# **Reconstructing DNA replication kinetics from small DNA fragments**

Hang and John Bechhoef and  $J$  and  $r^*$ 

*Department of Physics, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada*  $\zeta$  20 J<sub>b</sub> c 2006; published 5 May 2006)

In higher organisms, DNA replications, in provided a recent *in vitro* experiments have in vitro experiments have in vitro experiments have in vitro experiments have in the second in vitro experiments have in vitro experim  $\ell$  ded large amounts of data on the state of replication of the state of the time of the time  $\ell$ dependence of the average size of replication and non-replications, one can estimate the rate of initiation of initiation  $\mathcal{A}$  $\text{DNA}$  replication origins, as well as the average rate at  $\mu$  DNA bases are contributed. One problem in making  $\mu$ such estimates is that, in the experiments, the DNA is broken up into small fragments, whose finite size can bias downward the measured averages. Here, we present a systematic way of accounting for this bias by more leisurely discussion, see 13. The general situation is equal situation in the general situation is  $\mathbb{R}^n$  $i\nu$  f

 $m=1$ ization. Here, we use primes to denote distributions for density  $\mathbf{f}$  $\mathbf{r}$  is a general-cut case.

## **B. Distribution of edge domains**

To understand the distribution of edge domains, we first  $\mathbf{r}$ consider how a domain of size  $\mathfrak{a}$  is  $x' = \overline{y}$  is cut. Then we consider  $\overline{y}$ the probability density for a resulting edge domain to have a resulting edge domain to  $\mathbf{r}_i$ **s**  $x \times x'$ . In considering how a domain of  $x' = x' + y'$ , there are two cases  $x'$  1, then the probability density that  $\alpha$ it was  $x'$ . Since the cut position is assumed to be cut position is assumed to be assumed to be assumed to be

uniformly distributed along the probability density density density density density  $\mathbf{DNA}$ , the probability density densit that the cut creates a domain of  $\alpha$  is the  $\alpha$  is then  $\alpha$  is the  $\alpha$ factor of 2 arises because each cut creates *two* edges. Thus, the probability density density that the cut produces a domain of size  $\pi$ *x*  $(2/x')$   $x'=2$ .

In the second case, the original domain  $x'$  is larger than  $1$ and will always be cut. The probability of creating a domain  $\mathbb{R}^n$ **b**  $\overline{x}$  is now uniform over the fragment and is thus 2,  $\overline{y}$ the same result as we found in the first case. Finally, we first case  $\mathbf{r}$  as we first case. Finally, we find observe that an edge domain of size *x* can be created by any domain of  $x \geq x$ . This leads to the relative frequency of  $\Gamma$ observing an edge domain of  $x: x: x \mapsto x$ :

> $n_e(x) = 2 \int_{x}^{x}$  $(x')dx'.$  (3)

Normalizing Eq. 3-3-1 distribution dist  $e^{x}.$ 

To generalize the cuts are distributed as  $\tau$  $f(\ell), \qquad \mu \qquad \ell \qquad \qquad \text{. II A}, \quad \ell \qquad \text{E} \quad (3).$  $f(t)$  integrating over  $f(x, A, \ldots, 1)$  integrating  $x$ can lead to an edge of  $\mathfrak{p}(x)$  **x**. Thus, we have have  $\mathfrak{p}(x)$ 

$$
n'_e(x) = 2 \int_x^{\infty} f(\ell) d\ell \int_x^{\infty} (x') dx', \qquad (4)
$$

 $\int_{c}^{t} f(x) dx$  and  $\int_{c}^{t} f(x) dx$  are  $\int_{c}^{t} f(x) dx$  and  $\int_{c}^{t} f(x) dx$ 

#### **C. Distribution of oversized domains**

The simplest way to derive the distribution of oversized is derived in  $\mathcal{I}$ to recognize that there is a duality between domains and cut  $\epsilon$ fragments. That is, if one interchanges fragments with domains and cut locations with domain boundaries, then the oversized domains of the interior original case are just the interior origina  $d = \frac{1}{\sqrt{1 - (\mathbf{F}^2 - \mathbf{3})}}$ .

$$
A \quad i \qquad \qquad i \qquad \qquad \text{if } i \cdot r \qquad , \qquad \ldots
$$



FIG. 3. An illustration of the duality between domains and frage- $\mathbf{r}$  ( ) interior domains ( ) interior ( ) interior ( ) be derived from each other by interchanging fragments with domains. Shaded vertical wedges denote places where the molecule is  $\mathbf{r}_i$  $\mathbf{r}$ 

$$
n'_{o}(x) = f(x) \int_{x}^{\infty} (x' - x) (x') dx'.
$$
 (5)

 $E_f$  (5) interchanging from  $f = E_f$ , 2) interchanging from Eq. 2 $f(x) = \frac{1}{2\pi} \int_{0}^{x} f(x) \, dx$  $\ell'$  x'. To  $\mathbf{z}$  is the unit case, we simply let  $f_f$  be a single cut size  $f$  function, and is only a single cut size  $f$   $\mathcal{L}$ . Then

$$
n_o(x) = \delta(x - 1) \int_1^{\infty} (x' - 1) (x') dx'.
$$
 (6)

A distribution  $f(x)$  and  $f(x)$  and  $f(x)$  and  $f(x)$ ecule of DNA the total domain domain domain domain different domain different domain different domain different domain distribution of the total domain different domain distribution domain different domain different domain  $\mathbf{C}$  is the DNA. Table I summarizes on  $\mathbf{DNA}$ . Table I summarizes of  $\mathbf{DNA}$ . Table I summarizes of  $\mathbf{DNA}$ . Table the formulas describing the three different domains in the three d

 $u = \int_0^1 f(x) dx$  and  $\int_0^1 f(x) dx$  cases.

### **D. Example**

Finally, where  $\mathbf{r}$  is a example fig. 4). Let the original distribution between  $0$  and 100 into  $\sum_{i=1}^{n}$  by  $\sum_{i=1}^{n}$  be cut into fraction distributed between 50 and 150 units. Then  $x=X/100$ , the interior-domain frequencies are given by  $\mathbf{f}(\mathbf{r},t)$ 

$$
n_i(x) = \begin{cases} 1 & x, & 0 \quad x \quad 0.5, \\ (9/8) & (3/2)x + (1/2)x^2, & 0.5 < x \quad 1; \end{cases} \tag{7}
$$

$$
\mathbf{r} = \mathbf{r} \cdot \mathbf{r} \cdot \mathbf{r} \cdot \mathbf{r}
$$

TABLE I. Summary of the relationships between the fragment domain length distributions and the fragment domain length distributions and the fragment domain length distributions and the fragment distributions and the fragm original domain length distributions, for both unitors,  $f(x)$  for  $x$  and  $f(x)$  for  $x$   $x$  $1 \t 0 \t x>1.$ 

	Fr	$f - t$
	$f - r = 1$	$G \cap L$ $\vdots$
$I \cap C$	$n_i(x) = (x)(1 \ x) (1 \ x)$	$n'_i(x) = (x) \int_x^{\infty} (\ell \ x) \ f(\ell) d\ell$
Е	$n_e(x) = 2(1-x)\int_{x}^{\infty} (x')dx'$	$n'_e(x) = 2\int_{x}^{\infty} f(t) d\ell \int_{x}^{\infty} (x') dx'$
$\sqrt{1/2}$	$n_o(x) = \delta(x \ 1) \int_1^{\infty} (x' \ 1) \ (x') dx'$	$n'_{o}(x) = f(x) \int_{x}^{\infty} (x' - x) (x') dx'$

$$
n_e(x) = \begin{cases} 2(1-x), & 0 \le x \le 0.5, \\ 2[(3/2) - x](1-x), & 0.5 < x \le 1; \end{cases}
$$
 (8)  

$$
n_o(x) = \begin{cases} 0, & 0 \le x \le 0.5, \\ (1/2) -x + (1/2)x^2, & 0.5 < x \le 1. \end{cases}
$$
 (9)

 $T_{\text{r}}$  *n*<sub>i</sub>, *n<sub>e</sub>*, *n<sub>o</sub> n*<sub>i</sub> *n*<sub>*i*</sub>, *n<sub>o</sub> i n*<sub>*i*</sub> *n*<sub>*i*</sub>  $\lambda$  f in [13].

# **III. UNBIASED ESTIMATORS OF AVERAGE DOMAIN SIZES**

 $I \quad \text{I} \quad \text{I} \quad \text{DNA}$ into fragments led to three different types of subdomains: interior, edge, and oversized. We then  $i$ for calculating the frequency distributions of the frequency of the second terms of the these seconds of these seconds of these seconds of the second terms of the second terms of the second terms of the second terms of th domains, and  $\mathbf{f}$  and one is faced with the reverse problem:  $r$  and  $r$  and  $r$  and  $r$  $n_i$  **f i i n**<sub>*i*</sub>(x),  $n_e(x)$ ,  $n_o(x)$ ,  $\mathbf{r}$  can one inferred about the original distribution  $\mathbf{r}$  (x)?  $I \quad r \quad l \; , \; r \quad E \; . \; (2), \qquad \qquad \text{if} \qquad r \; r \; .$ ments of the fragment distribution  $f(\ell)$  $h_i(x) = 1$  for  $i = 1$  for  $i = 1$  $\hat{u}^{(x)}$ ; however,  $\hat{u}^{(x)}$  and  $\hat{u}^{$ algorithms for inferring replication in  $\mathbf{N}$ motivation for our study  $\mathbf{r}$  of the average only of the average of the average of the average of the average of the  $e$  replication and unreplication  $\mathcal{X}$  $\equiv \int_{0}^{\infty} x' (x') dx'$  if I in  $\equiv$  1 to estimate the average domain size  $\mathfrak{a}$  . In earlier work, we es  $\sum_{i=1}^{n}$   $\sum_{j=1}^{n}$   $(x_i)$  *j n*,  $\sum_{i=1}^{n}$   $(x_i)$  *j interior* -

$$
x_i^{tot} = \int_0^1 x (x) dx \int_0^1 x^2 (x) dx,
$$
  

$$
x_e^{tot} = \int_1^\infty (x) dx + \int_0^1 x^2 (x) dx.
$$
 (12)

Because the is only one size of the oversized domain,  $x_{o}^{tot}$  $= n_o,$  **E** . (10).  $x_i^{tot} + x_e^{tot} + x_o^{tot}$  $=\langle x \rangle$ . In dimensional units, we have  $\langle x \rangle$ 

$$
\bar{x}_2 = \frac{\bar{X}_2}{L} = \frac{X_i^{tot} + X_e^{tot} + X_o^{tot}}{N_i + N_e/2},
$$
\n(13)

- where  $X_i^{tot}$  is the total length of all observed domains, with order  $\mu$ analogous definitions for  $X_e^{tot}$  and  $X_o^{tot}$ . Intuitively, the sum of  $X_e^{tot}$ these quantities is just the total length of all the domains,  $\mathbf{u}$ where interior, edge, or oversized. Dividing this by interior,  $\mathbf{r}_i$ the total number of  $(N_i+N_e/2)$  and  $r$  is the gives our estimated on  $N_i$  $\mathbf{r}$
- As  $f(x) = f(x)$  distribution of fragmentalization of fragmental  $f(x)$
- $s$  is straightforward and leads  $i$  to  $r$  and  $r$  by  $r$
- $\langle L \rangle$  **E**. (13)  $\varphi$  **I** T 96 52 **F** 06.981 **II C**,  $x_i^{tot} = 71/3836270.333$  915 11971 T = T 915 11971 T = 11971 T = 11971 T + 1 139 .-356II-250C,-356 7584  $(13)91$ T, T 96 52 T 06.981

not a reliable indication of the  $\mathbf{DNA}$ 's status of the  $\mathbf{r}$ cation. A simple way to get a simple way to get a simple  $\mathbf r$  is to solution that  $\mathbf r$ data by replication fraction  $\mathbf{I}$  , we time that  $\mathbf{I}$  $u \in \mathcal{U}$  replication fraction fraction fraction  $\mathcal{U}$  $i$  and  $A$  more sophisticated way to deal with the data in- $\mathbf{r}$  is the starting-time distribution and decomposition and decompositio  $\mu$  [10].

 $S$  since our theory relies heavily on the assumption that cuts heavily on the assumption that cuts  $\mathcal{S}$ on the DNA of the Contract with extensive any  $\mathbf{p}$  and  $\mathbf{p}$  and the along molecule, we first show to test this assumption on the test this assumption on th

$$
= \int_0^\infty dx \, x \, (x) \int_x^\infty d\ell \, f(\ell) + \int_0^\infty dx \, (x) \int_0^x d\ell \, \ell \, f(\ell),
$$
\n(A4)

A 
$$
f
$$
  $f$   $f$   $f$   $f$   $f^{\infty} dx$   $f(x)=1$ .  
\n(A4)  $f$   $f$   $f(x)$   $f(x)$   $f(x)$   $f(x)$ 

$$
n'_i + n'_e/2 = \int_0^\infty dx \quad (x) \left[ \int_0^x d\ell \ \ell \ f(\ell) + \int_x^\infty d\ell \ \ell \ f(\ell) \right]
$$

$$
= \int_0^\infty dx \quad (x) = 1,
$$
 (A5)

$$
\langle \ell \rangle = 1 \qquad \text{if} \qquad i \qquad \text{if} \qquad \mathbf{E} \tag{5}
$$
\n
$$
\mathbf{f} \qquad \mathbf{f} \qquad \mathbf{f} \qquad \mathbf{E} \tag{6}
$$

$$
n'_{o} = \int_{0}^{\infty} d\ell \int_{\ell}^{\infty} dx \, x \, (x) \int_{0}^{\infty} d\ell \, \ell \int_{\ell}^{\infty} dx \, (x)
$$
\n(A6)

$$
= \int_0^\infty dx \, x \, (x) \int_0^x d\ell \, f(\ell) \int_0^\infty dx \, (x) \int_0^x d\ell \, f(\ell). \tag{A7}
$$

 $\mathfrak{X}^{\otimes n}$ 

 $n'_{o} + n'_{e}$