska, N. Jonoska, M. Saito, L.F. Landweber, *Recombination pathways in DNA rearrangements in ciliates,* in preparation.

[6] D.H. Ardell, C.A. Lozupone, L.F. Landweber, *Polymorphism, recombination and alternative unscrambling in the DNA polymerase alpha gene of the ciliate Stylonychia lemnae,* Genetics 165(4) (2003) Dec;1761 - 1777.

[7] J. Arsuaga, T. Blackstone, Y. Diao, K. Hinson, E. Karadayi, M. Saito, *Sampling large random knots in a confined space,* J. Phys. A: Math. Theor. 40 (2007), 11697 - 11711.

[8] J. Arsuaga, T. Blackstone, Y. Diao, E. Karadayi, M. Saito, *Linking of uniform random polygons in confined spaces,* J. Phys. A: Math. Theor. 40 (2007), 1925 - 1936.

[9] M. Asaeda, L.H. Przytycki, A.S. Sikora, *Categorification of the Kauffman bracket skein module of I-bundles over surfaces,* Algebr. Geom. Topol. 4 (2004), 1177-1210.

[10] J. C. Baez, *Link Invariants of Finite Type and Perturbation Theory,* Lett. Math. Phys. 26 (1992), no. 1, 43 - 51.

[11] D. Bar-Natan, *Khovanov's homology for tangles and cobordisms*, Geom. Topol. 9 (2005) 1443-1499.

[12] D. Bar-Natan, *Finite type invariants,* arXiv:math/0408182v3.

[13] J. S. Birman, *New Points of View in Knot Theory,* Bull. Am. Math. Soc. 28 (1993), no. 2, 253 - 287.

[14] S. Beermann, *The diminution of Heterochromatic chromosomal segments in Cyclops (Crustacea, Copepoda)*, Chromosoma, 1977. 60(4): p. 297 - 344.

[15] G.T. Bignell, T. Santarius, J.C. Pole, A.P. Butler, J. Perry, E. Pleasance, C. Greenman, A. Menzies, S. Taylor, S. Edkins, P. Campbell, M. Quail, B. Plumb, L. Matthews, K. McLay, P.A. Edwards, J. Rogers, R. Wooster, P.A. Futreal, M.R. Stratton, *Architectures of somatic genomic rearrangement in human cancer amplicons at sequence-level resolution*, .Genome Res, 2007. 17(9): 1296 - 303.

[16] P.J. Campbell et al., *Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing*, Nature Genetics 40 (2008), 722 - 729.

[17] J.S. Carter, A.S. Crans, M. Elhamdadi, M. Saito, *Cohomology of categorical self-distributivity,* J. Homotopy Relat. Struct. 3 (2008), no. 1, 13–63.

[18] J.S. Carter, A.S. Crans, M. Elhamdadi, M. Saito, *Cohomology of the adjoint of Hopf algebras,* J. Gen. Lie Theory Appl. 2 (2008), no. 1, 19–34.

[19] J.S. Carter, A.S. Crans, M. Elhamdadi, E. Karadayi, M. Saito, *Cohomology of Frobenius Algebras and the Yang-Baxter Equation,* to appear in Communications in Contemporary Mathematics.

[20] J.S. Carter, M. Elhamdadi, M. Graña, M. Saito, *Cocycle knot invariants from quandle modules and generalized quandle cohomology,* Osaka J. Math. 42 (2005), no. 3, 499–541.

[21] J.S. Carter, M. Elhamdadi, M. Saito, *Homology theory for the set-theoretic Yang-Baxter equation and knot invariants from generalizations of quandles,* Fund. Math. 184 (2004), 31-54.

[22] J.S. Carter, M. Elhamdadi, M. Saito, S. Satoh, *A lower bound for the number of Reidemeister moves of type III*, Topology and its Applications, 153 (2006), 2788–2794.

[23] J.S. Carter, S. Kamada, M. Saito, Surfaces in 4-space, Encyclopaedia of Mathematical Sciences, Vol. 142, Springer Verlag, 2004.

[24] J.S. Carter, M. Saito, *Set theoretic Yang-Baxter solutions via Fox calculus*, Journal of Knot Theory and its Ramifications, Vol. 15, No. 8 (2006) 949-956.

[25] J.S. Carter, M. Saito, *Cohomology of self-distributivity in coalgebras,* Preprint, submitted to conference proceedings of "Intelligence of Low Dimensional Topology 2006", Hiroshima, Japan, July 2006.

[26] J.S. Carter, M. Saito, *Cohomology for self-distributivity in coalgebras,* Intelligence of low dimensional topology 2006, 9–17, Ser. Knots Everything, 40, World Sci. Publ., Hackensack, NJ, 2007.

[27] J.S. Carter, M. Saito, S. Satoh, *Ribbon concordance of surface-knots via quandle cocycle invariants,* Journal of Australian Mathematical Society, 80 (2006), no. 1, 131–147.

[28] JS. Carter, M. Saito, S. Satoh,

Ribbon-moves for 2-knots with 1-handles attached and Khovanov-Jacobsson numbers, Proc. Amer. Math. Soc. 134 (2006) 2779-2783.

[29] A.R.O. Cavalcanti, L.F. Landweber, *Insights into a biological computer: detangling scrambled genes in ciliates*, Nanotechnology: Science and Computation (J. Chen, N. Jonoska, G. Rozenberg eds.) Springer (2006) 349–360.

[30] A.R.O. Cavalcanti, D.M. Dunn, R. Weiss, L.F. Landweber, T.G. Doak, *Sequence features of Oxytricha trifallax telomeric and subtelomeric sequences*, Protist 155 (2004) 311-22.

[31] A.R.O. Cavalcanti, L.F. Landweber, *GENE UNSCRAMBLER for detangling scrambled genes in ciliates,* Bioinformatics 20 (2004) 800-2.

[32] A.R.O. Cavalcanti, Stover, Orecchia, Doak, L.F. Landweber, *Coding properties of Oxytricha trifallax macronuclear chromosomes: analysis of a pilot genome project,* Chromosoma 113 (2004) 69.

[33] A.R.O. Cavalcanti, T.H. Clarke, L.F. Landweber, *MDSIESDB: a database of macronuclear and micronuclear genes in spirotrichous ciliates,* NAR 33:D (2005) 396?8.

[34] M. Cavaliere, N. Jonoska, S. Yogev, R. Piran, E. Keinan, N.C. Seeman *Biomolecular Implementation of Computing Devices with Unbounded Memory* (G. Ferretti, et. al. eds.) Springer LNCS 3384 (2005) 35–49.

[35] M. Cavaliere, N. Jonoska, P. Leupold, *DNA Splicing: Computing by Observing* Natural Computing (published online first DOI:10.1007/s11047-007-9062-8)

[36] B. Chakraborty, N. Jonoska, N.C. Seeman, *Programmable transducer by DNA self-assembly,* (submitted).

[37] W.J. Chang, P.D. Bryson, H. Liang, M.K. Shin, L.F. Landweber, *The evolutionary origin of a complex scrambled gene,* PNAS 102 (2005) 15149-54.

[38] W-J. Chang, S. Kuo, L.F. Landweber, *A new scrambled gene in the ciliate* Uroleptus, Gene 368 (2006) 72–77.

[39] W.J. Chang, Stover, Addis, L.F. Landweber, *A Micronuclear Locus Containing Three Protein-Coding Genes Remains Linked During Macronuclear Development,* Protist 155 (2004) 245-255.

[40] W.J. Chang, V.M. Addis, A.J. Li, E. Axelsson, D.H. Ardell, L.F. Landweber, *Intron Evolution and Information processing in the DNA polymerase alpha gene in spirotrichous ciliates: a hypothesis for interconversion between DNA and RNA deletion,* Biol Direct, 2007. 2: p. 6.

[41] J. Chen, N. Jonoska, G. Rozenberg, (eds) *Nanotechnology: Science and Computing*, Springer - Verlag 2006.

[42] L. Colson, N. Jonoska, M. Margenstern, *λP Systems and Typed Λ-calculus*, (invited paper) (G. Mauri et al. eds.) Springer LNCS 3365 (2005) 1–18.

[43] R. Daniel, *IES excision in* Sterkiella histriomuscorum *via a hypothetical nuclease mechanism,* (2006) senior thesis, Princeton University.

[44] C. Dennis, A. Fedorov, E. Käs., L. Salomé, M. Grigoriev, *RuvAB-directed branch migration of individual Holliday junctions is impeded by sequence heterology,* The EMBO Journal 23 (2004) 2413–2422.

[45] Doak, Cavalcanti, Stover, Dunn, Weiss, Herrick, Landweber, *Sequencing the Oxytricha trifallax macronuclear genome: a pilot project,* Trends Genetics 19 (2003) 603-7.

[46] T. Dobzhansky, *On the Sterility of the Interracial Hybrids in Drosophila Pseudoobscura,* Proc Natl Acad Sci U S A, 1933. 19(4): p. 397-403.

[47] A. Ehrenfeucht, T. Harju, I. Petre, D.M. Prescott, G. Rozenberg, *Computing in Living Cells* Springer 2005.

[48] The ENCODE Project Consortium (Ewan Birney et al.) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project, Nature 447 (2007) 799-816.

[49] E. Flapan, *When Topology Meets Chemistry, A Topological Look at Molecular Chirality,* Cambridge University Press, 2000.

[50] G. Franco, N. Jonoska, A. Plass, B. Osborn, *Knee Joint Injury and Repair Modeled by Membrane Systems* BioSystems 91 3 (2008) 473-488.

[51] G. Franco, N. Jonoska, *Forbidding-Enforcing Conditions in DNA Self-Assembly of Graphs*, in *Nanotechnology: science and computing* (J. Chen, et. al. eds.) Springer-Verlag (2006) 105 – 118.

[52] K. F. Gauss, *Zur mathematischen Theorie der electorodynamischen Wirkungen,* Werke Konigl. Gesell. Wiss. Gottingen 5:605.

[53] O. Garnier, V. Serrano, S. Duharcourt, E. Meyer, *RNA mediated programming of developmental genome rearrangements in* Paramecium tetraurelia*,* Mol. Cell Biol. 24 (2004) 7370–7379.

[54] D. Genova, N. Jonoska, *Defining Structures through Forbidding and Enforcing Constraints* to appear Phisica B 394, Issue 2, 15 May 2007, 306–310.

[55] D. Genova, N. Jonoska, *Topological properties of forbidding-enforcing systems* Automata, Languages and Combinatorics, 11 4 (2006) 375–397.

[56] S.A. Gerbi, *Unusual chromosome movements in sciarid flies,* Results Probl Cell Differ, 1986. 13: p. 71-104.

[57] M. Gerstenhaber, S.D. Schack, *Algebras, bialgebras, quantum groups, and algebraic deformations,* in: Deformation theory and quantum groups with applications to mathematical physics (Amherst, MA, 1990), 51–92, Contemp. Math., 134, Amer. Math. Soc., Providence, RI, 1992.

[58] A. Girard, R. Sachidanandam, G. J. Hannon, M.A. Carmell, *A germline-specific class of small RNAs binds mammalian Piwi proteins,* Nature. 442 (2006) 199–202.

[59] T. Head, *Formal language theory and DNA: an analysis of the generative capacity of specific recombinant behaviors*, *Bull. Math. Biology* 49 (1987) 737–759.

[60] L. Helme-Guizon, Y. Rong, *A categorification for the chromatic polynomial,* Algebr. Geom. Topol. 5 (2005), 1365–1388.

[61] B. P. Higgins, C.D. Carpenter, A. C. Karls, *Chromosomal context directs high-frequency precise excision of IS492 in* Pseudoalteromonas atlantica*,* PNAS 104 (2007) 1901–6.

[62] J. Hoste; M. Thistlethwaite; J. Weeks, *The first 1,701,936 knots,* Math. Intelligencer 20 (1998) 33-48.

[63] A. Ishii, K. Tanaka, *A categorification of the one-variable Miyazawa polynomial,* A pdf file of a talk at a conference via private communication, February 6, 2008.

[64] E.F. Jasso-Hernandez, Y. Rong, *A categorification for the Tutte polynomial,* Algebr. Geom. Topol. 6 (2006), 2031–2049.

[65] N. Jonoska, M. Saito, *Boundary components of thickened graphs,* Proceedings of DNA7, LNCS 2340, Springer-Verlag, (2002) 70-81. Editors: N. Jonoska, N. Seeman.

[66] N. Jonoska, G. McColm *Flexible versus Rigid Tiles*, Unconventional Computers (Gh. Paun et al. eds.) Springer LNCS 4135 (2006) 139–151.

[67] N. Jonoska, Gh. Păun, eds. Special issue on Membrane Computing, New Generation Computing, 22 No. 4 (2004).

[68] N. Jonoska, G. Wu, N.C. Seeman, *Existence of single-stranded reporters in DNA-based graph structures* to appear Theoretical Computer Science.

[69] E. Dolzhenko, N. Jonoska, *On complexity of two dimensional languages generated by transducers* in Implementation and Application of Automata (O. Ibaraa et al eds.) Springer LNCS 5148 (2008) 181 –190.

[70] N. Jonoska, G. McColm, *Complexity classes for self-assembling flexible tiles* to appear in Theoretical Computer Science.

[71] N. Jonoska, L. Kari, K. Mahalingam, *Involution join and solid codes,* Fundamenta Informaticae 86 1–2 (2008) 127–142.

[72] N. Jonoska, R. Twarock, *Blueprints for dodecahedral DNA cages*, J. of Physics A: Mathematical and Theoretical 41 (2008) 304043 (14pp) .

[73] N. Jonoska, A. Taormina, R. Twarock, *DNA cages with icosahedral symmetry in bionanotechnology* to appear.

[74] N. Jonoska, *Biomolecular Automata in* NanoBioTechnology, Bioinspired Devices and Materials of the Future, Chapter 11, (S. Oded, I. Levi eds.) Humana Press 2008.

[75] N. Jonoska, G. McColm, *Describing Self-assembly of Nanostructures*, SOFSEM (V. Geffert et al eds.) Springer LNCS 4910 (2008) 66–73.

[76] N. Jonoska, J. Pirnot, *Finite State Automata Representing Two-dimensional Subshifts,* CIAA 2007 (J. Holub, J. Žd'árek eds.) Springer LNCS 4783 (2007) 277–289.

[77] N. Jonoska, Gh. Paun, G. Rozenberg, (eds.) *Aspects of Molecular Computing*, Springer LNCS 2950, Springer-Verlag 2004.

[78] N. Jonoska, *Trends in computing with DNA*, *Journal of Computer Science and Technology* 19 No.1 (2004) 98-100.

[79] N. Jonoska, *Biomolecular Automata* Bulletin of EATCS 88 (2006) 141– 166.

[80] N. Jonoska, G. McColm, A. Staninska, *Spectrum of a pot for DNA complexes* in DNA Computing 12 (C. Mao, T. Yokomori eds.) Springer LNCS 4287 (2006) 83–94.

[81] N. Jonoska, G. McColm, A. Staninska, *Expectation and Variance of Self-Assembled Graph Structures* Proceedings of the 11th meeting of DNA based computers, in DNA Computing (DNA11), (A. Carbone N.A. Pierce, eds.) Springer LNCS 3892 (2006) 144–157.

[82] N. Jonoska, L. Kari, K. Mahalingam *Involution Solid and Join Codes*, Developments in Language Theory, (O.H. Ibarra, Z. Dang eds.) Springer LNCS 4036 (2006) 192-202.

[83] N. Jonoska, J. Pirnot, *Transitivity in two-dimensional local languages defined by dot systems*, Int. J. of Foundations of Computer Science, 17(2) (2006) 435 – 463.

[84] N. Jonoska, S. Liao, N.C. Seeman, *Transducers with Programmable Input by DNA Self-assembly*, in: *Aspects of Molecular Computing* (N. Jonoska, Gh. Paun, G. Rozenberg eds.), Springer LNCS 2950 (2004) 219-240.

[85] N. Jonoska, K. Mahalingam, *Languages od DNA based code words*, in: *DNA Computers*, Revised papers of the 9th meeting on DNA based computers (J.Chen, J. Reif eds.), Springer LNCS 2943 (2004) 61-73.

[86] N. Jonoska, K. Mahalingam, J. Chen, *Involution Codes; With Aplication to DNA Coded Languages* Natural Computing 4 (2) (2005) 141–162.

[87] N. Jonoska, M. Margenstern, *Tree operations in P-systems and* Λ*-calculus*, *Fundamenta Informaticae* 59 No.1 (2004) 67-90.

[88] N. Jonoska, M. Saito, *Algebraic and topological models for DNA recombinant processes* (invited paper) (C.S. Calude, E. Calude, M.J. Dinneen eds.) Springer LNCS 3340 (2004) 49—62.

[89] N. Jonoska, K. Mahalingam, *Methods for Constructing Coded DNA Languages*, in: *Aspects of Molecular Computing*, (N. Jonoska, Gh. Paun, G. Rozenberg eds.), Springer LNCS 2950 (2004) 241-253.

[90] N. Jonoska, S. Karl, M. Saito, *Three dimensional DNA structures in computing, BioSystems* 52 (1999) 143-153.

[91] N. Jonoska, G. McColm, *A Computational Model for Self-assembling Flexible Tiles* (C.S. Calude et. al. eds.) LNCS 3699 (2005) 142–156.

[92] F. Jonsson, G. Steinbruck, H.J. Lipps, *Both subtelomeric regions are required and sufficient for specific DNA fragmentation during macronuclear development in Stylonychia lemnae,* Genome Biology 2 (2001) research0005.1-0005.11.

[93] S.A. Juranek, S. Rupprecht, J. Postberg, H.J. Lipps, *snRNA and Heterochromatin Formation Are Involved in DNA Excision during Macronuclear Development in Stichotrichous Ciliates,* Eukaryot. Cell 4 (2005) 1934–1941.

[94] L. Kadison, *New examples of Frobenius extensions,* University Lecture Series 14, Amer. Math. Soc., 1999.

[95] S. Kamada, *Braid and Knot Theory in Dimension Four,* Surveys and monographs 95, Amer. Math. Soc. (2002).

[96] L. Kari, L.F. Landweber, *Computational power of gene rearrangement,* DNA Based Computers V (E. Winfree, D.K. Gifford eds.) AMS (1999) 207–216.

[97] L. Kari, L.F. Landweber, *Biocomputation in ciliates,* in Cellular Computing (M. Amos, ed.), Oxford University Press (2004) 202?216.

[98] L.H. Kauffman, *Virtual Knot Theory*, European J. Combinatorics 20 (1999) 663–690.

[99] L.H. Kauffman, *A self-linking invariant of virtual knots,* Fund. Math. 184 (2004), 135–158.

[100] L. H. Kauffman, *Knots and Physics,* Series on Knots and Everything 1, World Sci. Publishing (1991).

[101] L. H. Kauffman, *Invariants of graphs in* 3*-space,* Trans. Amer. Math. Soc. 311 (1989), 697–710.

[102] M. Khovanov, *A categorification of the Jones polynomial*, Duke Math. J. 101(3) (1999), 359–426.

[103] C.A. Klein, O. Schmidt-Kittler, J.A. Schardt, K. Pantel, M.R. Speicher, G. Riethmuller, *Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells,* Proc Natl Acad Sci U S A, 1999. 96(8): p. 4494-9.

[104] Y. Kirino, Z. Mourelatos, *Mouse Piwi-interacting RNAs are 2-O-methylated at their 3 termini,* Nature Structural and Molecular Biology 14 (2007) 347-348.

[105] J. Kock, *Frobenius algebras and 2D topological quantum field theories,* London Mathematical Society Student Texts (No. 59), Cambridge University Press, 2003.

[106] S. Kuo, W-J. Chang, L.F. Landweber, *Complex germline architecture: two genes interwined on two loci,* Mol. Biol. Evol. 23(1) (2006) 4–6.

[107] L.F. Landweber, L Kari, *Universal Molecular Computation in Ciliates,* in: Evolution as Computation (Landweber, Winfree, eds.) (2002) Springer-Verlag, Berlin.

[108] L.F. Landweber, E. Winfree, eds. Evolution as Computation, (2002) Springer-Verlag, Berlin.

[109] L.F. Landweber, T-C. Kuo, E.A. Curtis, *Evolution and assembly of an extremely scrambled gene,* PNAS **97** 7 (2000) 3298–3303.

[110] L.F. Landweber, *Why genomes in pieces?,* (2007) Science 318:405–406.

[111] L.F. Landweber, *Making sense of scrambled genomes,* Science, 2008. 319(5865): p. 901-2.

[112] L.F. Landweber, I.D. Pokrovskaya, *Emergence of a dual-catalytic RNA with metal-specific cleavage and ligase activities: the spandrels of RNA evolution,* PNAS 96 (1999) 173-8.

[113] Li et al., *A neoplastic gene fusion mimics trans-splicing of RNAi in normal human cells,* Science Sept. 5, Vol. 321, No. 5894. (2008) 1357-1361.

[114] S.J. Lolle, J.L. Victor, J.M. Young, R.E. Pruitt, *Genome-wide non-mendelian inheritance of extra-genomic information in Arabidopsis,* Nature 434, 505?509 (2005).

[115] R. Lucito, J. West, A. Reiner, J. Alexander, D. Esposito, B. Mishra, S. Powers, L. Norton, M. Wigler, *Detecting gene copy number fluctuations in tumor cells by microarray analysis of genomic representations,* Genome Res, 2000. 10(11): p. 1726-36.

[116] McFarland, Chang, Kuo, Landweber, *Conserved linkage of two genes on the same macronuclear chromosome in spirotrichous ciliates,* Chromosoma 115 (2006) 129-38.

[117] K. Mochizuki, N.A. Fine, T. Fujisawa, M.A. Gorovsky, *Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in* Tetrahymena*,* Cell, 110 (2002) 689–699.

[118] M. Mollenbeck, Y. Zhou, A.R.O. Cavalcanti, F. Jonsson, W.-J. Chang, S. Juranek, T.G. Doak, G. Rozenberg, H.J. Lipps, L.F. Landweber, *The pathway for detangling a scrambled gene,* (2008) PLoS ONE 3(6):e2330.

http://www.plosone.org/article/info:doi/10.1371/journal.pone.0002330

[119] K. Murasugi, *Knot theory and its applications,* Translated from the 1993 Japanese original by Bohdan Kurpita. Birkhuser Boston, Inc., Boston, MA (1996).

[120] M. Nowacki, V. Vijayan, Y. Zhou, T. Doak, E. Swart, L.F. Landweber, *RNA-template guided DNA recombination: Epigenetic reprogramming of a genome rearrangement pathway,* Nature (2008) 451:153–158.

[121] M. Nowacki, G. Maquilan, B. Higgins, E. Swart, T. Doak, L.F. Landweber. (in revision).

[122] T. Ohara, Y. Sakaguchi, T. Suzuki, H. Ueda, K. Miyauchi, T. Suzuki, *The 3 termini of mouse Piwi-interacting RNAs are 2-O-methylated,* Nature Structural and Molecular Biology 14 (2007) 349-350.

[123] H.A. Orr, Dobzhansky, Bateson, and the genetics of speciation. Genetics, 1996. 144(4): p. 1331-5.

[124] A.G. Paschka, V. Horejschi, F. Jonsson, A. Lindecke, G. Weier, C. Kaltschmidt, H.J. Lipps, *A microarray analysis of developmentally regulated genes during macronuclear differentiation in the stichotrichous ciliate Stylonychia lemnae,* Gene, 2005. 359: p. 81-90.

[125] H. Pearson, *What is a gene?,* Nature 441 (2006) 399–401.

[126] D.M. Prescott, *The DNA of ciliated protozoa,* Microbiol Rev, 1994. 58(2): p. 233-67.

[127] D.M. Prescott, A.F. Greslin, *Scrambled actin I gene in the micronucleus of* Oxytricha nova*,* Dev Genet. 13(1) (1992) 66–74.

[128] D.M. Prescott, A. Ehrenfeucht, G. Rozenberg, *Template-guided recombination for IES elimination and unscrambling of genes in stichotrichous ciliates,* J. of Theoretical Biology 222 (2003) 323–330.

[129] H.M. Robertson, J.A. Navik, K.K. Walden, H.W. Honegger, *The Bursicon Gene in Mosquitoes: An Unusual Example of mRNA Trans-splicing,* Genetics 176 (2007) 1351-3.

[130] D. Rolfsen, *Knots and Links*, Corrected reprint of the 1976 original. Mathematics Lecture Series, 7. Publish or Perish, Inc., Houston, TX, 1990.

[131] C. Rourke; B. Sanderson, *A new classification of links and some calculations using it,* arXiv:math/0006062.

[132] P. Sa-Ardyen, N. Jonoska, N. Seeman, *Self-assembly of graphs represented by DNA helix axis topology* Journal of American Chemical Society, 126(21) (2004) 6648–6657.

[133] M. Saito, S. Satoh, *The twist spun trefoil needs four broken sheets,* Journal of Knot Theory and its ramifications Vol. 14, No. 7 (2005), 853-858.

[134] Satoh, S., *On the chirality of Suzuki's θ_n-curves*, Preprint.

[135] N.C. Seeman, *DNA in a Material World,* Nature 421 (2003) 427–431.

[136] F. Storici, K. Bebenek, T.A. Kunkel, D.A. Gordenin, M.A. Resnick, *RNA-templated DNA repair,* Nature 447 (2007) 338-41.

[137] Stover, Cavalcanti, Li, Richardson, Landweber, *Reciprocal fusions of two genes in the formaldehyde detoxification pathway in ciliates and diatoms,* Mol. Biol Evol. 22 (2005) 1539-42.

[138] D.W. Sumners, Lifting the curtain: using topology to probe the hidden action of enzymes, *Notices of AMS* 42 (5) (1995) 528–537.

[139] S.L. Tausta, L.A. Klobutcher, *Detection of circular forms of eliminated DNA during macronuclear development in* E. crassus, Cell 59 (6) (1989) 1019–1026.

[140] M. Thistlethwaite, *Kauffman's polynomial and alternating links,* Topology 27 (1988), 311–318.

[141] J.W. Wen, C. Maercker, H.J. Lipps, *Sequential excision of internal eliminated DNA sequences in the differentiating macronucleus of the hypotrichous ciliate Stylonychia lemnae,* Nucleic Acids Research, 24(22) (1996) 4415–4419.

[142] L.C. Wong, L.F. Landweber, *Evolution of Programmed DNA Rearrangements in a Scrambled,* Gene, Mol. Biol. Evol. 23 (2006) 756-63.

[143] G. Wu, N. Jonoska, N.C. Seeman, *Construction of a DNA nano-object directly demonstrated computation,* (submitted).

[144] H. Yan, X. Zhang, Z. Shen, N.C. Seeman, *A robust DNA mechanical device controlled by hybridization topology*, Nature 415 (2002) 62–65.

[145] B. Yurke, A.J. Turberfield, A.P. Mills, F.C. Simmel Jr., *A DNA fueled molecular machine made of DNA*, Nature 406 (2000) 605–608.

[146] D. Zerbib, C. Mezard, H. George, S.C. West., *Coordinated actions of RuvABC in Holliday junction processing,* J Mol Biol. 281(4) (9998) 621–630.

[147] Y. Zhou, L.F. Landweber, *BLASTO: a tool for searching orthologous groups,* Nucleic Acids Res. (2007) 35:W678-82.