International Journal of Pure and Applied Mathematics
Volume 40 No. 1 2007, 103-114

HOMOLOGY SEARCHING AND INTERLOCKING OF CHROMOSOMES: ANALYTIC MODELING VERSUS COMPUTER SIMULATION

Günther Karigl
Institute of Discrete Mathematics and Geometry
Vienna University of Technology
Wiedner Hauptstraße 8-10, Vienna, A-1040, AUSTRIA
e-mail: g.karigl@tuwien.ac.at

Abstract: In order to simulate the process of homology searching and interlocking of chromosomes in meiotic pairing, a cellular automaton model is introduced. The model reflects both the spatial as well as the temporal behavior of chromosomes in the meiotic cell nucleus. Moreover, we compare the outcome of our simulations to results obtained by a stochastic model describing interlocking probabilities analytically. With the stochastic model we use our simulation results to estimate model parameters. Thus a far more general description of chromosomal behavior is achieved which relies on both simulation and analytic considerations.

AMS Subject Classification: 92C37, 37N25, 60J10
Key Words: chromosome pairing, cellular automaton model, Markov chain model, simulation model

1. Introduction

Sexual reproduction in populations enables continuous recombination of parental genetic make-up under the descendants and guarantees genetic variability. In sexually reproducing organisms, gametes of opposite sex fuse to produce organisms with two sets of homologous, i.e. nearly identical chromosomes, one of each parent. This doubling of chromosomes has to be compensated by the separation of chromosome sets prior to the production of gametes. This

Received: August 17, 2007

is achieved during meiosis. Each chromosome has to find its homologue and to pair with it along the whole length. How this process actually proceeds is largely unknown in detail (see [5], for a review). Moreover, during pairing of homologous chromosomes it may happen that foreign chromosomes will be entrapped between them. This process is called interlocking of chromosomes. It has been observed surprisingly rare and has given rise to various speculations on the course of homology searching, as well.

In Section 2 of this paper we introduce a cellular automaton model in order to simulate homology searching and interlocking of chromosomes, describing the spatial and temporal behavior of chromosomes during meiotic pairing. For a reasonable choice of the simulation parameters we orientated ourselves by experimental data from yeast and found the following results: Under the assumption that homology searching takes place mainly on the surface of the nuclear membrane (2D-model), the mean pairing time was about 3000 simulation steps (with considerable deviations above), and the mean frequency of interlocking was 14% for a single pair of chromosomes and 71% for all pairs in the nucleus. In case of a search within the whole nucleus (3D-model) the mean pairing time increased by a factor of three to four, and interlocking frequencies were 20% per pair and 83% per nucleus, respectively. This indicates that interlocking should occur in eight out of ten cell divisions.

In Section 3, finally, we set up a stochastic model allowing for an analytical derivation of the distribution of pairing times and interlocking events. We use the simulation results of Section 2 to estimate model parameters and obtain explicit expressions for mean pairing times, interlocking probabilities, etc. Moreover, it is possible to describe the dependence of our results on the number of chromosomes per nucleus. Thus a description of chromosomal behavior can be achieved which is valid for various species and is far more general than the simulation results.

2. A Cellular Automaton Model

2.1. Cell Geometry and Transitions

For our computer simulation of homology searching and interlocking we describe the movement of chromosomes in the cell nucleus by a cellular automaton model (see [3]). For that purpose we consider the nucleus as a sphere which is divided into volume elements of equal size by means of a three-dimensional grid, gen-

erated by l circles of longitude, $2b+1$ circles of latitude and $t+1$ spherical shells (cf. [2]). Thus we obtain $2bl+2$ cells on the surface of the sphere and

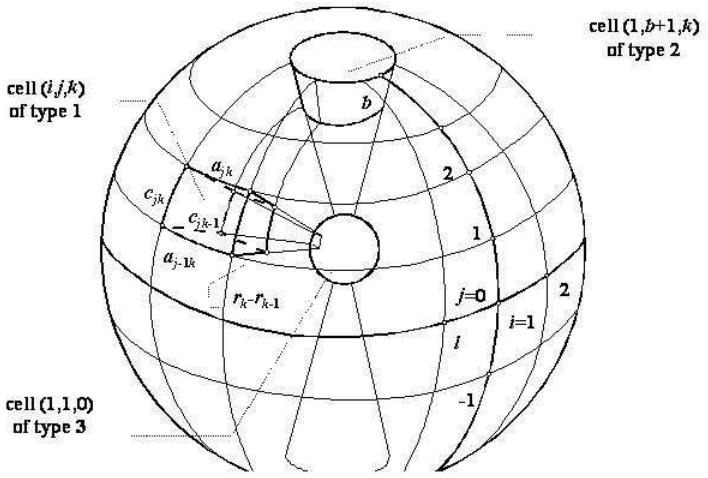

Figure 1: Cellular grid model of the cell nucleus

$(2bl+2)t+1$ volume cells which can be subdivided into three different types according to their neighborhood relations (see Figure 1).

Next we describe the transition behavior within the nuclear grid by a regular Markov chain proceeding in discrete space and time. We allow for transitions from each grid cell (i,j,k) to any neighboring cell $P(i,j,k)$. Thus, if e.g. (i,j,k) is a cell of type 1, then there are possible transitions to $(i\pm 1,j,k)$ along the j-th circle of latitude, to $(i,j\pm 1,k)$ along the i-th circle of longitude or to $(i,j,k\pm 1)$ from one spherical shell to another (if $k<t$). Moreover, we assume that: (i) the transition probabilities for two neighboring cells are proportional to their common boundary area, and (ii) the limit distribution of the regular chain (see [4]) is proportional to the cell volumes such that eventually each grid cell can be accessed with the same probability. More details concerning cell geometry and transition behavior can be found in [1] and [2].

2.2. Model Dynamics

Let us consider n pairs of homologous chromosomes m_ν and f_ν, $\nu=1,\ldots,n$ in each nucleus. We represent every male and female chromosome by three sites denoted by $m_{1\nu},m_{2\nu},m_{3\nu}$ and $f_{1\nu},f_{2\nu},f_{3\nu}$, respectively. Two sites are located at the ends of each chromosome and serve as recognition sites during homology

searching, and one site is located in the middle of the chromosome and has been introduced to check for chromosome interlocking (see Figure 2). After

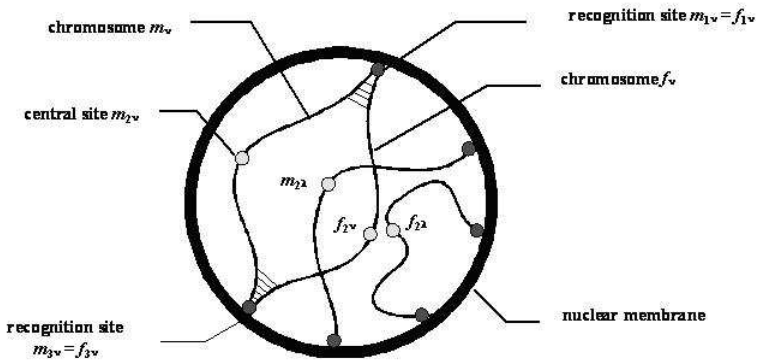

Figure 2: Interlocking of chromosomes

two corresponding recognition sites have met each other, the pairing process of the homologous chromosomes will begin and, depending on the position of the remaining chromosomes, interlocking may occur.

Starting with a discrete uniform distribution for all chromosome sites, the pairing process proceeds in discrete time steps $t = 0, 1, 2, \ldots$ according to the following model equations (where we identify the sites $m_{\mu\nu}$ and $f_{\mu\nu}$ with the cells of the nuclear grid containing these sites at time t, respectively):

$$\left.\begin{array}{l} m_{\mu\nu}(t+1) = P(m_{\mu\nu}(t)),\ \mu = 1,2,3 \\ f_{\mu\nu}(t+1) = P(f_{\mu\nu}(t)),\ \mu = 1,2,3 \end{array}\right\} \text{if } m_{\mu\nu}(t) \neq f_{\mu\nu}(t),\ \mu = 1,3 \tag{1}$$

$$\left.\begin{array}{l} m_{1\nu}(t+1) = f_{1\nu}(t+1) = m_{1\nu}(t) \\ m_{\mu\nu}(t+1) = P(m_{\mu\nu}(t)),\ \mu = 2,3 \\ f_{\mu\nu}(t+1) = P(f_{\mu\nu}(t)),\ \mu = 2,3 \end{array}\right\} \begin{array}{c} \text{if WLOG } m_{1\nu}(t) = f_{1\nu}(t) \\ \text{and } m_{3\nu}(t) \neq f_{3\nu}(t) \end{array} \tag{2}$$

$$\left.\begin{array}{l} m_{\mu\nu}(t+1) = f_{\mu\nu}(t+1) = m_{\mu\nu}(t), \\ \qquad\qquad\qquad \mu = 1,3 \\ m_{2\nu}(t+1) = P(m_{2\nu}(t)) \\ f_{2\nu}(t+1) = P(f_{2\nu}(t)) \end{array}\right\} \begin{array}{c} \text{if } m_{\mu\nu}(t) = f_{\mu\nu}(t),\ \mu = 1,3 \\ \text{and } m_{2\nu}(t) \neq f_{2\nu}(t) \end{array} \tag{3}$$

$$m_{\mu\nu}(t+1) = f_{\mu\nu}(t+1) = m_{\mu\nu}(t),\ \mu = 1,2,3 \text{ if } m_{\mu\nu}(t) = f_{\mu\nu}(t),\ \mu = 1,2,3 \tag{4}$$

for all $\nu = 1, \ldots, n$. Moreover, we restrict chromosome movements to

$$d(m_{1\nu}(t), m_{2\nu}(t)) \leq \frac{d_0}{2} \text{ and } d(m_{2\nu}(t), m_{3\nu}(t)) \leq \frac{d_0}{2} \tag{5}$$

for any t, and the same is supposed for $f_{1\nu}, f_{2\nu}$ and $f_{3\nu}$. In equation (5) $d(x, y)$ denotes the distance of cells x and y, and d_0 is the average chromosome length. After the first encounter of two corresponding recognition sites (i.e. in case of equation (2)) pairing of the involved homologous chromosomes will start and from that time on we require that

$$d(m_{\mu\nu}(t+1), f_{\mu\nu}(t+1)) \leq d(m_{\mu\nu}(t), f_{\mu\nu}(t)),\ \mu = 2, 3. \tag{6}$$

If both recognition sites of an homologous pair m_ν, f_ν have met independently (i.e. in case of equation (3)), a further chromosome, say m_λ or f_λ may be enclosed by m_ν and f_ν (see Figure 2). In order to decide whether interlocking takes place or not, our simulation program checks, if the central site $m_{2\lambda}$ or $f_{2\lambda}$ of chromosome m_λ or f_λ, respectively, belongs to the tetrahedron spanned by the cells $m_{1\nu}, m_{2\nu}, f_{2\nu}$ and $m_{3\nu}$. This interlocking test $IL(m_{1\nu}(t), m_{2\nu}(t), f_{2\nu}(t), m_{3\nu}(t), x)$ is carried out for all remaining central sites $x = m_{2\lambda}, f_{2\lambda}, \lambda \neq \nu$. To be more precisely, we check for interlocking by equation

$$IL(x_1, x_2, x_3, x_4, y) \Longleftrightarrow \begin{cases} \operatorname{sgn} D(x_1, x_2, x_3, x_4) = \operatorname{sgn} D(y, x_2, x_3, x_4) \\ = \operatorname{sgn} D(x_1, y, x_3, x_4) = \operatorname{sgn} D(x_1, x_2, y, x_4) \\ = \operatorname{sgn} D(x_1, x_2, x_3, y) \end{cases} \tag{7}$$

which involves the comparison of the signs of at most five 3×3-determinants $D(x, u, v, w) = |u-x, v-x, w-x|$. The simulation terminates if all homologous chromosomes are paired, i.e. if equation (4) holds for all $\nu = 1, \ldots, n$.

2.3. Simulation Results

In order to have proper values for the model parameters we use experimental data from yeast (see [1, 2, 6]). There are $2n = 32$ chromosomes in each yeast nucleus, about half of which have only one and half have two recognition sites. According to the dimensions of nuclei and chromosomes, we have chosen the net parameters as $b = 7, l = 28$ and $t = 4$ which results in a grid of 394 surface cells and 1577 volume cells per nucleus.

Altogether we compared four different approaches: Model $2D/1$ is based on the assumption that recognition of homologous chromosomes is bound to the nuclear membrane and there is only one recognition site per chromosome. In model $2D/1.5$ we assume that half of the chromosomes have one and half have two recognitions sites, respectively. In the same way we define models $3D/1$ and $3D/1.5$, where recognition of chromosomes takes place inside the nuclear lumen. The results of more than 2000 simulations per model are summarized in Table 1. It is evident that homology searching is more efficient for the $2D$-models

than for the $3D$-models, and it is also more efficient with two recognitions sites per chromosome instead of one. However, interlocking of chromosomes is more likely in case of $3D$-search than it is in case of the corresponding $2D$-search.

model	$2D/1$	$2D/1.5$	$3D/1$	$3D/1.5$
sample size	2341	2207	2150	1825
pairing time				
mean	3247.7	2966.1	12077.4	9698.3
min.–max.	765–10843	699–10442	3074–31320	2370–28925
std.dev.	1235.3	1136.8	4446.8	4190.1
interlocking prob.				
mean per pair		14.0%		20.1%
mean per nucleus		70.5%		83.1%

Table 1: Chromosome pairing times and interlocking probabilities

3. A Stochastic Approach

3.1. Mean Pairing Times

In the following section we extend our simulation results by analytic considerations. First of all we ask, how long does it take for the formation of the first, second, ... pair of homologous chromosomes? What is the total mean pairing time?

Let us denote by $S_1, S_2, \ldots, S_n$ the numbers of steps until the formation of chromosome pair number $1, 2, \ldots, n$ during cell division, and let $T_1, T_2, \ldots, T_n$ be the number of steps until the formation of the first, second, ..., n-th pair of chromosomes. If we assume that the random variables $S_1, S_2, \ldots, S_n$ are independent and identically distributed with geometric distribution $G(p)$, i.e. with generating function

$$\psi(s) = \sum_{k\geq 0} q^k p s^k = \frac{p}{1-qs} \tag{8}$$

(where $q = 1-p,\ p \ll 1$), then the mean searching time $E_0 = E(S_\omega)$ for any pair $\omega = 1, \ldots, n$ is

$$E_0 = \psi'(1) = \frac{q}{p} \approx \frac{1}{p}. \tag{9}$$

Further it can be shown that time differences between two successive pairing events $\Delta T_\omega = T_\omega - T_{\omega-1}$ for $\omega = 1, \ldots, n$ (where $\Delta T_1 = T_1$) are independent and approximately geometrically distributed according to $G(1-q^{n-\omega+1})$. Hence the distribution of T_ω, which is a sum of geometrically distributed random variables, is determined by its generating function

$$\psi_\omega(s) = \frac{(1-q^n)(1-q^{n-1})\cdots(1-q^{n-\omega+1})}{(1-q^n s)(1-q^{n-1}s)\cdots(1-q^{n-\omega+1}s)}. \tag{10}$$

The corresponding expectation, i.e. the mean pairing time until formation of the first ω chromosome pairs, turns out to be

$$\begin{aligned} E(T_\omega) = E(T_1) + E(\Delta T_2) + \cdots + E(\Delta T_n) &= \frac{q^n}{1-q^n} + \cdots + \frac{q^{n-\omega+1}}{1-q^{n-\omega+1}} \\ &\approx E_0(\frac{1}{n} + \cdots + \frac{1}{1-\omega+1}) \approx E_0 \ln \frac{n+1}{n-\omega+1} \end{aligned} \tag{11}$$

for $\omega = 1, \ldots, n$. In particular the total mean pairing time is given by $E(T_n) = E_0 \ln(n+1)$. Mean pairing times $E(T_\omega)$ as a function of ω (for fixed n) as well as total pairing times $E(T_n)$ as a function of n are visualized in Figure 3 and show good accordance to our results obtained in [1].

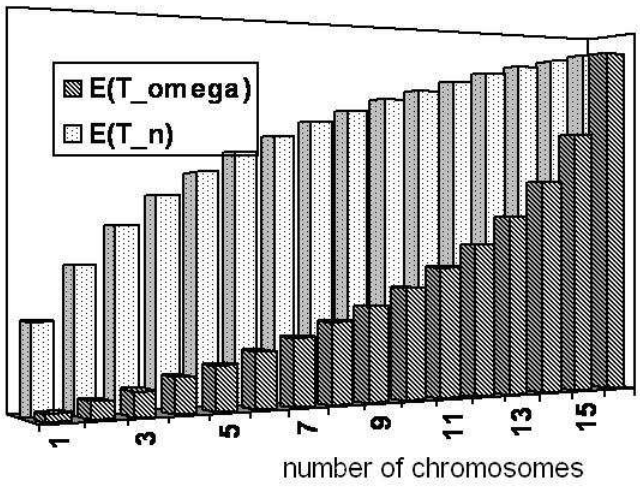

Figure 3: Pairing times $E(T_\omega)$ up to the occurrence of the first ω chromosome pairs, $\omega = 1, \ldots, n$ (front) and total pairing times $E(T_n)$ in case of $2n$ chromosomes (back)

3.2. Interlocking Probabilities

For the computation of interlocking probabilities let us consider the pairing process of the $2n$ homologous chromosomes once more, and distinguish between two possible alternatives during the formation of a pair of homologues: Either two homologues pair from both ends simultaneously and hence up to $2n-2$ chromosomes could be locked (we call this event E_1, see Figure 2), or homologues zip from one end to the other such that no interlocking occurs (we call this event E_2). We define the probabilities of these events by $P(E_1) = k$ and $P(E_2) = 1-k$, respectively, where $0 \leq k \leq 1$. Moreover, in case of E_1 we denote the probability of inclusion for each single chromosome or pair of homologues by p, and set $q = 1-p$. Under these assumptions the interlocking probability $p_{IL}(\omega)$ that chromosome interlocking occurs during formation of the ω-th pair of homologues derives to

$$\begin{aligned} p_{IL}(\omega) &= 1 - P(\text{no IL}|E_1)P(E_1) - P(\text{no IL}|E_2)P(E_2) \\ &= 1 - kq^{2n-\omega-1} - (1-k) \\ &= k(1-q^{2n-\omega-1}) \end{aligned} \tag{12}$$

for $\omega = 1, \ldots, n$, and hence the interlocking probability p_{IL} per nucleus is given by

$$\begin{aligned} p_{IL} &= 1 - \textstyle\prod_{\omega=1}^{n} (1 - p_{IL}(\omega)) \\ &\approx 1 - \exp\left(-\textstyle\sum_{\omega=1}^{n} p_{IL}(\omega)\right) \\ &= 1 - \exp\left(-kn + \frac{k}{p} q^{n-1}(1-q^n)\right). \end{aligned} \tag{13}$$

In order to obtain estimates for the parameters k and p, we make use of the relative frequencies $P(E_2)$ of pairings in case of event E_2 as well as the mean interlocking probabilities $\overline{p_{IL}}$ per pair from our simulations. With these frequencies we compute k and p from

$$P(E_2) = 1-k \text{ and } \overline{p_{IL}} = \frac{1}{n}\sum_{\omega=1}^{n} p_{IL}(\omega) = k\left(1 - \frac{q^{n-1}(1-q^n)}{np}\right) \tag{14}$$

(in the case of yeast for $n = 8$ pairs of homologues with two recognition sites).

Using equation (13) expected interlocking probabilities can be computed and compared to our simulation results (see Table 2). Moreover, by means of equation (13) it is also possible to consider the dependence of the interlocking probability p_{IL} on an arbitrary number of chromosomes, since the basic parameters k and p are related to the formation of an arbitrary pair of homologues and do not depend on the total number of chromosomes in the nucleus. It can

model	$2D/1.5$	$3D/1.5$
sample size	2207	1825
parameter estimates		
k	0.574	0.372
p	0.0265	0.0727
interlocking prob. per nucleus		
expected	67.4%	80.0%
observed	70.5%	83.1%

Table 2: Comparison of expected and observed interlocking probabilities

be seen from Figure 4 that p_{IL} closely comes up to 1 if $n \geq 15$.

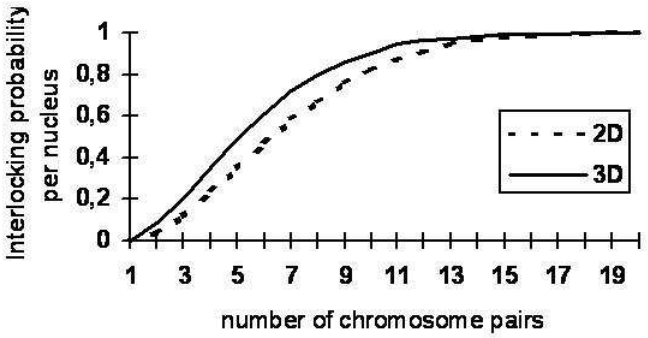

Figure 4: Interlocking probabilities for various chromosome numbers

3.3. Interlocking Distribution

Finally we consider the question, how many chromosomes can be locked up during the formation of one pair of homologues, if chromosome interlocking occurs? In order to derive this interlocking distribution let X_ω denote the number of entrapped chromosomes (in case of event E1, see Section 3.2) during the formation of the ω-th chromosome pair, $\omega = 1, \ldots, n$, and let X be the number of entrapped chromosomes for an arbitrary pair of homologous chromosomes. During the formation of the ω-th chromosome pair there are already $\omega - 1$ pairs, the two chromosomes belonging to pair ω, and $2(n - \omega)$ unpaired chromosomes in the nucleus. Thus X_ω can be composed as a sum of

$\omega - 1$ independent random variables which are distributed according to $B_2(p)$ and $2(n-\omega)$ independent random variables which are distributed according to $B_1(p)$, where $B_1(p)$ denotes the Bernoulli distribution Z on $\{0,1\}$ with $P(Z=0) = 1-p = q$, $P(Z=1) = p$ and $B_2(p)$ is defined in the same way on $\{0,2\}$, i.e. $P(Z=0) = q$, $P(Z=2) = p$. In particular X_1 has a standard binomial distribution $B(2n-2,p)$. The generating functions of $B_1(p)$ and $B_2(p)$ are $\psi_1(s) = ps + q$ and $\psi_2(s) = ps^2 + q$, respectively. Therefore the generating function $\psi_\omega(s)$ of X_ω is given by

$$\psi_\omega(s) = \psi_2(s)^{\omega-1}\psi_1(s)^{2(n-\omega)} = (ps^2+q)^{\omega-1}(ps+q)^{2(n-\omega)},\ \omega = 1,\dots,n. \quad (15)$$

It follows that the expectations $E(X_\omega) = \psi'_\omega(1) = 2(n-1)p$ of X_ω do not depend on the pair number ω.

Finally the distribution of X, i.e. the interlocking distribution under consideration, is determined by the generating function

$$\begin{aligned}\psi(s) &= 1 - k + \frac{k}{n}\sum_{\omega=1}^{n}\psi_\omega(s)\\ &= 1 - k + \frac{k}{n}\frac{(ps^2+q)^n - (ps+q)^{2n}}{pq(s-1)^2}.\end{aligned} \quad (16)$$

From equation (16) it follows that there will be $x = 0, 1, \dots, 2n-2$ chromosomes entrapped by interlocking with probability

$$\begin{aligned}P(X=x) &= \frac{\psi^{(x)}(0)}{x!}\\ &= \begin{cases} 1 - k + \frac{k}{np}q^{n-1}(1-q^n) & \text{if } x = 0\,,\\ \frac{2k}{np}q^{n-1}(1-q^n-npq^{n-1}) & \text{if } x = 1\,,\\ \frac{k}{np}q^{n-2}(np+3q-q^{n-1}(2n^2p^2+np(5q-1)+3q^2)) & \text{if } x = 2\,,\end{cases}\end{aligned} \quad (17)$$

and the mean number of locked chromosomes per pair of homologues is $E(X) = \psi'(1) = 2(n-1)kp$.

A linear approximation of equation (17) for small values of p shows that

$$\begin{cases} P(X=0) = 1 - \frac{3(n-1)}{2}kp\,,\\ P(X=1) = (n-1)kp\,,\\ P(X=2) = \frac{n-1}{2}kp\,,\\ P(X=x) = 0 \text{ if } x \geq 3\,,\end{cases} \quad (18)$$

which means that the probability for interlocking of two chromosomes is about half of the probability for interlocking of a single chromosome, whereas interlocking of three or more chromosomes hardly happens in reality.

Our simulation results concerning the distribution of interlocking compared to the theoretical probabilities given by equation (17) are summarized in Table 3. It can be seen that there is a rapid decrease of interlocking probabilities with

model	$2D/1.5$		$3D/1.5$	
interlocking	exp.	obs.	exp.	obs.
distribution $P(X = x)$				
$x = 0$	86.0%	86.0%	79.9%	79.9%
$x = 1$	7.9%	8.1%	8.3%	4.6%
$x = 2$	5.2%	4.4%	8.0%	2.8%
$x \geq 3$	0.8%	1.5%	3.8%	12.7%

Table 3: Interlocking distribution $P(X = x)$ for yeast

increasing number of locked chromosomes. In our yeast example the expectation $E(X)$, i.e. the mean number of locked chromosomes per pair in the $2D$-model turns out to be 0.21 which means that on average there is one chromosome entrapped out of five pairs of homologues.

References

[1] D. Dorninger, G. Karigl, J. Loidl, Simulation of chromosomal homology searching in meiotic pairing, *Journal of Theoretical Biology*, **176** (1995), 247-260.

[2] D. Dorninger, G. Karigl, J. Loidl, A cellular automaton model for chromosome interlocking in meiotic pairing, *Simulation Practice and Theory*, **6** (1998), 269-280.

[3] G.B. Ermentrout, L. Edelstein-Keshet, Cellular automata approaches to biological modeling, *Journal of Theoretical Biology*, **160** (1993), 97-133.

[4] W. Feller, *An Introduction to Probability Theory and its Applications*, Wiley and Sons, New York (1971).

[5] J. Loidl, The initiation of meiotic chromosome pairing: the cytological view, *Genome*, **22** (1990), 759-778.

[6] P.B. Moens, M.L. Ashton, Synaptonemal complexes of normal and mutant yeast chromosomes (Saccharomyces cerevisiae), *Chromosoma*, **91** (1985), 113-120.