Persistence Length of Chromatin Determines Origin Spacing in *Xenopus* Early-Embryo DNA Replication

KEY WORDS

completion problem, persistence length, replication \mathbf{r}_1 factory, $\mathbf{r}=\mathbf{r}$ for $\mathbf{r}=\mathbf{r}$

ACKNOWLEDGEMENTS

 $\bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet \quad A \quad \bullet \quad \bullet \quad B. \quad A_{\bullet} \rightarrow \bullet \quad \bullet \quad B \rightarrow \bullet$ $H, A.$ \ldots \ldots \ldots J. \ldots \ldots B. \ldots and \ldots E. $\mathcal{L}(\mathcal{S})$ for calculation for containing of the manuscript. We are set the manuscript. We are set the manuscript. are particularly grateful to $J.A.H.$, and $P.A.A.$ as the form of the invariance of \mathbf{r}_1 and invariance comments and invariance comments and invariance comments and discussions.

 \mathcal{T} we now \mathcal{T} E C(Canada), E rance \mathbf{F} ation de France, and NIH.

DA replication, *in provisiae*, *random-simples in simple entering in simple euranteering* in somation in some such as *S. cerevisiae*, an ically set of the sequence plays and interest role in defining original P . Δn replication. In *Xenopus* and *Drosophila* early embryos, by contrast, replication or any specific DNA sequences. If potential original are distributed randomly along the specific genome, one expects a geometric (exponential) distribution of separations. Because the length of S phase is determined by the replication of the entire genome, even relatively rare long gaps could be phase beyond its observed $10-20$ minutes for complete duplication of the whole generated $(3, \ldots, \ell, \ell, \ell)$. 2,3 The problem is all the more acute in the that early embryo cells lack an efficient $\mathcal{A}=\mathcal{A}^{\mathcal{A}}$ which is used by many many eukaryotic cells to delay entry into mitosis into menyer punto mitosis DNA.
This entry into mitosis in the presence of unreplicated DNA . problem is formally stated as the the σ "random-completion problem, because of the reasons of the r explained above, its solution requires a mechanism that regulates replication θ

sequence.

oughly, two approaches have been advanced to resolve the random-completion to resolve the random-completion to p the first scenario ("original origin"), potential origin redundance and initiate stochastically throughout S phase. This allows large gaps to be "filled in" during the later stages of $\sqrt{38}$ $\overline{\text{L}}$, i.e. ϵ , ϵ , ϵ , ϵ and ϵ , ϵ $m_{\tilde{g}}$ in the distribution of a potential original origins, thus potential origins, thus preventing the formation of problematic large gaps between origins. $9 \, \rm L_{\odot}$, we shall show $\rm v$ that consideration of recent experimental results on early embryo *Xenopus* replication leads to a more number of the number of the second scenarios elements of both scenarios θ both scenarios and θ m_{α} important, suggests a biological picture in which the secondary structure of chromosome of chromosome of chromosome of chromosome in ℓ mating in particular and in particular role in D An replication.

ne recent development is that new experimental techniques now make it possible to extract large amounts of data from the replication process. For example, molecular-combing 10 and direct visual hybridization (DIRVISH) 3,11 techniques can give detailed statistics about detailed statistics a n numbers and sizes of replications as averaged over the genome, as well as n

In previous analysis, we examined the average lengths of eyes and holes are average lengths of eyes and holes a at different times during S $\rm H$ $\rm \sigma$, we focuse that distribution $I_{\rm 12i}$ of eye-to-eye distances in order to the origin spacing ϵ wormlike chain model of chromatin fibers, as well as well as the origin synchrony. We also generalized the correlation measurements of Blow et al. 9 In our simulations, we can detect origin synchrony through correlations in the sizes of nearby replications (or eyes). Adjacent (small) eyes of sizes of sizes size will have interested at about the same time. The correlation coefficient is con defined as

dynamical configurations of chromosomes, using crosslinking to α measure interaction frequencies between different graphs $\frac{15}{\pi}$ amount and quality of the data from the data from the second experiments is α stimulation of the formation of \mathcal{S}_1 of \mathcal{S}_2 . \mathcal{D} And \mathcal{D} $\cdot\cdot\cdot^8$

Here, we show that recently obtained molecular-combing data on $\mathcal{F}_\mathbf{q}$ D A θ **b** θ **replication** *Xenopus laevis* **and most naturally**explained by postulating that chromatin forms loops at θ for $f \in \mathbb{R}^{16,17}$ and these loops control origin spacing $f \in \mathbb{R}^{16}$ spacing (Fig. 1); It is important to note that the size of such a loop is not arbitrary. The size of arbitrary. The size of arbitrary. The size of $\mathcal{A}_\mathbf{z}$ stiffness of the polymer means that loops that are too small cost too small cost too small cost too much energy. If a loop is too large, there will be too many conformation $I = \frac{1}{2}$ mations to explore for the ends to meet, and it thus costs to much contain the ends to much contain the ends to entropy. Balancing these effects gives an optimal loop size, $\frac{18}{\sqrt{18}}$ which leads to an original variable \mathcal{E} and \mathcal{E} at a since origins are connected by at a since \mathcal{E} $\overline{\mathbf{r}}$ and single loop.

The sizes of the postulated by fitting to experimental by fitting to experimental to experimental to experimental the sizes of the sizes of the size , then we called the turn of $\mathbf{r}_1 = \mathbf{r}_2$, then $\mathbf{r}_1 = \mathbf{r}_2$ obtained independent dently in single-molecule molecule measurements of chromatin stiffness in \mathcal{E}_c other systems. 15,19 Because the size of a polymer loop is controlled in polymer loop is controlled in the size of a polymer loop is controlled in the size of a polymer loop is controlled in the size of a polymer loop by its stiffness, we can link the physical properties of chromatin, ϵ when considered as a semiflexible polymer, the polymer spacing during duri D And P replication. As we shall see, the physical properties of chromosomers of chromoso matin loops can explain both the observed regularity of initiation of initiation of \mathcal{F}_max s_p and $\frac{9}{3}$ and the existence of an $\frac{1}{3}$ where $\frac{1}{3}$ where $\frac{1}{3}$ where $\frac{7}{3}$ where $\frac{7}{3}$ where ℓ is inhibited, reconciling apparently contradictory views ℓ in ℓ in ℓ on the nature of the mechanism that ensures $r_{\rm eff}$ and completely genome replication in early embryos. Although our results concerns one particular system, there is reason to suspect that they $m \rightarrow m \rightarrow m$

MATERIALS AND METHODS

Analysis of Molecular Combing Experiments on Early-Embryo *Xenopus.* We analyze data from the recent molecular combine ℓ $\mathbf{B}n_{\mathbf{x}}$ the \mathbb{P}^2 Theodore are available on request. These experiments used. $Xenopus \rightarrow \infty$, $\bullet \quad Xenopus \rightarrow \infty$ labeling D -A_b bases to study the kinetics of D -A n replication in this system. ne by labeling by labeling the species of the species θ (biotin-d_u during as ϵ , some Δ as ϵ , some time ϵ R replication process, one adds the second dye (dig-dUTP, green) and thus thus θ DNA replicated after *t'* are labeled with two colors (predominantly green). $F = n - \frac{1}{2}$ steed D A are strengthed out uniformly on a glass surface using u $m_{\tilde{g}}$ combined under a microscope (stretching factor: 1 μ) $= 2.0$ $\sqrt{0.1}$ k in Fig. 8 in n_1 , 20). The alternative regions reflected regions regions regions regions regions in \vec{r} form a snapshot of the replication state of the D -A fragment at time the time \mathbf{r}_g and time in different runs added. Varying that time in different runs allows one to \mathcal{A} is the problematically look at the progression of \mathcal{A} phase. The phase of replication throughout S phase.

where $s_j(s_j)$ is the *i*-th (*j*-th) eye size and brackets (< \Rightarrow) denote a size and brackets (T is neighborhood distance i -j j indicates j indicates α and apart. For example, $C(1)$ is the correlation coefficient for nearest nearest nearest $C(2)$ for next-nearest, and so on.

Looping of a Helical, Wormlike Polymer Chain: Statistics and Dynamics. In forming loops (see Fig. 1, for example), polymers that have an intrinsic stiffness such as chromatin cannot have arbitrary loop such as chromatin can not size α optimal loop size is 3–4, the persistence length (and persistence of the poly m stiffness). 18 Previous work dealing with looping in biological contexts with looping in biological contexts has implicitly assumed that looping is a reaction-limited process, i.e., \mathcal{A} where κ reactive groups meat meet many times before actually binding. limit, the kinetic distribution of loop sizes is identical to the distribution of Γ

loops in the third equilibrium. For this case, Shimada and Yamakawa (SY) derived an approximate θ and θ is θ and θ is θ

$$
(1)
$$

Here, $G(t)$ *dl* is the probability for finding a loop whose size is between ϵ *l* and *l* and *l* = *L*/*l_p*, with L the polymer and L the polymer and L the persistence length. Notice that for small *l*, the loop-formation probability is exponentially suppressed, which provides a natural explanation for an original explanation for an origininitiation exclusion zone. The peak of the SY distribution at *l* = 3.4 can be e_1 ected initiative to construct the probe to enhanced initiations. Finally, for f_1 = ability decreases rapidly, which makes the formation of single large chromating ℓ arge chromating chromating ℓ loops undirected that Equation 1 does not accurately described that $\mathcal{L}_\mathbf{a}(\mathbf{r},\mathbf{r})$ large-*l* limit, which has been modeled more accurately as a Gaussian chain.²³ $I_{\rm eff}$ the dynamics are diffusive, i.e., i.e. the $\epsilon_{\rm e}$ encounter each other with some some small reaction $\#$ (< 1), we can show that the SY approximation continues to hold in the regime ℓ loop-size is less than a few times the persistence length, and the loop-forma- $\mathbf{r} \in \mathbb{S}_c$ is given

$$
(2)
$$

where C is a dimensionless prefactor that is proposed that is proposed (10^{-1}), a all l_1 and D is the diffusion constant.²⁴ \ldots first return time" \mathcal{S}_c predicted. \mathbf{E}_1 is very short very short (10⁻³ to 10⁻² seconds for chromatin, comparable to the that of A^{25}), implying that loop-formation dynamics are ℓ must faster than $r \mapsto r$ is the set of $(20 - r)$ is r .

Finally, one further approximation that has been made in this and previous n_{α} in work on looping is that the reaction \mathcal{M} is the polymer assumed to be the polymer assu ends, whereas in the case of case of chromatin, original $\mathbb{E}[\mathbf{A}(\mathbf{r}_{0},\mathbf{r}_{0},\omega_{0}%]^{T}$ the ends of the D -A) are assumed to bind to r and the plication factories that have \mathcal{A} already bound a neighboring origin, which is also in general notation, which is also in general not at the end of the chromatin molecule. We believe that the this is unlikely to be an important that the an important \mathcal{L}_max $\mathcal{L}_{\mathbf{z}}$ is a set of $\mathcal{L}_{\mathbf{z}}$

other that while the loop-size distribution does not accurately follow the size of accurately follow the α System outside the so-called Kramers regime $\mathbb{E}_{\mathbf{A}}$ was regime where 2 was derived, the following of chromatin falls with the following $\frac{24}{\sqrt{3}}$

Computer Simulations. The effect of adding the effect of adding characteristic of adding characteristic of adding chromating characteristic of adding characteristic order of adding characteristic order of adding contract

The difference between the distributions, ∆ρ*i2i* = ρ*i2i_exp* - ρ*i2i_random*, is shown in (Fig. 3B). Notice that there are two clearly distinct regimes. In the following for $r = \left(\frac{l_{12i}}{2} < 20\right)$, the experimental data clearly differ $i = 0$ $s_n = (P = 4 \quad 10^{-33}; '^2=165, r = 6, r = m$ are inhibited over original distances smaller than 8 kb (mostly smaller than 8 kb (mostly than 4–5 kb). This is consistent with both the observation that the observation that the observation that the o one origination in the smaller than 10×2 and the speculation that an exclusion \mathcal{L} and \mathcal{L} and \mathcal{L} On the other hand, activation of one origin appears to stimulate the activation of neighboring original separated by a distance of $8, 16$ kb (peak at 13 kb). This number is consistent with the previously reported origin spacing r of 5 –15 $h^{9,12}$ and the saturation density of *Xenopus* rigin ϵ $C \longrightarrow \bullet$ (C)^{26,27} along specified spe

The second regime (*l*₁₂₁ⁿ 20 kb) shows that for simulation and experimental for simulation and experimental simulation and experimental for simulation and experimental for simulation and experimental for simulation a the distribution of large eye-to-eye distances is statistically similar (*P* = 0.14; $x^2 = 34$, $r \le n = 26$), which is the random-initiation for random-initiation hypothesis that the random-initiation hypothesis that the random-initiation $r = r \cos \theta$. holds for this regime, even as it fails at smaller 28

Eye-Size Correlations and Origin Synchrony. We see the test of the t ρ respectively. The sizes of correlations between the sizes of nearby eyes. Figure 4 shows that there is a weak but statistically significant positive correlation: larger executive correlation: larger eyes tend to have larger neighbors, and vice versa. Because domains grow at constant velocity, size correlations may be interpreted as origin synchrony. The interpreted as origin synchrony. $v_{\rm c}$ value for the nearest-neighbor correlation is consistent with that $r_{\rm c}$ B $\cdot \cdot \cdot (0.16)^9$

The observation of eye-size correlation of executive significance in \mathcal{F} that no local initiation $I(x,t)$ when ℓ is form can probability ℓ correlations. 29 Intuitively, the presence of eye-size correlations means that size correlations means that the probability of initiating and origin is enhanced by the presence of nearby the presence of nearby $\mathcal{L}_\mathbf{z}$ active origins and thus cannot be a function on \overline{X} and \overline{t} (position along the generation of α phase α , we can calculate α in α months of α months of α months of α Carlo simulation that eye-size correlations assuming that origins are placed at σ random along the genome (■) and intiations are independent from one α and α expected, the correlations are consistent with α

Origin Spacing, Loops, and Replication Factories. *incluent to the experiment* and the to-executive is not consistent with the random-initiation is not consistent with the random initiation hypothesis for short distances (< 20 kb) and since eye-size correlations imply s , and called an alternative between ℓ interaction between ℓ hypothesis that chromatin folding can lead to a replication folding can lead to a replication factory with α $\left(\mathbf{m} \cdot \mathbf{r}, \mathbf{r}, \mathbf{r}, \mathbf{r}, \cdots, \mathbf{r}\right),$ 16,17,30 and $\mathbf{r}, \mathbf{r}, \cdots, \mathbf{r}$ initiations occur at the replication factory, and there must be a correlation between the loop sizes and the distances between replication origins. As mentioned earlier, because of the intrinsic stiffness of chromatin, loops have \mathcal{E}_c a preferred size: activated origins will tend to occur at a characteristic separation from the replication forks of already activated replication origins. In the Monte Carlo simulations, we compute for each time interval ∆*t* the number of initiations ∆*N(t)* = *I(t)* x ∆*t x L*' (where *L*' is the length of DA, is unreplied at time t , using the published result of $I(t)$, see **I**

Refe e ce

- 1. Gien D \ldots making sense of euranorial DNA replication original ℓ 2001; 294:96-100.
- 2. Hyrien \mathcal{C}_ℓ , \mathcal{C}_ℓ and terminates at random initiates and terminates at random in s_z intervals but at regular intervals in the ribosomal DA_p *Xenopus* early embryos. EMBO
- J 1993; 12:4511-20. 3. Blow JJ. Control of chromosomal DNA replication in the early *Xenopus* embryo. EMBO J 2001; 20:3293-3297.
- 4. Raff JW, Glover DM. Nuclear and cytoplasmic mitotic cycles continue in *Drosophila* embryos in $D \cap A$ synthesis is independent of $D \cap A$ synthesis in the J Cell Biol 1988; 48:399-407.
- 5. Laskey R. Chromosome replication in early development of *Xenopus laevis*. J E_{mbr} E , $1985; 89:285-96.$
- 6. Hyrien o, Marheine K, Goldar A. Paradoxes of european DNA replication: MCM proteins and the random completion problem. BioEssay 2003; 25:116-25.
- 7. $L_{\rm H}$, $L_{\rm H}$ Capacitier Miller Miller M, Sogo JM, D A *n* **c** replication in *x emopus* extensive random in *B* 2000;
296:769-86. 296:769-86.
- 8. Herrich J, Jun S, Bensimon A. Kinetic model of D. American and the μ \blacksquare , \blacksquare , \blacksquare J B 2002; 320:741-50.
- 9. B JJ, Gillespie J, E_{ran}cially, DA. **Replication origins in** *Xenopus* egg extract Are 5-15 kilobases apart and are activities and are at different times. J G Biology α 2001; 152:15-25.
- 10. Bensimon A, Simon A, Croquette A, Croquette v, Hersch, F, Bensimon D. Alignment and ϵ_1 , and ϵ_2 and $D-A$ by a moving interface. Science 1994; 265:2096-8.
- 11. γ γ γ γ γ C. Visualization of D A α replication on *Epstein-Barr* γ \cdots , \cdots , 2001; 294:2361-4.
- 12. Herrick J, Stanislav J, Hyrien J, Bensimon A. Replication for the replication for \mathbb{R} \cdot D A \cdot \cdot \cdot *X. laevis* \cdot \cdot \cdot \cdot \cdot J B 2000; 300:1133-42.
- 13. $R(x, y, \ldots, y, K_n)$ see EA, Collingwood D, Hunt S, $\sum_{i=1}^n C_i$ S, $\sum_{i=1}^n A_i$, \ldots eplication dynamics of the year ϵ as ϵ as ϵ 2001; 294:115-21.
- 14. \mathcal{W}_s JJ, Aprilog JG, Chen JB, Jennings EG, Young A. Genome-wide distribution of C and C and α is *S. cerevisiae*: Higherens in α and α replica- \cdots , \cdots , \cdots , 2001; 294:2357-60.
- 15. Decker J, κ K, Decker M, Ke_{ker} et al. Capturing chromosome conformation. \cdots , 2002; 295:1306-11.
- 16. Cook PR. The organization of replication and transcription and transcription Γ
- 17. \mathcal{A} \mathcal{A} \mathcal{A} \mathcal{A} \mathcal{A} replication or sequence specificity to existence specific to equation or equation of \mathcal{A} $\tilde{G}_{\cdot \cdot \cdot}$ 2001; 2:640-5.
- 18. Yamaka H. Helical wormlike chains in polymer solutions. Berlin: Springer, 1997.
- 19. G_{α} , B_{α} is given C_{α} by the fulling a single chromatin fiber reveals that maintain its maintain its $h = e^{-\mu}$. $h = \mu \cdot \mu$. Proc A_1 . Acad 2000; 97:127-32.
- 20. Allemand JF, Bensimon D, Julien L, Bensimon A, Croquette L, Hallenberg specific $b \cdot b \cdot b \cdot c$, $c \cdot b \cdot D \cdot A \cdot B \cdot J \cdot 1997; 73:2064-70.$
- 21. **R.** Making contacts on a nucleic acid polymer. Biology 2001; 26:733-40.
- 22. himada J, Yamakawa H. Ring-closure probabilities for twisted words. A_{pp} \cdot D A_{cro} \cdot \cdot 1984; 17:689-98.
- 23. Ringrose L, $C_{\alpha,\beta}$ and $C_{\alpha,\beta}$ and $C_{\alpha,\beta}$ are as AF. African propagative comparison $C_{\alpha,\beta}$ of DNA looping in vitro and in viros and increases effective DNA flexibility at short- \ldots , EBJ 1999; 18:6630-41.
- 24. Jen \mathbf{B}_k ere J, Ha By. Diffusion-limited loop formation of semiflexible polymers: Kramers the intertwined theory and the intertwined time scales of chain $\mathbb{E}[r]$ \ldots 2003; 64:420-6.
- 25. Merlit H , Rippe K, Kerlitz K, Langowski J. Looping dynamics of linear DNA molecules and the effect of D -A μ μ and μ as μ by Brownian Dyanmics D and μ μ and μ and μ 1998; 74:773-9.
- 26. **e** A, Choi_l J, Brown He, $E \rightarrow GH$, GI, Blog J. Interaction between the original recognition complex and the replication licensing system in *Xenopus*. Cell 1996; 87:287-96.
- 27. Walter J, Newport J, Newport J, Newport J, Newport and χ *Xenopus* enter the size in χ and 1997; 275:993-5.
- 28. The agreement between the two curves (experiment and random initiation) becomes between ϵ ter as the eye-to-eye-to-eye-to-eye-to-eye-to-eye-to- P_{π} and the two-eyeregimes (inhibition and enhancement) are most distinguishable when they are divided after when they are divided after α the first two oscillations, i.e., at around *l i2i* = 20 kb. On the other hand, we also note that the simple random-initiation hypothesis reproduces all mean quantities such as the mean eye size throughout $\mathcal{L}_{\mathcal{A}}(t)$ as shown in ref. 8.
- 29. Sekimoto K. Evolution of the domain structure during the nucleation-and-growth process $\mathcal{L}_{\mathbf{z}}$ with $n_1 + n_2 + n_3$ with $n_1 + n_2 + n_3$ and n_3 \ldots B 1991; 5:1843-69.
- 30. Berezney R, Dubey DD, Huberman JA. Heterogeneity of european generations, replications, rep ϵ ^clusters, and ϵ ²⁰⁰⁰; 108:471-84.
- 31. Buongiorno-Nardelli M, Micheli M, Marilley M, Mar size and supercoiled loop domains in the european of the european $\frac{1}{2}$ 82; 298:100-2.
- 32. Blumenthal AB, Kriegstein HJ, Hogeness DS. The units of D. A. M. _{The units} of *Drosophila melanogaster* chromosomes. Cold Spring Harbor Symphosomes. B 1973; 38:205-23.
- 33. Edga κ C, Tutter Av, Cotting C, Gilbert CH, Proton Ta, Michael JC. MC 2-7 complexes bind chromatin in a distribution surrounding the origin recognition σ
- **complex in** *Xenopus* et al., J B C **c** 2002; 277:33049-57.

34. \bullet . D A \bullet Annu Rev B \bullet 1992; 61:199-223. 35. **Rippe K, von Hipper Hipper J. Action at a distance:** \mathbf{F}_i \cdots transformation. TIBS 1995; 20:500-6.

- 36. Horn PJ, Peterson CL. Molecular biology. Chromatin here for r_a and the set of $t \cdot t$ since $t \cdot t$ since 2002; 297:1824-7.
- 37. Heun P, Laroche T, Shimada K, Furrer P, Gasser SM. Chromosome dynamics in the yeast \cdots interpreted. Science 2001; 294:2181-6.