

---

## Modeling of Chromosome Intermingling by Partially Overlapping Uniform Random Polygons

T. Blackstone<sup>‡</sup> · R. Scharein<sup>†</sup> · B. Borgo<sup>†</sup> ·  
R. Varela<sup>‡</sup> · Y. Diao<sup>\*</sup> · J. Arsuaga<sup>†</sup>

Received: date / Accepted: date

**Abstract** During the early phase of the cell cycle the eukaryotic genome is organized into chromosome territories. The geometry of the interface between any two chromosomes remains a matter of debate and may have important functional consequences. The Interchromosomal Network model (introduced by Branco and Pombo) proposes that territories intermingle along their periphery. In order to partially quantify this concept we here investigate the probability that two chromosomes form an unsplittable link. We use the uniform random polygon as a crude model for chromosome territories and we model the interchromosomal network as the common spatial region of two overlapping uniform random polygons. This simple model allows us to derive some rigorous mathematical results as well as to perform computer simulations easily. We find that the probability that one uniform random polygon of length  $n$  that partially overlaps a fixed polygon is bounded below by  $1 - O(\frac{1}{\sqrt{n}})$ . We use numerical simulations to estimate the dependence of the linking probability of two uniform random polygons on the amount of overlapping. We propose that this dependence relation may be modeled as  $f(\epsilon, m, n) = 1 - \frac{a(\epsilon)}{b(\epsilon)\sqrt{mn} + c(\epsilon)}$ . Numerical evidence shows that this model works well when  $\epsilon$  is relatively large ( $\epsilon \geq .5$ ). We then use these results to model the data published by Branco and Pombo and observe that for the amount of overlapping observed experimentally the URPs have a non-zero probability of forming an unsplittable link.

**Mathematics Subject Classification (2000)** MSC 57M25 · MSC 92B99

---

<sup>‡</sup> Department of Computer Science  
San Francisco State University  
1600 Holloway Ave.  
San Francisco, CA 94132

<sup>†</sup> Department of Mathematics  
San Francisco State University  
1600 Holloway Ave  
San Francisco, CA 94132

<sup>\*</sup> Department of Mathematics and Statistics  
University of North Carolina Charlotte  
9201 University City Blvd  
Charlotte, NC 28223

## 1 Introduction

During the  $G_0/G_1$  phase of the cell cycle the eukaryotic genome is organized into distinct sub-nuclear regions called chromosome territories (reviewed in [10]). DNA organization within the territories is believed to be closely related to basic cellular functions such as gene expression and repair [5,31]. The organization of the DNA at the interface between any two chromosomes is particularly important since it has been associated with the formation of chromosome aberrations [7], gene expression [36] and diffusion of large macromolecular complexes [10]. Two models have been proposed to describe the geometry of the interface between two chromosomes: the Interchromatin Compartment model [4] and the Interchromosomal Network model. In this paper, we investigate the Interchromosomal Network Model introduced by Branco and Pombo in which chromosome territories significantly intermingle along their periphery [7]. One of the implications of this model is that non-trivial linking may occur.

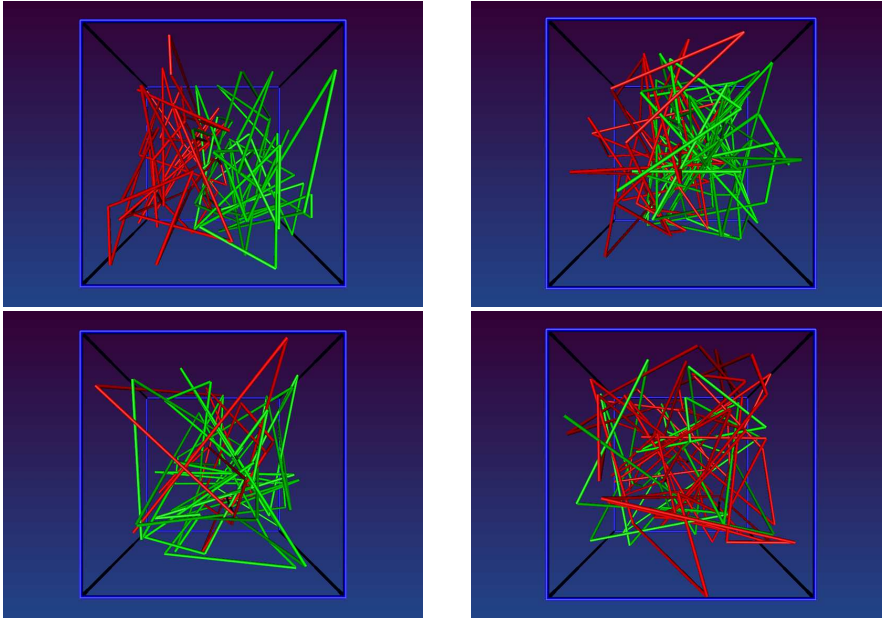
Linking is in fact a common phenomenon in biology. Linking has been observed both in proteins [6,9,16,43] and nucleic acids. In particular DNA links are commonly observed as products of *in-vitro* site-specific recombination reactions (e.g. [42]), as replication products in *Escherichia coli* (e.g. [21,45]) and as patterns of organization of mitochondrial DNA of trypanosomes (reviewed in [27]). Because of these and other observations from polymer science (e.g. [15,26]) the problem of linking of polymer chains has been theoretically addressed. These works have estimated the linking probability for fully as well as partially overlapping chains in free space [25,41]. In this later case it has been proposed that the unlinking probability increases with the distance between their centers of masses as  $1 - \exp(-r^u)$ , where  $r$  is the distance between the two centers of masses, or as a linear combination of exponentials [13,22,?,?]. Chromosomes however are found under strict conditions of confinement and paradoxically little is known about the entanglement of DNA or polymer chains in confined volumes [2,38,39]. In [2], the uniform random polygon model was used to study linking of polymers in confined volumes and it was rigorously shown that the linking probability between a chain of fixed length and a random chain of length  $n$  increases to 1 at least as fast as  $1 - O(1/\sqrt{n})$ . This study also provided numerical evidence that the linking probability of two random chains of lengths  $n$  and  $m$  increases to 1 at a rate at least at the order  $1 - O(1/\sqrt{nm})$ .

In order to provide an initial estimation of the linking probability of intermingling chromosomes we here extend the results presented in [2] by studying uniform random polygons (URP) with various degrees of overlapping. First we introduce the concept of partially overlapping uniform random polygons and rigorously show that the linking probability between any fixed chain and a random chain of length  $n$  approaches one at a rate  $1 - O(1/\sqrt{n})$  as  $n$  approaches infinity. Second we perform numerical studies to estimate the coefficients of this asymptotic formula and propose that the linking probability of two chains may be modeled by the fitting function  $1 - \frac{a(\epsilon)}{b(\epsilon)\sqrt{mn}+c(\epsilon)}$ .

## 2 Two Uniform Random Polygons in Confined Spaces with Overlapping

For the sake of simplicity, throughout this paper, we will use the unit cube  $[0, 1]^3$  as the common confining space for the two random polygons. Let  $\epsilon \in [0, 1]$  be a prefixed constant. For  $i = 1, 2, \dots, n$ , let  $U_i = (u_{i1}, u_{i2}, u_{i3})$  be a three-dimensional random point that is uniformly distributed in the space  $C_\epsilon^1 = [0, (1 + \epsilon)/2] \times [0, 1] \times [0, 1]$

such that  $U_1, U_2, \dots, U_n$  are independent. Let  $e_i$  (called the  $i$ -th edge) be the line segment joining  $U_i$  and  $U_{i+1}$ , then the edges  $e_1, e_2, \dots, e_n$  define a uniform random polygon  $R_n$  in the confined space  $C_\varepsilon^1$ , where  $e_n$  is the line segment joining  $U_n$  and  $U_1$ . Similarly, let  $V_i = (v_{i1}, v_{i2}, v_{i3})$  be a three-dimensional random point that is uniformly distributed in the space  $C_\varepsilon^2 = [(1-\varepsilon)/2, 1] \times [0, 1] \times [0, 1]$  such that  $V_1, V_2, \dots, V_n$  are independent. Let  $e'_i$  be the line segment joining  $V_i$  and  $V_{i+1}$ , then the edges  $e'_1, e'_2, \dots, e'_n$  define a uniform random polygon  $R'_n$  in the confined space  $C_\varepsilon^2$ . Notice that  $C_\varepsilon^1 \cap C_\varepsilon^2 = [(1-\varepsilon)/2, (1+\varepsilon)/2] \times [0, 1] \times [0, 1]$ . That is, the volume of the overlapping space for  $R_n$  and  $R'_n$  is exactly  $\varepsilon$ . Figure 1 shows a few examples of randomly generated URPs with various overlapping parameter  $\varepsilon$  and polygon lengths. Notice that the volumes around the polygons are included only for illustrative purposes and are not considered in this study.



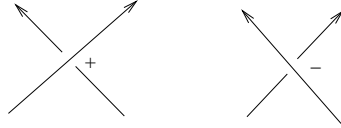
**Fig. 1** Examples of overlapping URPs. The top left panel shows two URPs of length 50 with an overlapping parameter of  $\varepsilon = 0.2$ . The top right panel shows two URPs of lengths 70 (green) and 30 (red) with an overlapping parameter of  $\varepsilon = 0.5$ . The Bottom left panel shows two URPs of lengths 40 (green) and 20 (red) with an overlapping parameter of  $\varepsilon = 0.75$ . The bottom right panel shows two URPs of lengths 40 (green) and 80 (red) with an overlapping parameter of  $\varepsilon = 1$ .

### 3 The mean squared linking number of two equal length uniform random polygons in a confined space

The approach we use in this section is quite similar to that of our previous paper on the linking of URPs [2], however for the sake of readability, we decide to include the details. Use the  $xy$ -plane as the projection plane, we will need to study the projection diagrams of  $R_n$  and  $R'_n$  under this projection. Notice that under this projection,  $C_\varepsilon^1$  becomes  $[0, (1+\varepsilon)/2] \times [0, 1]$  and  $C_\varepsilon^2$  becomes  $[(1-\varepsilon)/2, 1] \times [0, 1]$ . With probability 1, this

projection is a regular projection, namely there are no crossing points of multiplicity more than 2 in the projection.

Let us consider the case when there are only two (independent) oriented random edges  $\ell \subset C_\varepsilon^1$  and  $\ell' \subset C_\varepsilon^2$ . Since the end points of the edges are independent and are uniformly distributed in  $C_\varepsilon^1$  and  $C_\varepsilon^2$ , the probability that the projections of  $\ell$  and  $\ell'$  intersect each other is a positive number, which we will call  $2p$ . Assume that  $\ell$  and  $\ell'$  are oriented. Define a random variable  $\delta$  in the following way:  $\delta = 0$  if the projection of  $\ell_1$  and  $\ell_2$  have no intersection,  $\delta = -1$  if the projection of  $\ell_1$  and  $\ell_2$  has a negative intersection and  $\delta = 1$  if the projection of  $\ell_1$  and  $\ell_2$  has a positive intersection. For the definition of positive and negative intersections, see Figure 2.



**Fig. 2** Assignment of  $\pm 1$  at a crossing. The sign for any two edges whose relative position is as shown in the figure is determined by the right hand rule.

By symmetry, we then see that  $P(\delta = 1) = P(\delta = -1) = p$ . It follows that  $E(\delta) = 0$  and  $Var(\delta) = E(\delta^2) = 2p$ .

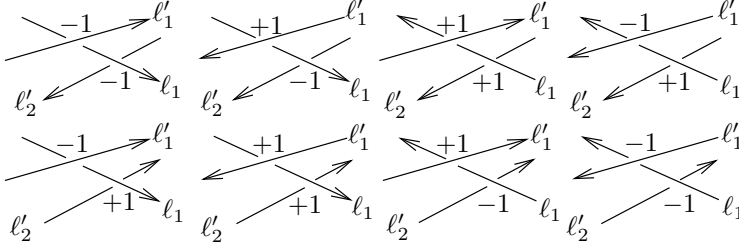
The linking number of two random polygons will be used as our primary tool to detect the non-trivial linking between the two polygons. Therefore, we need to consider the mean squared linking number of two uniform random polygons  $R_n$  and  $R'_n$  (confined in  $C_\varepsilon^1$  and  $C_\varepsilon^2$  separately). We will need the following lemma, which concerns the case when there are four edges (some of them may be identical or have a common end point) involved:  $\ell_1, \ell_2 \subset C_\varepsilon^1$  and  $\ell'_1, \ell'_2 \subset C_\varepsilon^2$ . Let  $\delta_1$  be the random number  $\delta$  defined above between  $\ell_1$  and  $\ell'_1$ ,  $\delta_2$  be the random number defined between  $\ell_2$  and  $\ell'_2$ .

**Lemma 1** Let  $\ell_1, \ell_2, \ell'_1$  and  $\ell'_2$  be four random edges such that the end points of  $\ell_1$  and  $\ell_2$  are uniformly distributed in  $C_\varepsilon^1$  and the end points of  $\ell'_1$  and  $\ell'_2$  are uniformly distributed in  $C_\varepsilon^2$ . Let  $\delta_1$  be the random number  $\delta$  defined above between  $\ell_1$  and  $\ell'_1$  and let  $\delta_2$  be the random number defined between  $\ell_2$  and  $\ell'_2$ .

(1) If the end points of  $\ell_1, \ell_2, \ell'_1$  and  $\ell'_2$  are distinct, then  $E(\delta_1\delta_2) = 0$  (this is the case when there are eight independent random points involved); (2) If  $\ell_1 = \ell_2$ , and the end points of  $\ell'_1$  and  $\ell'_2$  are distinct (this reduces the case to where there are only three random edges, with six independent points, involved), then  $E(\delta_1\delta_2) = 0$ ; (3) In the case that  $\ell_1 = \ell_2$  and  $\ell'_1$  and  $\ell'_2$  share a common point (so there are only five independent random points involved in this case), let  $u = E(\delta_1\delta_2)$  and in the case that  $\ell_1$  and  $\ell_2$  share a common point,  $\ell'_1$  and  $\ell'_2$  also share a common point (so there are four edges defined by six independent random points involved in this case), let  $E(\delta_1\delta_2) = v$ . We have  $q = p + 2(u + v) > 0$ , where  $p$  is as defined before.

*Proof* (1) This is obvious since  $\delta_1$  and  $\delta_2$  are independent random variables in this case; (2) For each configuration in which the projections of  $\ell'_1$  and  $\ell'_2$  both intersect the projection of  $\ell_1$  (since otherwise  $\delta_1\delta_2 = 0$ ), there are eight different ways to assign the orientations to the edges (see Figure 3. Four of them yield  $\delta_1\delta_2 = -1$  and four of them lead to  $\delta_1\delta_2 = 1$ . Since the joint density function of the vertices involved is simply  $\frac{1}{V^6}$

where  $V$  is the volume of the confined space  $C^3$ , thus by a symmetry argument, we have  $E(\delta_1\delta_2) = 0$ ;



**Fig. 3** For each configuration of  $\ell_1, \ell'_1$  and  $\ell_2$  for which  $\delta_1\delta_2 \neq 0$ , there are eight symmetric ways of assigning the order of the vertices to them (i.e., the orientations of them). Four of them yield  $\delta_1\delta_2 = -1$  and four of them lead to  $\delta_1\delta_2 = 1$ .

(3) Consider the case when there are two random triangles involved (one each in  $\ell'_1$  and  $\ell'_2$ ). Orient them and name the edges of the first triangle (in  $\ell'_1$ )  $\ell_1, \ell_2$  and  $\ell_3$ . Similarly, name the edges of the second triangle  $\ell'_1, \ell'_2$  and  $\ell'_3$ . Let  $\delta_{ij}$  be the crossing sign number  $\delta$  between the edges  $\ell_i$  and  $\ell'_j$ . Consider the variance of the summation  $\sum_{i,j=1}^3 \delta_{ij}$  (the summation indices are taken using mod(3)):

$$\begin{aligned} V\left(\sum_{i,j=1}^3 \delta_{ij}\right) &= E\left(\left(\sum_{i,j=1}^3 \delta_{ij}\right)^2\right) \\ &= \sum_{i,j=1}^3 E(\varepsilon_{ij}^2) + 2\sum_{i=1}^3 \left(E(\delta_{ij}\delta_{i(j-1)}) + E(\delta_{ij}\delta_{i(j+1)})\right) \\ &\quad + 2\sum_{i=1}^3 \left(E(\delta_{ij}\delta_{i+1,j+1}) + E(\delta_{ij}\delta_{i-1,j+1})\right). \end{aligned}$$

Since the  $\delta_{ij}$ 's are identical random variables, i.e., they have the same distributions, each term in the first summation of the right hand side in the above yields  $2p$ , each term in the second summation yields  $u$  (see Lemma 1) and each term in the third summation yields  $v$ . There are 9 terms in the first summation and 18 terms in the second and third summation. This leads to  $V(\sum_{i,j=1}^3 \delta_{ij}) = 18p + 36(u + v) = 18(p + 2(u + v))$ . Since  $V(\sum_{i,j=1}^3 \delta_{ij}) > 0$ , this implies that  $p + 2(u + v) > 0$ , as claimed.  $\square$

We will now pass to the case of two uniform random polygons  $R_1 \subset C_\varepsilon^1$  and  $R_2 \subset C_\varepsilon^2$ , each of  $n$  vertices. Name the edges of  $R_1$  and  $R_2$  by  $\ell_1, \ell_2, \dots, \ell_n$ , and  $\ell'_1, \ell'_2, \dots, \ell'_n$  such that this order matches the order inherited from the orientations of the polygons. As in the proof of Lemma 1, let  $\delta_{ij}$  be the crossing sign number between  $\ell_i$  and  $\ell'_j$ . Then the linking number between  $R_1$  and  $R_2$  is defined as  $\frac{1}{2} \sum_{i,j=1}^n \delta_{ij}$ . The linking number between two polygons is a well known topological invariant. That is, the linking number computed from any regular projection (in which only double crossings occur) of two polygons topologically equivalent to the original ones stays the same. Applying Lemma 1 to this case, we obtain the following theorem.

**Theorem 1** *The mean squared linking number between two uniform random polygons  $R_1$  and  $R_2$  of  $n$  edges each (in the confined space  $C^3$ ) is  $\frac{1}{2}n^2q$  where  $q = p+2(u+v) > 0$  is as defined in Lemma 1.*

*Proof* Name the edges of  $R_1$  and  $R_2$  by  $\ell_1, \ell_2, \dots, \ell_n$ , and  $\ell'_1, \ell'_2, \dots, \ell'_n$  such that this order matches the order inherited from the orientations of the polygons. Since the linking number between  $R_1$  and  $R_2$  is  $\frac{1}{2} \sum_{i,j=1}^n \delta_{ij}$ , it is easy to verify that

$$E \left( \left( \frac{1}{2} \sum_{i,j=1}^n \delta_{ij} \right)^2 \right) = \frac{1}{2} n^2 q.$$

□

Table 1 summarizes the estimated  $q$  values for different  $\epsilon$ .

**Table 1** The estimated  $q$  values obtained using  $n = 150$  and sample sizes of 50000.  $e$  is the error bar at the 95% confidence level.

$\epsilon$	.55	.6	.65	.7	.75	.8	.85	.9	.95	1
$q$	.0005	.0039	.0091	.0141	.0190	.0230	.0266	.0283	.0299	.0303
$e$	< .0001	< .0001	.0001	.0002	.0002	.0003	.0003	.0004	.0004	.0004

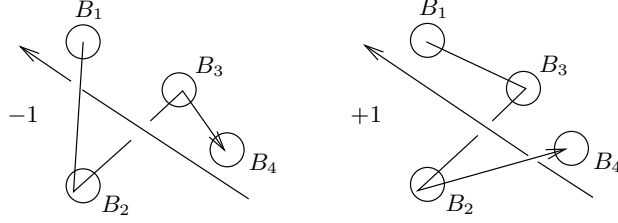
#### 4 The linking probability of a uniform random polygon with a fixed simple closed curve with overlapped volume in a confined space

While it is intuitively obvious that the linking probability of two URPs ( $C_\epsilon^1$  and  $C_\epsilon^2$  respectively with an  $\epsilon > 0$  overlapping) should go to 1 as their lengths tend to infinity, proving it rigorously turned out to be quite difficult. In fact, there is no known rigorous proof for this asymptotic linking probability behavior for any other model either. Here, we consider a similar but simpler problem where one of the two components is a non-degenerate fixed polygon in  $C_\epsilon^1$  with part of it in the overlapped volume  $C_\epsilon^1 \cap C_\epsilon^2$ . A result similar to this one was given in [2] with the proof given for a special case of the closed curve. Here we will provide a different proof under the general setting for the fixed closed curve.

**Theorem 2** *Let  $S$  be a fixed non-degenerate polygon in  $C_\epsilon^1$  such that part of it is within the overlapped volume  $C_\epsilon^1 \cap C_\epsilon^2$ , and let  $R_n$  be a uniform random polygon of  $n$  vertices confined in  $C_\epsilon^2$ , then as  $n$  approaches infinity, the probability that  $S$  and  $R_n$  form an unsplittable link approaches one. More specifically, the probability that the linking number between  $S$  and  $R_n$  does not equal to one is bounded below by  $1 - O(\frac{1}{\sqrt{n}})$ .*

*Proof* Since  $S$  is non-degenerate and has a part in  $C_\epsilon^1 \cap C_\epsilon^2$ , it is a well-known result that  $S$  can be projected to some plane such that no more than two strands can be projected to the same point in the projection. It follows that there exists a small segment  $\ell$  on  $S$  such that: (1) the ball  $B$  with  $\ell$  as a diameter is completely within  $C_\epsilon^1 \cap C_\epsilon^2$ ; (2) the circle  $C$  in the projection plane with the projection of  $\ell$  as a diameter does not intersect any other edges of  $S$  (other than the edge containing  $\ell$ ).

Assign  $S$  an arbitrary orientation. Choose four small balls  $B_1, B_2, B_3$  and  $B_4$  in  $B$  with the following properties: (1) they are all within  $B$ ; if  $Y_i \in B_i$  for  $i = 1$  to 4 and  $Y_1Y_2Y_3Y_4$  forms three consecutive steps in  $R_n$ , then it will contribute a  $-1$  to the linking number between  $R_n$  and  $S$ , if  $Y_i \in B_i$  for  $i = 1, 4, Y_2 \in B_3, Y_3 \in B_2$  and  $Y_1Y_2Y_3Y_4$  forms three consecutive steps in  $R_n$ , then it will contribute a  $+1$  to the linking number between  $R_n$  and  $S$ . The existence of these balls is rather obvious since their radius can be arbitrarily small. Figure 4 is an illustration of these balls and the corresponding paths  $Y_1Y_2Y_3Y_4$ .



**Fig. 4** The  $Z$ -pattern in which there is an equal chance of obtaining  $-1$  or  $+1$ .

Without loss of generality, let us assume that  $n = 4k$  for some integer  $k$  (this is purely for the sake of convenience in order to avoid using the integer value round up or down functions). We will name the vertices of  $R_n$  as (following the order they appear in  $R_n$ ):  $X_{11}, X_{12}, X_{13}, X_{14}, X_{21}, X_{22}, X_{23}, X_{24}, \dots, X_{k1}, X_{k2}, X_{k3}, X_{k4}$ . Let  $A_i$  ( $1 \leq i \leq k$ ) be the event:  $X_{i1} \in B_1, X_{i4} \in B_4, X_{i2}, X_{i3} \in B_2 \cup B_3$  but each ball contains only one of them. It is obvious that the events  $A_i$ 's are independent of each other and  $P(A_i) \geq \alpha$  for some (small) positive constant (that depends only on the balls  $B_1$  to  $B_4$ ). We will say that the four vertices  $X_{i1}, X_{i2}, X_{i3}, X_{i4}$  form a  $Z$ -pattern if the event  $A_i$  happens. By symmetry, we see that  $P(a = 1|A_i) = P(a = -1|A_i) = 1/2$ , where  $a$  is the linking number between  $X_{i1}X_{i2}X_{i3}X_{i4}$  and  $\ell$ .

It is an easy exercise in probability to show that there exist positive constants  $\beta$  and  $\gamma$  such that with probability at least  $1 - e^{-\gamma n}$ , a randomly generated  $R_n$  contains at least  $\beta n$   $Z$ -patterns.

For any  $t \geq \beta n$ , let  $(j_1, j_2, \dots, j_t)$  be a  $t$ -tube such that  $1 \leq j_1 < j_2 < \dots < j_t \leq k$  and let  $\mathcal{A}$  be the set of all such  $t$ -tubes (with  $t \geq \beta n$ ) and for any  $(j_1, j_2, \dots, j_t) \in \mathcal{A}$ , define

$$T(j_1, j_2, \dots, j_t) = \left( \bigcap_{i \in \{j_1, j_2, \dots, j_t\}} A_i \right) \cap \left( \bigcap_{m \notin \{j_1, j_2, \dots, j_t\}} \overline{A_m} \right).$$

By this definition, the event that  $R_n$  has at least  $\beta n$   $Z$ -patterns is simply the disjoint union of the  $T(j_1, j_2, \dots, j_t)$ 's with  $(j_1, j_2, \dots, j_t)$  running over  $\mathcal{A}$ . We then have ( $Lk$  is the linking number between  $R_n$  and  $S$ ):

$$\begin{aligned} & P(Lk \neq 0) \\ & \geq P(Lk \neq 0 \text{ and } R_n \text{ has at least } \beta n \text{ } Z \text{- patterns}) \\ & = \sum_{(j_1, j_2, \dots, j_t) \in \mathcal{A}} P(Lk \neq 0 \cap T(j_1, j_2, \dots, j_t)) \\ & = \sum_{(j_1, j_2, \dots, j_t) \in \mathcal{A}} P(Lk \neq 0 | T(j_1, j_2, \dots, j_t)) P(T(j_1, j_2, \dots, j_t)). \end{aligned} \quad (1)$$

Now, for each  $T(j_1, j_2, \dots, j_t)$ , consider  $P(Lk \neq 0 | T(j_1, j_2, \dots, j_t))$  in two steps. First take any configuration of  $R_n$  not involved in  $A_i$  with  $i \in \{j_1, j_2, \dots, j_t\}$  and compute the total linking number contribution of these segments with  $S$ . Let this number be  $Lk'$ . Then  $P(Lk \neq 0 | T(j_1, j_2, \dots, j_t))$  becomes  $P(Lk \neq -Lk' | \cap_{1 \leq i \leq t} A_{j_i})$ . However, under the condition  $\cap_{1 \leq i \leq t} A_{j_i}$ , each  $Z$ -pattern involved has a probability of  $1/2$  of making a  $+1$  or  $-1$  contribution in the linking number and these events are independent of each other. Thus  $P(Lk = -Lk' | \cap_{1 \leq i \leq t} A_{j_i})$  is equivalent to  $P(x = b)$  for some  $b$ , where  $x$  is the binomial random variable with parameters  $1/2$  and  $t$ . It follows that

$$\begin{aligned} P(Lk = -Lk' | \cap_{1 \leq i \leq t} A_{j_i}) &= P(x = b) \\ &= \binom{t}{b} \frac{1}{2^t} \leq \binom{t}{t/2} \frac{1}{2^t} < \frac{2}{\sqrt{t}} = O\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

Combining this with (1), we get

$$\begin{aligned} P(Lk \neq 0) &\geq (1 - O(\frac{1}{\sqrt{n}})) \sum_{(j_1, j_2, \dots, j_t) \in \mathcal{A}} P(T(j_1, j_2, \dots, j_t)) \\ &\geq (1 - O(\frac{1}{\sqrt{n}}))(1 - e^{-\gamma n}) = 1 - O(\frac{1}{\sqrt{n}}). \end{aligned}$$

Our numerical results in the next section strongly suggest that the linking probability between two uniform random polygons  $R_1 \subset C_\varepsilon^1$  and  $R_2 \subset C_\varepsilon^2$  ( $\varepsilon > 0$ ) with  $n$  and  $m$  vertices respectively is bounded below by  $1 - O(\frac{\varepsilon}{\sqrt{mn}})$ . However, we are unable to prove this rigorously at this time.

## 5 Numerical methods and results

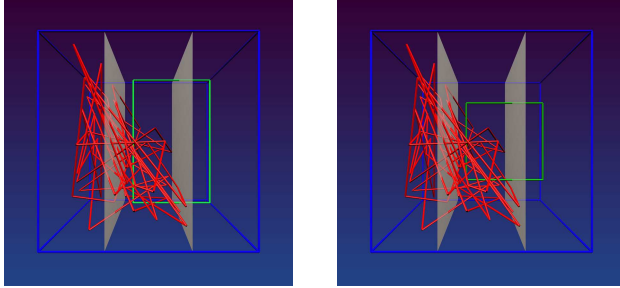
The generation of uniform random polygons is straightforward and has been explained in previous studies [2, 3, 33]. For partially overlapping polygons, we select the  $x$ -coordinate of each vertex uniformly from  $[0, (1+\varepsilon)/2]$  for one polygon and from  $[(1-\varepsilon)/2, 1]$  for the other polygon. The  $y$  and  $z$  coordinates of the vertices (for both polygons) are selected from  $[0, 1]$  uniformly. As in the previous studies we estimated the linking probability of two polygons by computing their linking number [2]. If the linking number between two polygons is not zero, then the two polygons are non-trivially linked. In this case the two polygons are said to be homologically linked. There exist nontrivial links whose linking number is zero so there are polygons that are topologically linked but not homologically linked. In this study we make no effort in trying to determine topologically linked polygons as that would require the use of various knot invariant polynomials, which is beyond the scope of this paper. Instead, we focused our investigation to  $P(Lk \neq 0)$ , since  $P(Lk \neq 0)$  is a lower bound for the actual linking probability and we have a fairly good understanding of the asymptotic behavior of  $P(Lk \neq 0)$  in terms of the numbers of vertices in the two random polygons involved. The linking number is computed following the Gaussian integral formulation [24]. Sample sizes are set to achieve convergence. For most cases, it turns out that 50,000 is enough.

### 5.1 The case of one fixed polygon

Our first numerical study supports the analytical results presented above and it concerns the linking between a uniform random polygon and a fixed equatorial flat curve.



This is illustrated in Figure 5 for  $\epsilon = 0.4$  and two different equatorial curves. To support our analytical result, we calculated the probability of having a non-zero linking number between a uniform random polygon of varying length and a fixed closed rectangular curve lying flat in the equatorial plane of the cube  $E = \mathbb{R}^2 \times \{\frac{1}{2}\}$  for different overlapping volumes.



**Fig. 5** One URP and one fixed equatorial curve with  $\epsilon = 0.4$ ,  $N = 50$ . The fixed equatorial rectangle is 0.1 (left) and 0.25 (right) away from the boundary.

By Theorem 2, this linking probability should grow at a rate of at least  $1 - O(\frac{1}{\sqrt{n}})$  and it is independent of the dimensions of the fixed curve as shown in Theorem 2. For simplicity we considered rectangles of different sizes in the plane  $E$ . Some examples are given below by their vertices for the case  $\epsilon = .25$ ):

- a)  $(.1, .1, .5)$ ,  $(.9, .9, .5)$ ,  $(.9, .625, .5)$ ,  $(.1, .625, .5)$ ;
- b)  $(.2, .1, .5)$ ,  $(.8, .9, .5)$ ,  $(.8, .625, .5)$ ,  $(.2, .625, .5)$ ;
- c)  $(.3, .1, .5)$ ,  $(.7, .9, .5)$ ,  $(.7, .625, .5)$ ,  $(.3, .625, .5)$ ;
- d)  $(.4, .1, .5)$ ,  $(.6, .9, .5)$ ,  $(.6, .625, .5)$ ,  $(.4, .625, .5)$ .

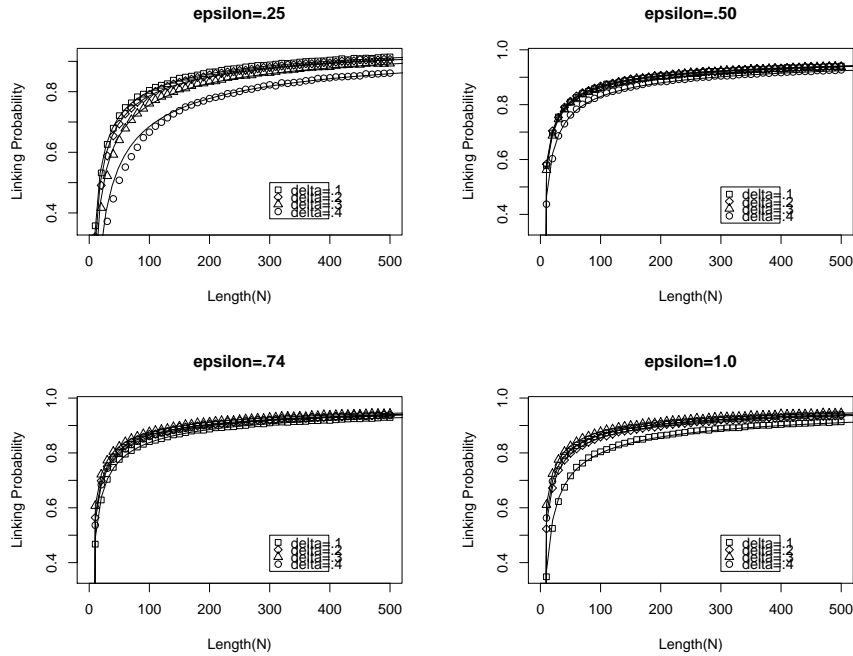
Results for different  $\epsilon$  values are shown in Figure 6. Each curve is fitted to a function of the form  $1 - \frac{a}{\sqrt{n}}$ , where the best-fit value for  $a$  is shown in table 2.

**Table 2** Best-fit values for the parameter  $a$  by non-linear least squares.

$\delta$	.1	.2	.3	.4	$\epsilon$
	$1.991 \pm .001587$	$2.138 \pm .002483$	$2.437 \pm .005167$	$3.148 \pm .037930$	.25
	$1.348 \pm .000109$	$1.323 \pm .000077$	$1.380 \pm .000123$	$1.705 \pm .001427$	.50
	$1.623 \pm .000754$	$1.356 \pm .000194$	$1.247 \pm .000067$	$1.417 \pm .000422$	.75
	$2.007 \pm .00215$	$1.451 \pm .000575$	$1.234 \pm .000074$	$1.350 \pm .0000001$	1.0

While the amount of overlapping between the polygon and the fixed closed curve, the position and size of the fixed close curve all affect the linking probability, the overall trend of the form  $1 - O(\frac{1}{\sqrt{n}})$  is strongly supported by the data.

This study provides an estimation of the values of the coefficients in  $1 - O(\frac{1}{\sqrt{n}})$  and also confirms the hypothesis that it is more likely for polygons with small overlapping volume to have a zero linking number than those with large overlapping volume. These numerical results are consistent and extend those obtained in [2].



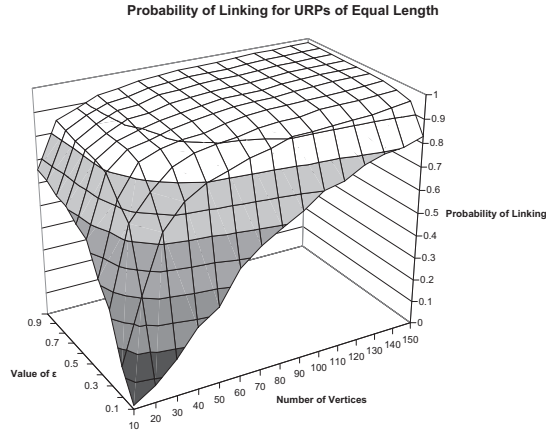
**Fig. 6** The plot of  $P(\text{Lk} \neq 0)$  for the cases  $\epsilon = 0.20, 0.50, 0.75$  and  $1$  respectively. The  $x$ -axis represents the number of vertices and the  $y$ -axis the probability  $P(\text{Lk} \neq 0)$ .

### 5.2 The case of two URP's with the same number of vertices

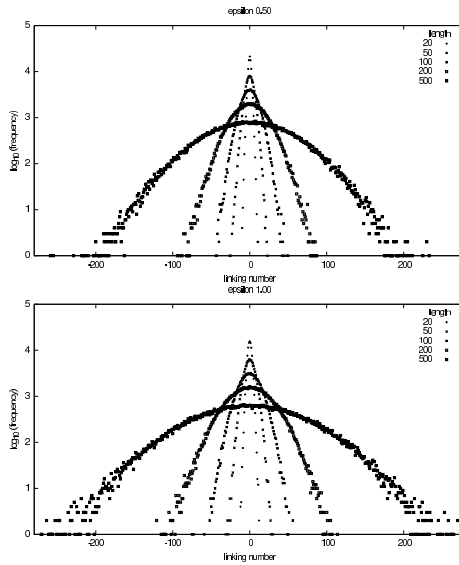
Our second numerical study concerns the linking probability and the distribution of linking numbers of two URPs with the same number of vertices. Figure 7 shows the linking probability as a function of the length of the two chains and of the overlapping value  $\epsilon$ . As suggested by our previous results the linking probability increases as  $1 - O(1/n)$  for fixed  $\epsilon$ . Furthermore, an increase in the amount of overlapping (for fixed  $n$ ) leads to an increase of the linking probability. Next we studied the distribution of linking numbers. We observed that for a fixed  $\epsilon$ , shorter polygon pairs tend to produce zero linking numbers with a higher frequencies in comparison with longer polygon pairs. We also confirmed that the distribution of linking numbers is more spread for longer polygons. A similar effect is observed when the value of  $\epsilon$  is varied. That is, the greater the amount of overlapping, the wider the distribution of linking numbers. The plot of distribution of linking numbers is shown in Figure 8 for two different values of  $\epsilon$ .

### 5.3 The case of two URP's with different number of vertices

According to our previous results for full overlapping polygons (i.e.  $\epsilon = 1$ )[2] and the results obtained in the sections above the linking probability of two random polygons with fixed overlapping should grow as  $1 - O(\frac{1}{\sqrt{mn}})$  for large values of  $m$  and  $n$ , and this growth is modulated by the amount of overlapping. In order to confirm this and



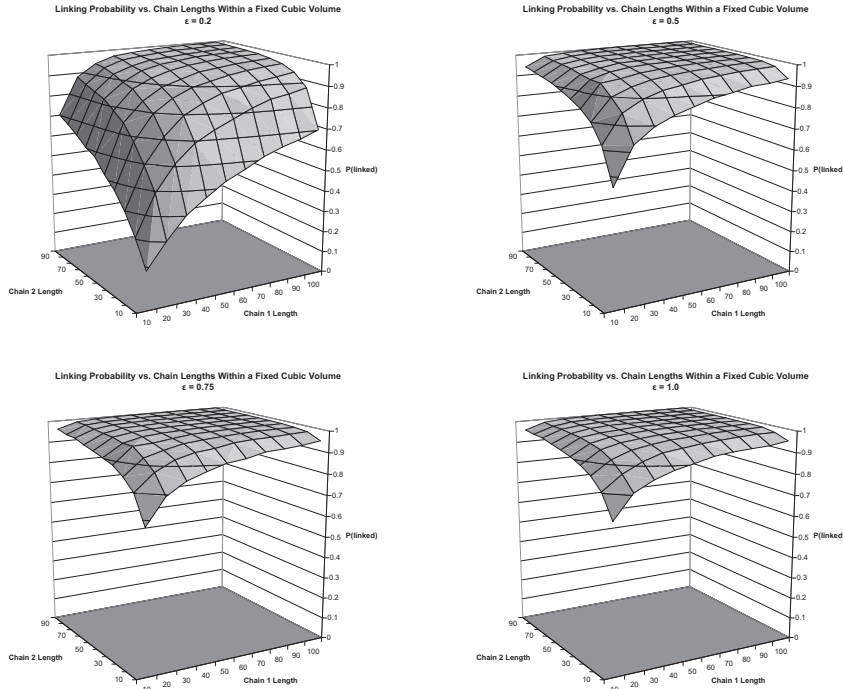
**Fig. 7** Probability of getting a non-zero linking number between two uniform random polygons as a function of  $\epsilon$ . The number  $n$  of vertices (x-axis) considered in the simulations varied from  $n = 10$  up to 150 and the values of  $\epsilon$  from .1 to .9 (y-axis).



**Fig. 8** Distribution of linking numbers between two uniform random polygons from  $n = 10$  to 360 for the values of  $\epsilon = 0.5$  and 1.

to determine the dependence on the amount of overlapping we performed numerical studies where the length of the polygons as well as the amount of overlapping changed. For a fixed  $\epsilon > 0$  our numerical results support that  $P(\text{Lk} \neq 0)$  has the general asymptotical behavior of  $1 - O(\frac{1}{\sqrt{mn}})$  as shown in Figure 9. However, the rate of convergence is clearly slower when  $\epsilon$  is small. We propose a fitting function of the form  $f(\epsilon, m, n) = 1 - \frac{a(\epsilon)}{b(\epsilon)\sqrt{mn} + c(\epsilon)}$ . Since  $P(\text{Lk} \neq 0) = 0$  when  $\epsilon = 0$ , an ideal fitting function  $f(\epsilon, m, n)$  should have the property that  $f(0, m, n) = 0$  and  $f(\epsilon, m, n) =$

$1 - O(\frac{1}{\sqrt{mn}})$  for any fixed  $\varepsilon > 0$ . This could be achieved by setting  $b(0) = 0$  and  $a(0) = c(0)$ . Ideally, we would like to see that  $b(\varepsilon) \rightarrow 0$  and  $a(\varepsilon)/c(\varepsilon) \rightarrow 1$  as  $\varepsilon \rightarrow 0$ . Several best-fit curves from non-linear least squares regression are shown in Table 3, where rows represent values of  $n$  and the columns represent values of  $m$ . As the numerical results show, this suggested fitting works pretty well for  $\varepsilon \geq .5$ , but not so well in the case that  $\varepsilon$  is small. This indicates that for small values of  $\varepsilon$ , the dependence of  $P(\text{Lk} \neq 0)$  on  $\varepsilon$  is not as simple as the fitting model we have suggested and more careful analysis will be needed in the future in order to find a better fit.



**Fig. 9** Linking probabilities for two different polygons with different number of vertices. The  $x$  and  $y$  axes represent the number of vertices for the two polygons. Panel A shows the linking probability for a value of  $\varepsilon = 0.6$ , Panel B shows the values for a value of  $\varepsilon = 0.75$ , Panel C shows values for an  $\varepsilon = 0.87$  and Panel D for an  $\varepsilon = 1.0$ .

#### 5.4 Predictions for chromosome linking

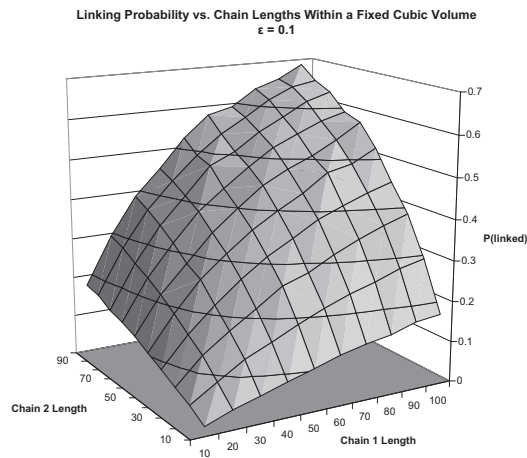
In Branco and Pombo [7] a new Fluorescence In-Situ Hybridization (FISH) technique to examine thin slides of frozen nuclei was used to measure the degree of overlapping in human lymphocytes. It was found that chromosome territories overlap and that the degree of overlapping varies between pairs of chromosomes (overlapping for three or more chromosomes was considered negligible). More specifically, it was found that overlapping varies between 0.20% and 0.01% of the volume of the nucleus which was

**Table 3** Non-linear least squares regression results for the fitting function  $f(\varepsilon, m, n) = 1 -$ 

$\frac{a(\varepsilon)}{b(\varepsilon)\sqrt{mn+c(\varepsilon)}}$										
$\varepsilon = .2: a = 2.3645, b = .3520, c = -1.2200$ and SSerr = .5719										
	10	20	30	40	50	60	70	80	90	100
10	.2633	.5927	.6968	.7505	.7842	.8077	.8252	.8388	.8498	.8589
20	.5927	.7505	.8077	.8388	.8589	.8732	.8841	.8926	.8995	.9053
30	.6968	.8077	.8498	.8732	.8886	.8995	.9079	.9145	.9199	.9244
40	.7505	.8388	.8732	.8926	.9053	.9145	.9215	.9270	.9316	.9354
50	.7842	.8589	.8886	.9053	.9164	.9244	.9305	.9354	.9394	.9427
60	.8077	.8732	.8995	.9145	.9244	.9316	.9371	.9414	.9450	.9480
70	.8252	.8841	.9079	.9215	.9305	.9371	.9421	.9461	.9494	.9521
80	.8388	.8926	.9145	.9270	.9354	.9414	.9461	.9498	.9528	.9554
90	.8498	.8995	.9199	.9316	.9394	.9450	.9494	.9528	.9557	.9581
100	.8589	.9053	.9244	.9354	.9427	.9480	.9521	.9554	.9581	.9603
$\varepsilon = .5: a = 1.1971, b = .7707, c = -5.2121$ and SSerr = .0613										
	10	20	30	40	50	60	70	80	90	100
10	.5203	.7895	.8529	.8827	.9004	.9124	.9211	.9278	.9332	.9375
20	.7895	.8827	.9124	.9278	.9375	.9443	.9493	.9533	.9564	.9591
30	.8529	.9124	.9332	.9443	.9514	.9564	.9602	.9632	.9656	.9676
40	.8827	.9278	.9443	.9533	.9591	.9632	.9663	.9688	.9708	.9725
50	.9004	.9375	.9514	.9591	.9641	.9676	.9704	.9725	.9742	.9757
60	.9124	.9443	.9564	.9632	.9676	.9708	.9732	.9752	.9767	.9780
70	.9211	.9493	.9602	.9663	.9704	.9732	.9754	.9772	.9786	.9798
80	.9278	.9533	.9632	.9688	.9725	.9752	.9772	.9788	.9801	.9812
90	.9332	.9564	.9656	.9708	.9742	.9767	.9786	.9801	.9813	.9824
100	.9375	.9591	.9676	.9725	.9757	.9780	.9798	.9812	.9824	.9833
$\varepsilon = .75: a = 1.3030, b = 1.1429, c = -7.8037$ and SSerr = .0344										
	10	20	30	40	50	60	70	80	90	100
10	.6406	.8441	.8913	.9134	.9266	.9355	.9419	.9469	.9508	.9540
20	.8441	.9134	.9355	.9469	.9540	.9590	.9627	.9656	.9680	.9699
30	.8913	.9355	.9508	.9590	.9643	.9680	.9708	.9730	.9747	.9762
40	.9134	.9469	.9590	.9656	.9699	.9730	.9753	.9771	.9786	.9798
50	.9266	.9540	.9643	.9699	.9736	.9762	.9782	.9798	.9811	.9822
60	.9355	.9590	.9680	.9730	.9762	.9786	.9803	.9817	.9829	.9839
70	.9419	.9627	.9708	.9753	.9782	.9803	.9820	.9832	.9843	.9852
80	.9469	.9656	.9730	.9771	.9798	.9817	.9832	.9844	.9854	.9862
90	.9508	.9680	.9747	.9786	.9811	.9829	.9843	.9854	.9863	.9871
100	.9540	.9699	.9762	.9798	.9822	.9839	.9852	.9862	.9871	.9878
$\varepsilon = 1.0: a = 1.7903, b = 1.5582, c = -1.1805$ and SSerr = .0316										
	10	20	30	40	50	60	70	80	90	100
10	.6686	.8490	.8935	.9147	.9274	.9360	.9423	.9472	.9510	.9542
20	.8490	.9147	.9360	.9472	.9542	.9591	.9628	.9657	.9680	.9699
30	.8935	.9360	.9510	.9591	.9643	.9680	.9708	.9729	.9747	.9762
40	.9147	.9472	.9591	.9657	.9699	.9729	.9752	.9770	.9785	.9797
50	.9274	.9542	.9643	.9699	.9736	.9762	.9782	.9797	.9810	.9821
60	.9360	.9591	.9680	.9729	.9762	.9785	.9803	.9817	.9828	.9838
70	.9423	.9628	.9708	.9752	.9782	.9803	.9819	.9832	.9842	.9851
80	.9472	.9657	.9729	.9770	.9797	.9817	.9832	.9844	.9853	.9861
90	.9510	.9680	.9747	.9785	.9810	.9828	.9842	.9853	.9862	.9870
100	.9542	.9699	.9762	.9797	.9821	.9838	.9851	.9861	.9870	.9877

observed for the chromosome territory pairs 2 and 3 and 1 and 21 respectively. In order to determine whether the observed overlapping may imply some non-trivial linking we represented the chromatin as a polygonal trajectory in space (otherwise the trajectories could always be unlinked) and measured the probability that two of these trajectories are linked. We considered the case of chromosomes 2 and 3 (those showing the highest

overlapping). Each of these chromosome territories occupies about 2% of the total volume of the nucleus and their overlapping volume is 0.20% of the total volume of the nucleus. We modeled this situation by two overlapping random polygons in which the bounding box represents the spatial constraints imposed on the chromosomes due to the nuclear membrane, nuclear organelles or other chromosomes. We estimate that the corresponding overlapping value is  $\epsilon = 0.10$ . Since there is not an exact correspondence between the length of the URPs and the statistical length of a chromosome we show results for two independent polygons of various lengths. Figure 10 shows the results. Interestingly we observed that even for small  $\epsilon$  if the length of the polygon is long enough the linking probability tends to 1 suggesting that the linking probability of two chromosomes may be non-trivial.



**Fig. 10** Linking probability between two uniform random polygons with  $\epsilon = .10$ . The number  $n$  of vertices (x-axis) considered in the simulations varied from  $n = 10$  up to 150.

## 5.5 Conclusions

The eukaryotic genome is organized in territories during the early phases of the cell cycle [10]. An important problem, due to its functional implications, is to determine the geometry and topology of the regions interfacing any two neighboring chromosomes [8, 10]. Here we have investigated if the Interchromosomal Network model proposed by Branco and Pombo [7] predicts non-trivial linking of chromosomes. In order to better quantify the linking between two chromosomes we have assumed that the simulated trajectories are closed loops. A similar result exists for open trajectories however it would require more convoluted computations without providing more information. Second we have performed this study using a very simplistic model, the uniform random polygon (URP). The main advantage of this model is that it allows developing the asymptotic theory both analytically and numerically. Nevertheless it is important for us to point out that the URP model is not intended as a realistic model for chromosomes because the flexibility of the generated chains is not defined. However, results using the URP model are qualitatively very similar to those using other polymer models when studying knotting of polymers and of DNA [1, 29, 33]. In a preliminary study we

also compared our results with those of equilateral random polygons (or freely jointed chains) in confined volumes and results were also consistent (unpublished results). In future studies we intend to implement more realistic models of chromosome territories as those presented in [11, 34, 35, 40, 44] and test how the linking probability increases with chromosome sizes and degree of overlapping. Besides the theoretical considerations discussed here one may still wonder if chromosomes may link (or of this phenomenon can be experimentally observed) since one would expect other biological factors, such as the volume exclusion of the chromatin fibers [12] or the action of type II topoisomerases, may contribute to the unlinking of chromosomes. While these are definitely factors that may hamper the formation and/or observation of linked chromosomes it is worth mentioning that our results are consistent with observations published by the Ingber lab in which it was experimentally observed that genomes extracted from living cells behave as a single unit instead of as separate entities [28].

Finally we would like to emphasize that our results can also be applied to study other situations (to be discussed elsewhere) in which long polymer chains are bound to different spatial regions. Such is the case of semicrystalline polyethylene in amorphous regions [26]. For instance in [15] it was quantified how the linking probability of random walks over the diamond lattice decreases when the thickness of the amorphous regions increases.

**Acknowledgement** This work was supported in part by NIH grant 2S06GM52588-12 (J. Arsuaga), NSF grants DMS-0712958 and DMS-0920880 (Y. Diao), NSF grant DMS-0920887 (J. Arsuaga and R. Scharein). In addition, T. Blackstone was supported by a grant of the SFSU center for computing for the Life Sciences (CCLS), B. Borgo was a Howell fellow, R. Scharein was supported by a CSUPERB grant to J. Arsuaga and R. Varela was a fellow of the Louis Stokes Alliances for Minority Participation (LSAMP) Program 2006-08.

## References

1. Arsuaga J et al 2005, *Proc Natl Acad Sci USA* **102**(26) 9165-9.
2. Arsuaga J et al 2007, *J Physics A* **40** 1925–36.
3. Arsuaga J et al 2007, *J Physics A* **40** 11697–11711.
4. Albiez H et al 2006, *Chromosome Res* **14**(7) 707-33.
5. Fraser P and Bickmore W 2007, *Nature* **447**(7143) 413-7.
6. Boutz D R et al 2007, *J Mol Biol* **368** 1332-44.
7. Branco M and Pombo A 2006, *PLoS Biol* **4** 780–8.
8. Branco M and Pombo A 2007, *Trends Cell Biol* **17**(3) 127–34.
9. Connolly M L, Kuntz I D and Crippen G M 1980, *Biopolymers* **19** 1167-82.
10. Cremer T, Küpper K, Dietzel S and Fakan S 2004, *Biol Cell* **96** 555–67.
11. Lieberman-Aiden E et al. *science* **326**(5950) 289-93.
12. Dorier J and Stasiak A 2009, *Nucleic Acids Res* **37**(19) 6316–22.
13. Deguchi T and Tsurusaki K 1997, *Lectures at Knots 96* (edited by S. Suzuki), World Scientific Publishing Co., 95–122.
14. Dobay A, Dubochet J, Millett K, Sottas P E and Stasiak A 2003, *Proc Nat Acad Sci USA* **100**(10) 5611–5.
15. Duan Z H and Howard L N 1998, *J Comput Chem* **20**(3) 348–53.
16. Duda R L 1998, *Cell* **94** 55-60.
17. Edwards S F 1967, *Proc Phys Soc* **91** 513–9.
18. Edwards S F 1968, *J Phys A* **1** 15–28.
19. Flammini A, Maritan A and Stasiak A 2004, *Biophys J* **87**(5) 2968–75.
20. Flory J P 1953, *Principles of Polymer Chemistry*, Cornell University Press.
21. Grainge I et al 2007, *EMBO J* **26** 4228–38.
22. Hirayama N, Tsurusaki K, Deguchi T *J Physics A* **42**(10) 105001.

23. Iwata K and Kimura T 1981, *J Chem Phys* **74** 2039–48.
24. Klenin K and Langowski J 2000, *Biopolymers* **54**(5) 307–17.
25. Klenin K V et al 1989, *J Biomol Struct Dyn* **6**(4) 707–14.
26. Lacher R C, Bryant J L, Howard L N and Sumners D W 1986, *Macromolecules* **19**(10) 2639–43.
27. Liu B et al 2005, *Trends Parasitol* **21** 363–9.
28. Maniotis A J, Bojanowski K, Ingber D E 1997, *J Cell Biochem* **65**(1) 114–30.
29. Mansfield M L 1994, *Macromolecules* **27** 5924–6.
30. Mathews D H and Turner D H 2006, *Curr Opin Struct Biol* **16**(3) 270–8.
31. Meaburn K J, Misteli T and Soutoglu E 2007, *Semin Cancer Biol* **17**(1) 80–90.
32. Men Y, Rieger J and Strobl G 2003, *Phys Rev Lett* **91**(9) 095502–2.
33. Millett K 2000, *Knots in Hellas'98 (Delphi)*, Series on Knots and Everything **24**, World Scientific, 306–34.
34. Mateos-Langerak J et al 2009, *Proc Natl Acad Sci USA* **106**(10) 3812–7.
35. Munkel C et al 1999, *J Mol Biol* **285**(3) 1053–65.
36. Noordermeer D et al 2008, *PLoS Genet* **4**(3) e1000016.
37. Orlandini E, Janse van Rensburg E J, Tesi M C and Whittington S G 1994, *J Physics A* **27**(2) 335–45.
38. Orlandini E and Whittington S G 2004, *J Chem Phys* **121** 12094–9.
39. Orlandini E, Tesi M C and Whittington S G 2000, *J Physics A* **33** 181–6.
40. Sachs R K et al 1995, *Proc Natl Acad Sci USA* **92**(7) 2710–4.
41. Soteros C E, Sumners D W and Whittington S G 1999, *J Knot Theory Ramifications* **8** 49–70.
42. Sumners D W, Ernst C, Spengler S J and Cozzarelli N R 1995, *Q Rev Biophys* **28**(3) 253–313.
43. Wikoff W R et al 2000, *Science* **289** 2129–33.
44. Yokota H et al 1995, *J Cell Biol* **130**(60) 1239–49.
45. Zechiedrich E L, Khodursky A B, Cozzarelli N R 1997, *Genes Dev* **11** 2580–92(19)