Graphs for Modeling DNA Recombination Processes

— Background on Assembly Graphs and Related Concepts —

August 30, 2011

1 Introduction

This document is posted at http://math.usf.edu/∼saito/DNAweb. This is an overview of a series of work [2, 3, 4] on template models of DNA recombination and their combinatorial studies from point of views of graphs and knot theory. 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 Assembly graphs

Definitions of assembly graphs and related concepts are listed below from [4] for the rest of this subsection.

- An *assembly graph* is a finite connected graph, where all vertices are rigid vertices of valency 1 or 4. A vertex of valency 1 is called an *end point*.

  Note that the definition of assembly graph implies that the number of end points is always even.

  In Fig. 1 examples of assembly graphs are depicted. Note that the graphs in (B) and (C) do not have any end points, and the broken under-arc depicts the crossing information as in standard knot diagrams as already mentioned.

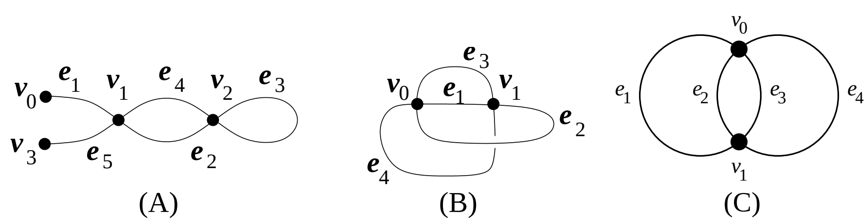

Figure 1: Examples of assembly graphs. Simple assembly graphs (A) and (B), and a non-simple assembly graph (C).

- A *rigid* vertex v of valency 4 is a pair $(v, (e_1, e_2, e_3, e_4)^{\text{cyc}})$ the equivalence class of $(e_1, e_2, e_3, e_4)^{\text{cyc}}$ of a cyclic order (e_1, e_2, e_3, e_4). Specifically, the equivalence relation $\sim$ of cyclic order for quadruplets $((e_1, e_2, e_3, e_4))$ is generated by $(e_1, e_2, e_3, e_4) \sim (e_2, e_3, e_4, e_1)$ and $(e_1, e_2, e_3, e_4) \sim (e_4, e_3, e_2, e_1)$.

  Topologically, a 4-valent rigid vertex is regarded as a fat vertex (disk) with four edges attached in order.

- The number of 4-valent vertices in Γ is denoted with $|\Gamma|$. The assembly graph is called *trivial* if $|\Gamma| = 0$.

  Topologically, a trivial assembly graph is regarded as a set of linear segments, with end points identified with 1-valent vertices, and segments as edges.

- Two assembly graphs are *equivalent* if there is an isomorphism of graphs between them that preserves the cyclic order of edges at each rigid vertex.

- Denote by $E(v)$ the set of edges incident to a vertex v.

- For each rigid vertex $(v, (e_1, e_2, e_3, e_4)^{\text{cyc}})$ of an assembly graph, the edges e_2 and e_4 are called *neighbors* of e_1.

  Neighbors are adjacent edges, and forms "90 degree". An edge and its non-neighbor form a "straight" line through the vertex.

- A *transverse path* in an assembly graph Γ is a sequence $\gamma = (v_0, e_1, v_1, e_2, \ldots, e_n, v_n)$ if v_0, v_n are endpoints, or $(v_0, e_1, v_1, e_2, \ldots, e_n)$, if v_0 is a 4-valent vertex and $e_n \in E(v_0)$, satisfying the following conditions:

  (1) $\{v_0, \ldots, v_n\}$ is a sequence of a subset of vertices of Γ, with possible repetition of the same vertex at most twice,

  (2) $\{e_1, \ldots, e_n\}$ is a set of distinct edges, and

  (3) each e_i is not a neighbor of e_{i-1} with respect to the rigid vertex v_{i-1}, $i = 2, \ldots, n$, and in the case where v_0 is a 4-valent vertex, e_1 is not a neighbor of e_n with respect to the rigid vertex v_0.

  A transverse path can be considered as an image of a map from the unit interval $[0, 1]$ to Γ, where the image of the boundary points $(\{0\} \cup \{1\})$ consists of either two end points of Γ, or a single 4-valent vertex.

  A transverse path for each example is given by

$$\begin{array}{cc} (A) & (v_0, e_1, v_1, e_2, v_2, e_3, v_2, e_4, v_1, e_5, v_3), \\ (B) & (v_0, e_1, v_1, e_2, v_0, e_3, v_1, e_4), \\ (C) & (v_0, e_2, v_1, e_4). \end{array}$$

  In the case of (C), there is another transverse path (v_0, e_3, v_1, e_1), which is not equivalent to (v_0, e_2, v_1, e_4).

- Each transverse path is also called a *transverse component* when it is regarded as a topological object (the image of an immersed interval or circle on a graph).

- Let $\gamma = (v_0, e_1, v_1, e_2, \ldots, e_n, v_n)$ or $(v_0, e_1, v_1, e_2, \ldots, e_n)$ be a transverse path, with or without end points, respectively. The *reverse* γ^R of γ is the transverse path $(v_n, e_n, \ldots, e_1, v_0)$ or $(e_n, \ldots, e_1, v_0)$, respectively.

- Two transverse paths with end points are *equivalent* if they are either identical, or, one is the reverse of the other. Two transverse paths γ, γ' without end points are *equivalent* if they have the same cyclic order: $\gamma^{\text{cyc}} = \gamma'^{\text{cyc}}$.

- An assembly graph Γ is called *simple* if there is a transverse *Eulerian* path in Γ, meaning, there is a transverse path γ that contains every edge from Γ exactly once.

- Given a simple assembly graph Γ with two endpoints, choose one of them to be initial (i) and the other to be terminal (t). We call such Γ a *directed* simple assembly graph with direction from i to t.

  We consider the transverse path of a directed simple assembly graph as a path starting at the vertex i and terminating at the vertex t.

- A *composition* $\Gamma_1 \circ \Gamma_2$ of two (directed simple) assembly graphs Γ_1 and Γ_2 with end points is the directed simple assembly graph with end points, obtained by identifying the terminal vertex of Γ_1 with the initial vertex of Γ_2.

  Note that initial vertex of $\Gamma_1 \circ \Gamma_2$ is the initial vertex of Γ_1 and terminal vertex of $\Gamma_1 \circ \Gamma_2$ is the terminal vertex of Γ_2. In general, $\Gamma_1 \circ \Gamma_2$ is not isomorphic to $\Gamma_2 \circ \Gamma_1$.

- A *polygonal* path is a path $(v_{i_1}, e_{i_2}, \ldots, v_{i_k})$ such that for any $j = 1, \ldots, k-1$, e_{i_j} and $e_{i_{j+1}}$ are neighbors.

- A *Hamiltonian* path is a path that visits every vertex exactly once.

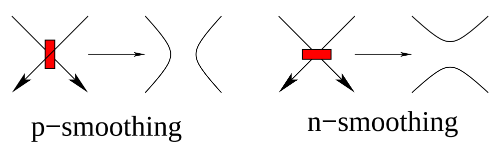

Figure 2: Smoothings of a vertex

- A *smoothing* of a rigid vertex is a replacement of the vertex by two parallel strings without crossings as depicted in Fig. 2. If an assembly graph is directed, then one is compatible with the orientation (Fig. 2 left), called the *p-smoothing*, and the other is not compatible (Fig. 2 right), called the *n-smoothing*.

3 Assembly words

In this section we review definitions of words related to assembly graphs, see [4] for more details.

- Let A be an alphabet set. A *double occurrence word* (DOW) over A is a word which contains each symbol of A exactly 0 or 2 times.

  Words w_1 over A_1 and w_2 over A_2 are *equivalent* if there is a bijection between A_1 and A_2 under which w_1 agrees with w_2.

- If $w = a_1 \dots a_k$ is a word over A, its reverse w^R is defined by $a_k \dots a_1$.

  Two words w_1, w_2 are reverse equivalent if w_1 is equivalent to w_2 or w_2^R.

- An *assembly word* is a double occurrence words with reverse equivalence.

- Let Γ be a simple directed assembly graph with two end points, the initial vertex i and the terminal vertex t, and let γ be the transverse path from i to t. Identify the set of 4-valent vertices $V = \{v_1, \dots, v_n\}$, where $n = |\Gamma|$, with an alphabet set. Starting from i, read off the vertices along γ that are encountered in order, to obtain a word in V called an *assembly word* of Γ.

  An assembly word is a double occurrence word over V such that each symbol of V appears exactly twice.

- Conversely, for an assembly word w over A such that each symbol of A appears exactly twice in w, there is an assembly graph Γ with the specified assembly word.

  The equivalence classes of assembly words and those of assembly graphs are in one-to-one correspondence. The assembly graph corresponding to an assembly word w is denoted by Γ_w.

- Let $[n] = \{1, \dots, n\}$. A DOW $w = a_1 \cdots a_{2n}$ over $[n]$ written in such a way that the first appearance of a symbol must be one greater than the largest of all preceding symbols, is an assembly word in *ascending order.*

- Let γ^R be the reverse of γ. Let w and w^R be assembly words over V that are obtained from γ and γ^R. Then w and w^R are defined to be equivalent. We say that two assembly words are equivalent if it is the same as itself or its reverse.

- A *circular word* over an alphabet A is an equivalence class of words over A, where two words are equivalent if they are related by a reverse and/or a cyclic permutation. We sometimes denote the circular word by $[w]$ for a word w, but often abbreviate the notation and say "a circular word w" if no confusion arises.

  A regular (non-circular) word a *linear* word when we need to specify.

- A (*multi-component*) *assembly word* over an alphabet A is a set $W = \{w_0, [w_1], [w_2], \dots, [w_n]\}$, for a positive integer n consisting of one linear word w_0 and n circular words $[w_1], [w_2], \dots, [w_n]$ over A, such that each letter of A appears exactly twice (or 0 times) in W. We say that two (multi-component) assembly words are equal if they are equal as sets. Each word in W is a *component* of W, and $n + 1$ is the *number of components* of W.

Next we define various properties of assembly words.

- If a double occurrence word w can be written as a product $w = uv$ of two non-empty double occurrence words u, v, then w is called *reducible*, otherwise it is called *irreducible*. Similarly, if $\Gamma = \Gamma_1 \circ \Gamma_2$ for some non-trivial directed assembly graphs Γ_1 and Γ_2, then Γ is called *reducible*. Otherwise it is called *irreducible*.

- Reducibility of simple, directed assembly graphs and words are generalized as follows. An assembly graph Γ is reducible if there are two assembly graphs Γ_i, $i = 1, 2$, with at least one 4-valent vertices, such that $\Gamma = \Gamma_1 \cup \Gamma_2$ when the graphs are regarded as topological spaces (1-dimensional complexes), and $\Gamma_1 \cap \Gamma_2 = \{v_0\}$, a single point. In the last expression of the intersection, Γ_1 and Γ_2 are regarded as subsets of Γ, and $v_0 \in \Gamma$ is a point on an edge of Γ, as well aa an end point of Γ_1 and Γ_2.

  If Γ is not reducible, it is called *irreducible*. A (multiple-component) assembly word is *reducible* and *irreducible*, respectively, if the corresponding assembly graph has the corresponding property.

- An assembly word is a palindrome if it is equal to its reverse written in ascending order. The graph Γ_w is said to be palindromic if w is a palindrome.

- An assembly word is *strongly-irreducible* if it does not contain a proper subword that is an assembly word. An assembly graph that corresponds to a strongly-irreducible word is said to be strongly-irreducible.

4 MIC sequences

5 Assembly numbers and polynomials

In [4], the assembly number is defined and studied for assembly graphs. This subsection is a summary of definitions and results from [4].

Definitions.

- Let Γ be an assembly graph. An *open path* in Γ is a homeomorphic image of the open interval $(0, 1)$ in Γ.

  An open path is also represented by a sequence:

$$(e_1 \setminus v_0), v_1, e_2, v_2, e_3, \ldots, v_{m-1}, e_m, v_m, (e_{m+1} \setminus v_{m+1}),$$

  where v_i's are vertices in Γ for $i \in \{1, 2, \ldots, m\}$ such that $v_i \neq v_j$ when $i \neq j$, e_i's are edges in Γ for $i \in \{1, 2, \ldots, m\}$ with end points v_{i-1} and v_i respectively, such that the initial vertex of e_1 (and possibly part of e_1) and the terminal vertex of e_{m+1} (and possibly part of e_{m+1}) are not included. We say that the open path is a *cycle* if $e_1 = e_{m+1}$

- Two open paths are *disjoint* if they do not have a vertex in common.

- A set of pairwise disjoint open paths in Γ $\{\gamma_1, \ldots, \gamma_k\}$ is called *Hamiltonian* if their union contains all 4-valent vertices of Γ. An open path γ is called *Hamiltonian* if the set $\{\gamma\}$ is Hamiltonian.

- A *polygonal* path is an open path γ: $(e_1 \setminus v_0), v_1, e_2, \ldots, v_{m-1}, e_m, v_m, (e_{m+1} \setminus v_{m+1})$, such that e_i and e_{i+1} are neighbors for every $i \in \{1, 2, \ldots, m\}$.

  (Intuitively, the polygonal paths indicate the way a DNA rearrangement can occur, and indicate the types of smoothings that correspond to these rearrangements.)

- Let Γ be an assembly graph. The *assembly number* of Γ, denoted by $\mathrm{An}(\Gamma)$, is defined by $\mathrm{An}(\Gamma) = \min\{\ k \mid$ there exists a Hamiltonian set of polygonal paths $\{\gamma_1, \ldots, \gamma_k\}$ in $\Gamma\}$.

- Motivated by realizable words discussed in [47], a simple assembly graph Γ is called *realizable* if $\mathrm{An}(\Gamma) = 1$, otherwise it is called *unrealizable.*

- For a positive integer n, we define *minimal realization number for* n to be $R_{\min}(n) = \min\{|\Gamma| : \mathrm{An}(\Gamma) = n\}$, where $|\Gamma|$ is the number of 4-valent vertices in Γ. A graph Γ such that $R_{\min}(n) = |\Gamma|$ is called *a realization of* $R_{\min}(n)$.

- Each smoothing s of all vertices in an assembly graph Γ gives rise to a set of circles (components) obtained from Γ by performing s. Let $\mu(s)$ denote the number of resulting components (representing the number of molecules after the recombination). The *assembly polynomial* of a given assembly graph Γ is the polynomial $S_\Gamma(p, t) = \sum_s p^{\mathrm{ps}(s)} t^{\mu(s)-1}$ where $\mathrm{ps}(s)$ denotes the number of p-smoothings in s.

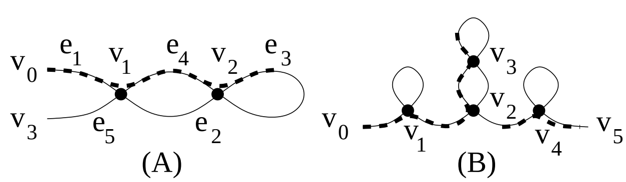

Figure 3: Polygonal Hamiltonian paths

The assembly graph depicted in Fig. 1 (A) has a Hamiltonian path represented by $(e_1 \setminus v_0), v_1, e_2, v_2, (e_3 \setminus v_2)$ that is not polygonal. In Fig. 3, polygonal Hamiltonian paths are depicted by thick dotted curves. Note that for the graph Γ depicted in (B), there is no single component polygonal Hamiltonian path, and therefore $\mathrm{An}(\Gamma) = 2$, and Γ is unrealizable.

Properties.

- For each pair of directed simple assembly graphs Γ_1 and Γ_2, one of the following equalities hold: $\mathrm{An}(\Gamma_1 \circ \Gamma_2) = \mathrm{An}(\Gamma_1) + \mathrm{An}(\Gamma_2)$, or $\mathrm{An}(\Gamma_1 \circ \Gamma_2) = \mathrm{An}(\Gamma_1) + \mathrm{An}(\Gamma_2) - 1$.
- For any positive integer n, there exists
  - (i) a reducible assembly graph Γ such that $\mathrm{An}(\Gamma) = n$,
  - (ii) an irreducible assembly graph Γ such that $\mathrm{An}(\Gamma) = n$, and
  - (iii) an assembly graph Γ with no endpoints such that $\mathrm{An}(\Gamma) = n$.

- The following properties hold for $R_{\min}$.
  - (i) For every positive integer n, $R_{\min}(n) < R_{\min}(n + 1)$.

(ii) If $R_{\min}(n) = k$ for some n and k, then for every $s \geq k$ there is an assembly graph Γ with s 4-valent vertices such that $\mathrm{An}(\Gamma) = k$.

(iii) $R_{\min}(n) \leq 3(n-1) + 1$ for every positive integer n.

Through case by case inspection, we find that $\mathrm{An}(\Gamma) = 1$ for all assembly graphs Γ with $|\Gamma| \leq 3$, and the graph Γ in Fig. 3 (B) has $\mathrm{An}(\Gamma) = 2$ and $|\Gamma| = 4$, so we have $R_{\min}(1) = 1$ and $R_{\min}(2) = 4$.

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.