

# The influence of volume exclusion by chromatin on the time required to find specific DNA binding sites by diffusion

S. A. Isaacson<sup>a,1</sup>, D. M. McQueen<sup>b</sup>, and Charles S. Peskin<sup>b</sup>

<sup>a</sup>Department of Mathematics and Statistics, Boston University, Boston, MA 02215; and <sup>b</sup>Courant Institute of Mathematical Sciences, New York University, New York, NY 10012

Contributed by Charles S. Peskin, December 17, 2010 (sent for review July 24, 2010)

Within the nuclei of eukaryotic cells, the density of chromatin is nonuniform. We study the influence of this nonuniform density, which we derive from microscopic images [Schermelleh L, et al. (2008) *Science* 320:1332–1336], on the diffusion of proteins within the nucleus, under the hypothesis that chromatin density is proportional to an effective potential that tends to exclude the diffusing protein from regions of high chromatin density. The constant of proportionality, which we call the volume exclusivity of chromatin, is a model parameter that we can tune to study the influence of such volume exclusivity on the random time required for a diffusing particle to find its target. We consider randomly chosen binding sites located in regions of low (20th–30th percentile) chromatin density, and we compute the median time to find such a binding site by a protein that enters the nucleus at a randomly chosen nuclear pore. As the volume exclusivity of chromatin increases from zero, we find that the median time needed to reach the target binding site at first decreases to a minimum, and then increases again as the volume exclusivity of chromatin increases further. Random permutation of the voxel values of chromatin density abolishes the minimum, thus demonstrating that the speedup seen with increasing volume exclusivity at low to moderate volume exclusivity is dependent upon the spatial structure of chromatin within the nucleus.

first passage time | gene regulation | stochastic reaction-diffusion

How do regulatory proteins and transcription factors find specific DNA binding sites? In considering this question, it is often remarked that the rate at which proteins find specific DNA binding sites can “exceed the diffusion limit.” This statement is normally interpreted to mean that the association rate for a protein to find a specific binding site is faster than the predicted rate for the protein to reach the binding site by diffusion (2). The question of whether proteins, in vivo, generally find binding sites faster than the diffusion limit is still an area of active research. One potential difficulty in addressing this problem is in understanding precisely what is meant by the term “diffusion limited” binding rate. Here we adopt the viewpoint that a diffusion limited rate refers only to the rate at which a protein undergoing pure diffusive motion in a spatially homogeneous environment finds a target binding site. This corresponds to the standard Smoluchowski diffusion limited reaction model (3). In the present paper, we consider the influence of a heterogeneous environment on the time to find a target by diffusion.

A number of mechanisms that could potentially decrease the search time for a binding site, in comparison to the search time in models involving only diffusion in an empty nucleus, have been proposed and studied in experimental assays and mathematical models. For example, in ref. 2 it was discussed how the inclusion of electrostatic interactions between the protein and binding site may make predicted association rates comparable to those measured experimentally. Several alternative mechanisms are based on the knowledge that many regulatory proteins and transcription factors have nonspecific DNA-binding interactions. For example,

in ref. 4 a model was developed in which proteins could undergo a mixed search process involving periods of three-dimensional diffusion, coupled to periods of one-dimensional diffusion (or sliding) along DNA fibers during which the protein is nonspecifically bound. This idea, and variants that take into account effects such as hopping between DNA strands, has been studied extensively in theoretical models (for example, see refs. 2 and 4–8). Experimental studies have also begun to investigate whether sliding occurs in vivo, and its relative importance. In ref. 9 it was shown experimentally by single-molecule imaging studies that sliding can occur in *Escherichia coli* cells. As most of the current studies have been focused on prokaryotic cells, it remains to be seen whether sliding along chromatin in eukaryotic cells can noticeably reduce the time required for regulatory proteins to locate specific binding sites. Several other proposed mechanisms that could decrease the search time, such as direct “jumping” between different regions of chromatin fibers, are discussed in ref. 2. (More complete references for both theoretical models and previous experimental work can also be found in ref. 2.)

Within the nucleus, proteins are moving through a complex spatial domain comprised of chromatin fibers with spatially varying compaction levels, nuclear bodies, and fibrous filaments (such as the nuclear lamina). This spatially inhomogeneous environment provides another possible influence on the process by which proteins search for specific binding sites. In ref. 10 the role of spatial differences in chromatin density, and of volume exclusion by chromatin, on the motion of proteins within the nucleus was investigated. Using a combination of experimental and computational studies, including photo-activation experiments, the authors concluded that chromatin dense regions, such as heterochromatin, exhibited noticeable volume exclusion compared with less dense regions (such as euchromatin). The authors’ photo-activation experiments gave similar fluorescence activation curves in heterochromatin and euchromatin (when normalized to the different steady-state fluorescence levels in each region). This was interpreted to mean that heterochromatin is not substantially more difficult for proteins to enter than euchromatin, but just had a smaller amount of free space in which proteins could accumulate. In contrast, some molecules may have difficulty moving into denser chromatin regions. For example, in the supplemental movies of ref. 11 individual mRNAs that are observed to move freely appear restricted within regions of low histone-GFP fluorescence.

Author contributions: S.A.I. and C.S.P. designed research; S.A.I. and D.M.M. performed research; S.A.I. and D.M.M. contributed new reagents/analytic tools; S.A.I. and C.S.P. analyzed data; and S.A.I., D.M.M., and C.S.P. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

<sup>1</sup>To whom correspondence may be addressed. E-mail: isaacson@math.bu.edu or peskin@cims.nyu.edu.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1018821108/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1018821108/-DCSupplemental).

In this work we develop a mathematical model to investigate the possible influence of volume exclusion by chromatin and also of binding site location in relation to the chromatin, on the time required for an individual protein to find a specific binding site. The model resolves the entire nuclear volume, and represents chromatin as a continuous field (based on the DAPI stain fluorescence imaging data of ref. 1). It is assumed that regions of increased chromatin density, as determined by the DAPI stain intensity, are more difficult to enter than regions of low density. By varying a parameter that determines the overall strength of volume exclusion in a global manner we study how the search time changes when there is no volume exclusion (i.e., the nucleus is spatially homogeneous), weak volume exclusion, and very strong volume exclusion of chromatin dense regions. At the whole nucleus scale it is not clear, a priori, what the influence of volume exclusion will be. For example, it may be that volume exclusion helps funnel proteins toward active binding sites by increasing the difficulty for them to enter regions of heterochromatin and/or by creating effectively one- or two-dimensional channels in which diffusive search is much faster than in a three-dimensional volume. In contrast, perhaps this same funneling effect could trap proteins in channels of low chromatin density, causing the protein to wander far from target binding sites. We find that for binding sites located within regions of low DAPI stain intensity, moderate volume exclusivity leads to the fastest search times (faster than in the case of zero volume exclusivity, and also faster than in the case of strong volume exclusivity). Moreover, we find that the benefit of moderate volume exclusivity is abolished by a random permutation of the voxel values of imaged chromatin density. This shows that the benefit is somehow related to the spatial structure of the chromatin (i.e., to the spatial correlations in chromatin density) because it is precisely these correlations that are destroyed by the randomization procedure, which preserves the overall chromatin density distribution intact. In contrast to the case of a binding site in a region of low or moderate chromatin density, binding sites within chromatin dense regions simply become more inaccessible as the volume exclusivity is increased.

In the next section we begin by formulating our mathematical model, based on mouse myoblast structured illumination microscopy data from ref. 1. We then describe the numerical method we used to simulate the random walk defined by our mathematical model, followed by a discussion of the results we observe for the behavior of the search time as a function of the overall volume exclusivity and also as a function of the chromatin density at the binding site location.

### Mathematical Model

Our goal is to study the time needed for a diffusing regulatory protein to find a specific binding site within the nucleus of a eukaryotic cell. We assume that if there were no chromatin within the nucleus the protein would undergo diffusive motion with a fixed, constant diffusion coefficient. To model volume exclusion by chromatin we use a repulsive potential. The strength of the potential will be chosen to vary as a function of the density of chromatin at a given location. Regions of higher chromatin density will be more difficult to diffuse into than regions of low density. We stress that this model incorporates only the influence of volume exclusion, and not other effects such as trapping or DNA sliding.

Our mathematical model is diffusion in a potential,  $U(x)$ , within a bounded domain,  $\Omega \in \mathbb{R}^3$ , representing the nucleus. We model the specific binding site the protein is searching for as a small absorbing sphere of radius  $r_b$  about the point  $x_b \in \Omega$ . Let  $D$  be the diffusion constant of the protein (with units of  $\mu\text{m}^2/\text{s}$ ),  $k_B$  Boltzmann's constant, and  $T$  temperature (in Kelvin). We denote by  $p(x,t|x_b)$  the probability density that the molecule has not yet bound to the binding site, and is located at  $x \in \Omega$  at time  $t$ , given that the binding site is at  $x_b$ . The time evolution of  $p(x,t|x_b)$  is then given by the Fokker–Planck equation

$$\frac{\partial p}{\partial t}(x,t|x_b) = D\nabla \cdot \left( \nabla p(x,t|x_b) + \frac{1}{k_B T} p(x,t|x_b) \nabla U(x) \right) \quad [1]$$

for  $x \in \Omega$  and  $|x - x_b| > r_b$ . Note, here the spatial derivatives are with respect to the  $x$  coordinate. Although it appears here that we are using a constant diffusion coefficient, see *SI Text* for a change of variables that leads to an alternate interpretation of Eq. 1.

Let  $\partial\Omega$  denote the nuclear membrane and  $\eta(x)$  the outward unit normal vector to the membrane at  $x \in \partial\Omega$ . The associated boundary conditions to [1] are then

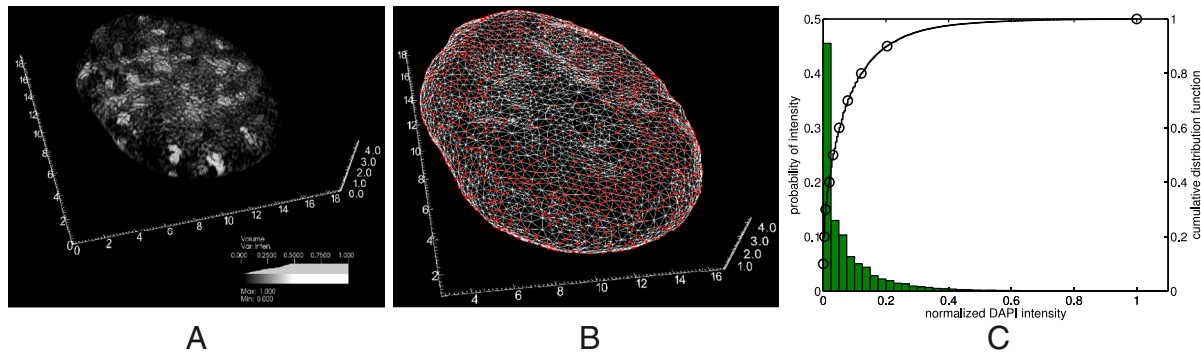
$$\begin{aligned} p(x,t|x_b) &= 0, & |x - x_b| &= r_b, & [2] \\ \nabla p(x,t|x_b) \cdot \eta(x) &= 0, & x &\in \partial\Omega. & [3] \end{aligned}$$

The first, Dirichlet, boundary condition models the binding reaction, whereas the second, Neumann, boundary condition models the assumed impermeability of the nuclear membrane to the regulatory protein.

As described in *Materials and Methods*, from the data of ref. 1, we were able to reconstruct a triangulated surface representation for the nuclear membrane (see Fig. 1B), and a discrete intensity field for the DAPI stained DNA (see Fig. 1A). Let  $i = (i_1, i_2, i_3)$  denote the multiindex labeling the  $i$ th voxel of the mesh. (Each two-dimensional image is assumed to lie in a plane perpendicular to the  $z$  axis, with each intensity value of a pixel corresponding to the intensity value of a three-dimensional voxel centered in  $z$  on the pixel plane.) Based on the data in ref. 1, the voxels were assumed to have spatial dimensions of approximately .0397 by .0397 by .125  $\mu\text{m}$ . We subsequently label these dimensions by  $h = (h_1, h_2, h_3)$ . Denote by  $I_i$  the normalized DAPI stain intensity in the  $i$ th voxel, given by [9]. As the imaging data is defined on the mesh given by the collection of voxels, we work with a spatially discrete reaction-diffusion master equation (RDME) model for the motion of the protein (instead of the spatially continuous Fokker–Planck equation [1]).

Using the Cubes software program (12) we calculated the intersection of the mesh of imaging voxels with the nuclear membrane surface. (See *Materials and Methods* for details on the resulting embedded boundary mesh.) Denote by  $x_i$  the centroid and by  $V_i$  the volume of the portion of the  $i$ th voxel located within the nucleus. We let  $A_{ij}$  represent the area of the portion of the face separating voxels  $i$  and  $j$  that is within the nucleus. Finally, we define  $P(i,t|i_b)$  to be the probability that the regulatory protein has not yet found its binding site and is within the  $i$ th voxel at time  $t$ , given that the binding site,  $x_b$ , is located within voxel  $i_b$ . To obtain an RDME for the time evolution of  $P(i,t|i_b)$  we combine the finite volume discretization method of ref. 13 for obtaining RDMEs with pure-diffusive motion, with a discretization method similar to that of ref. 14 for discretizing Fokker–Planck equations (see *SI Text* for details). The method described in *SI Text* is used to discretize the spatial fluxes associated with [1], and the resulting expressions are used in the method of ref. 13 to derive the transition rates within the RDME.

We assume that within the voxel,  $i_b$ , a binding reaction may occur with bimolecular reaction rate  $k$  (having units of  $\mu\text{m}^3/\text{s}$ ). In our actual numerical simulations we take  $k = \infty$  so that the protein binds instantaneously upon reaching the voxel,  $i_b$ . The “binding times” we subsequently study then represent the time for the protein to first find a small region (with the size of a voxel). This would (approximately) correspond to choosing  $r_b$  in [2] to define a spherical binding site with the same volume as a voxel. If we had instead assumed the binding site is substantially smaller than the size of a voxel, we have previously shown (15–17) that the choice  $k = 4\pi D r_b$  makes the solution to the RDME an asymptotic approximation in  $r_b$  of  $p(x,t|x_b)$ .



**Fig. 1.** (A) DAPI intensity field reconstructed from data in ref. 1. Note, the volume rendering of the field is partially transparent to allow the viewer to see through it. This effect causes the field to appear sparser than it is in actuality. [Movie S2](#) shows a rotating view of the volume rendering. Axis units are in  $\mu\text{m}$ . (B) Surface triangulation of the nuclear membrane. Reconstructed from the fluorescent nuclear pore imaging data of ref. 1 as described in [Materials and Methods](#). [Movie S1](#) shows a rotating view of the surface mesh. Axis units are in  $\mu\text{m}$ . The surface triangulation contains approximately 5,000 vertices. Of these, approximately 2,000 are the sites of nuclear pores, the other approximately 3,000 are the result of the triangulation process. The pores are too small to resolve as holes, in the scale of the figure, but the locations of the pores are indicated by small red spheres. (C) Histogram of the probability distribution of intensity values within voxels. Each bar height gives the probability that intensity values fall between the intensity values at the bar's edges. Note the edges are separated by 0.025 intensity units for each bar. The left y-axis corresponds to the bar heights. A graph of the cumulative distribution function (cdf) for the normalized DAPI fluorescence intensity within voxels is overlaid on the histogram. Let  $I$  denote a value of the normalized fluorescence intensity and  $\bar{I}$  the random variable for the intensity within an arbitrary voxel. The cdf is the  $\text{Prob}[\bar{I} \leq I]$ . Circles denote every tenth percentile. In our study, binding sites are selected based on these percentiles, as explained in the text. The right y-axis gives the values of the cdf.

The final discretized RDME model we obtain from [1] is then

$$\frac{dP}{dt}(i,t|i_b) = \sum_j [\alpha_{ij}P(j,t|i_b) - \alpha_{ji}P(i,t|i_b)] - \frac{k}{V_{i_b}}\delta_{i,i_b}P(i_b,t|i_b), \quad [4]$$

where the jump rates,  $\alpha_{ij}$ , give the probability per unit time of the protein hopping from the  $j$ th voxel to the  $i$ th voxel when the protein is within the  $j$ th voxel. These hopping rates incorporate both diffusion and drift due to the potential,  $U(\mathbf{x})$ . Note that the RDME [4] is a coupled system of ordinary differential equations (ODEs), with one ODE for each voxel location,  $i$ .  $\alpha_{ij}$  will be zero unless voxels  $i$  and  $j$  are direct neighbors in the  $d$ th direction ( $d = 1, 2, \text{ or } 3$  corresponding to the  $x, y, \text{ and } z$  directions). The jump rates for voxels that are neighbors along the  $d$ th coordinate, as determined by our discretization procedure, are

$$\alpha_{ij} = \frac{2DA_{ij}}{h_d V_j} \frac{1}{\exp((U(x_i) - U(x_j))/k_B T) + 1}. \quad [5]$$

For voxels that are not cut by the nuclear membrane this expression reduces to

$$\alpha_{ij} = \frac{2D}{h_d^2} \frac{1}{\exp((U(x_i) - U(x_j))/k_B T) + 1}. \quad [6]$$

Note that when the potential at the neighbor,  $i$ , is substantially larger than the potential at the current location,  $j$ , then the hopping rate from  $j$  into  $i$  approaches zero. It is more difficult to hop to voxels with higher potential values.

The volume exclusion potential,  $U(\mathbf{x})$ , was chosen based on the intensity of the DAPI stain within the  $i$ th voxel. We assumed that regions of higher DAPI stain intensity should correspond to regions with a higher density of chromatin, and hence be more difficult for a regulatory protein to move into. Several functional relationships between the intensity field and the volume exclusion potential were tried, however, for the remainder we assume a linear scaling,

$$U(x_i) = \bar{U}I_i. \quad [7]$$

Here the scaling constant,  $\bar{U}$ , is considered a model parameter that determines the maximum “repulsiveness” of the potential. We subsequently refer to  $\bar{U}$  as the “volume exclusivity” of the chromatin. When  $\bar{U}$  is zero the regulatory protein will simply diffuse within the nucleus as if the nuclear volume were empty. In contrast, when  $\bar{U}$  is

large it will be very difficult for the protein to move into regions of high DAPI stain intensity.

Regulatory proteins begin their search process after entering the nucleus through a nuclear pore. Instead of restricting the initial location of the protein to a specific pore or collection of pores, in the model the initial position of the protein was chosen from a uniform distribution among all the pore locations. The probability the regulatory protein was initially in voxel  $i$  was therefore

$$P(i,0) = \frac{\text{number of nuclear pores in voxel } i}{\text{total number of nuclear pores}}. \quad [8]$$

To study the effect of varying binding site position within different regions of the nucleus, we also allowed the binding site location,  $i_b$ , to be a random variable. The set of voxels in which the binding site could be placed were determined by specifying an allowable range of intensity values. For each simulation the voxel representing the binding site was then chosen from a uniform distribution over all voxels having intensity values within the given range. We specified two percentiles of the intensity value distribution, shown in Fig. 1C, to determine an interval of allowable intensities. When choosing lower percentiles, the binding site was prevented from being placed in voxels of very high DAPI stain intensity. This was used to model that such voxels may not contain active binding sites (for example, because these voxels may contain silenced heterochromatin). Similarly, using intervals that only contained nonzero intensity values was used to model that regions of zero, measured, DAPI stain intensity may not actually contain DNA (and hence have no active binding sites).

### Numerical Implementation

One method to study the time required for a regulatory protein to find a specific binding site would be to solve numerically the system of ODEs given by the RDME [4]. By allowing  $i_b$  to be a random variable, the term  $\delta_{i,i_b}$  causes [4] to contain a random coefficient. We would therefore need to solve numerically [4] for many choices of  $i_b$  sampled from within the range of allowable intensity values. An alternative approach is to simulate instead the stochastic process described by the RDME [4]. This process models the regulatory protein as undergoing a continuous time random walk between voxels, with hopping rates between voxels given by  $\alpha_{ij}$ . When located in the voxel with the binding site,  $i_b$ , the protein may also bind with probability per unit time,  $k/V_{i_b}$ . As we took  $k = \infty$  for the simulations of the next section, this

corresponded to ending the simulation once the protein hops into the voxel,  $i_b$ . Realizations of the stochastic process described by the RDME can be created through the use of the Gillespie method (18) [also known as kinetic Monte Carlo (19)]. With the exception of floating point error in arithmetic operations and the error induced through the use of pseudorandom number generators, the Gillespie method is exact in simulating this stochastic process.

Using the Gillespie method, our numerical simulation algorithm can be summarized as follows

1. Precalculate the jump rates  $\alpha_{ij}$ .
2. Sample the binding site location,  $i_b$ , from a uniform distribution among the voxels within the specified intensity range.
3. Sample the initial position for the protein from  $P(i,0)$ .
4. Use the Gillespie method to simulate the motion of the protein until the time,  $T$ , it binds (i.e., first hops into voxel  $i_b$ ).
5. Repeat from step 2 until the desired number of simulations have been run.

For certain parameter choices, in a small fraction of simulations, always less than  $10^{-4}$ , we observed that the protein could take such a long time to find the binding site that for practical purposes the binding site was never found. We believe that this behavior arose for one or more of the following reasons: (i) pure chance, as it is always possible in a stochastic simulation for anything to happen (or in this case fail to happen); (ii) the binding site being chosen in a voxel that is cut off by voxels of high potential from the nuclear pore where the search is chosen to begin; and/or (iii) the trapping of the diffusing protein within regions of high potential that it may happen to enter and have difficulty in exiting. These individual simulations would take several orders of magnitude longer computing time than those in which the protein bound on physically relevant timescales. To avoid computational slow downs, we stopped any simulation where the protein had not bound by a prespecified time,  $t = T_{\max}$ . We generally chose  $T_{\max}$  to be  $10^7$  seconds. In those parameter regimes where a small subfraction of simulations were stopped before binding, our observed values for  $T$  represented censored data. Because estimation of the sample median and its standard error is unaffected when the number of censored samples is small, we use the median binding time rather than the mean as an overall measure of the time required to find a binding site. Ninety-five percent confidence intervals of the median were estimated using the  $PB_2$  estimator for the variance of the sample median (20). Survival distribution functions for the probability the binding time random variable,  $T$ , is greater than  $t$ ,  $\text{Prob}[T > t]$ , were estimated using MATLAB's `ecdf` routine. Associated 95% confidence intervals were estimated with the same routine.

## Results

We now study how the random variable,  $T$ , for the time at which the protein first binds to the binding site varies as a function of the maximal potential strength and binding site location. The bimolecular reaction-rate,  $k$ , is chosen to be infinite, so that the binding reaction occurs instantaneously upon the protein entering the voxel containing the binding site. This assumption effectively chooses the binding site to be the size of one voxel. Note the model does not account for the kinetics of binding once near the site, or for secondary effects (such as whether the binding site is in an open, binding accessible state (21, 22).

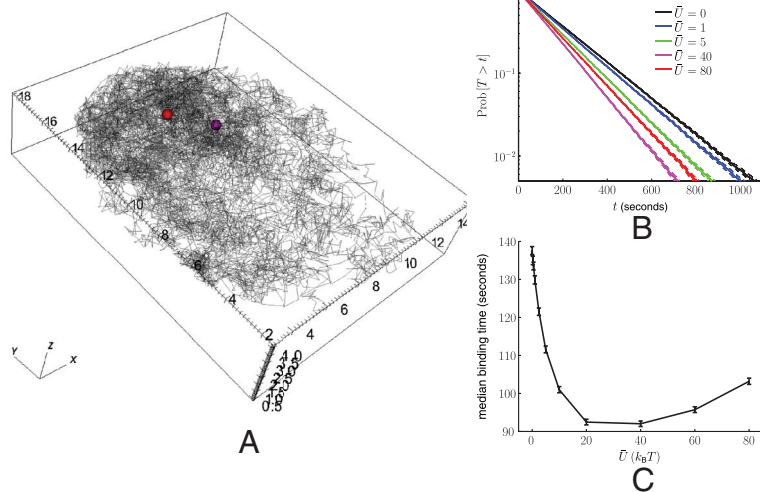
An example illustrating the effect of volume exclusion for a specific initial protein position and binding site location when  $D = 10 \mu\text{m}^2/\text{s}$  is shown in Fig. 2. The specific pore shown in red in Fig. 2A and binding site shown in purple, were used as the initial and binding site positions for all simulations. The continuous time random walk of the protein within the potential [7] was then simulated as described in the previous section. An individual simulation completed when the protein first reached the voxel

representing the binding site (and the time,  $T$ , at which this occurred was recorded), or was terminated if the the binding time exceeded a maximum time,  $T_{\max} = 10^7$ . Fig. 2A shows a typical trajectory of one protein undergoing the search process (Movie S3 shows the motion of a protein within the volume exclusion potential field).

Fig. 2C shows the median time to find the binding site as the magnitude of the volume exclusivity of the chromatin is increased. The graph illustrates that the median binding time decreases as the volume exclusivity is increased, until a minimum binding time is reached. As the volume exclusivity is further increased the median binding time then increases. The same effect is visible in the survival time curves,  $\text{Prob}[T > t]$ , shown in Fig. 2B. Note these curves appear well-approximated by an exponential process (as the curves are linear when the y-axis uses a logarithmic scale). For  $\bar{U} = 0$  the rate constant of the exponential is approximately  $.005 \text{ s}^{-1}$ . Using that the volume of the nucleus is  $528 \mu\text{m}^3$ , the binding process can be approximated by a well-mixed reaction with an association rate of  $2.63 \mu\text{m}^3 \text{ s}^{-1}$ . Given the dependence of this rate on diffusion constant, binding site size, nuclear geometry, and many other factors, this prediction compares favorably with other experimentally determined in vivo association rates. For example, in ref. 23 the binding of glucocorticoid receptor to a tandem array of mouse mammary tumor virus promoter sites within mouse adenocarcinoma cells was studied by fluorescence recovery after photobleaching (FRAP). The authors' analysis of the FRAP measurements predicts a lower bound on the association rate to a single binding site within the array of approximately  $.1 \mu\text{m}^3 \text{ s}^{-1}$ . An effective diffusion constant for glucocorticoid receptor of  $1.2 \mu\text{m}^2 \text{ s}^{-1}$  was also predicted from the experimental data. If we assume the association rate is diffusion limited, and hence proportional to the diffusion constant (3), then rescaling the association rate for a diffusion constant of  $10 \mu\text{m}^2 \text{ s}^{-1}$  gives a lower bound for the association rate of  $.83 \mu\text{m}^3 \text{ s}^{-1}$ .

Fig. 3 shows the statistics of the binding time as a function of both the binding site position and the volume exclusivity,  $\bar{U}$ . For each simulation one nuclear pore was chosen randomly from a uniform distribution among all pores, and the protein was initially placed in the voxel containing that pore. Likewise, for each individual simulation the binding site position was chosen from a uniform distribution over all voxels in a given intensity range. For example, "20 to 30" specified that the voxel representing the binding site should be sampled from those voxels with intensity values between the twentieth and thirtieth percentiles of the intensity value distribution (shown in Fig. 1C). The labels "40 to 50" and "70 to 80" were defined similarly. For each intensity range the binding site was localized within, and each value of  $\bar{U}$ , 128,000 simulations were run.

Fig. 3A shows several survival time distributions,  $\text{Prob}[T > t]$ , as  $\bar{U}$  is varied from zero to  $80k_B T$ . For these simulations, the binding site position was sampled from those voxels with DAPI stain intensities within the twentieth to thirtieth percentile of intensity values. As the volume exclusivity is increased from zero the binding time distribution initially shifts down and to the left (so that the binding site is found more quickly). Note, however, as the volume exclusivity is further increased the distributions ultimately shift back upward and to the right (so that the binding site is found more slowly). In contrast to Fig. 2B, where both the binding site and initial position were fixed, the survival time curves no longer appear well-approximated by an exponential distribution. This difference arises because the binding site position is no longer fixed as in Fig. 2B, but is instead chosen from a probability distribution. The survival time distributions shown in Fig. 3A are now averages of the survival time distribution for each fixed binding site over all possible binding sites (namely those within voxels in the 20th to 30th percentiles of intensity values).



**Fig. 2.** Binding time statistics for a specific fixed initial position and specific fixed binding site when  $D = 10 \mu\text{m}^2/\text{s}$ . (A) A typical path of the protein's random walk. The red sphere is drawn centered about the centroid of the voxel containing the pore from which the protein began its search. Similarly, the purple sphere is drawn centered about the centroid of the voxel that represented the binding site. The size of the spheres is purely for illustrative purposes. Note the path is a piecewise linear curve connecting the centroids of the voxels in which the protein was located approximately every one hundredth of a second. Between any two points of the path the protein actually underwent many hops between voxels. For this simulation the maximum of the potential was chosen to be  $40 k_B T$ . [Movie S3](#) shows the motion of a protein during the search process. (B) For the binding site position and initial position shown in (A) the probability the protein has not bound at time  $t$  as  $\bar{U}$  is varied. Note that each line corresponds to statistics determined from 128,000 simulations. The dashed lines above and below each solid line correspond to 95% confidence intervals. Note that the y-axis is a logarithmic scale, showing that the survival probability is approximately exponential in time. (C) Median binding time with 95% confidence intervals as  $\bar{U}$  is varied. For each data point 128,000 simulations were run. As the volume exclusivity is increased from zero (no volume exclusion), a minimum median binding time is reached. Beyond about  $\bar{U} = 40 k_B T$  the median binding time increases.

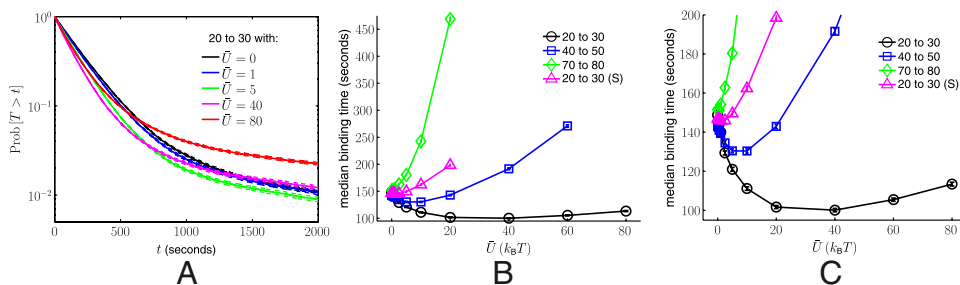
Fig. 3 *B* and *C* show the median binding time as a function of the volume exclusivity. Different curves correspond to different choices of the intensity range the binding site position was sampled from. Fig. 3*C* is the same graph as Fig. 3*B*, but with the y-axis rescaled to better show the curves that decrease in value. When the binding site is restricted to the twentieth to thirtieth or fortieth to fiftieth percentiles of intensity values, the protein initially finds the binding site more quickly as the volume exclusivity is increased. That is, the presence of volume exclusion by chromatin helps decrease the time needed for the protein to find the binding site (vs. when there is no volume exclusivity; i.e.  $\bar{U} = 0 k_B T$ ). As  $\bar{U}$  is further increased the median binding time ultimately begins to increase. For binding sites within regions of sufficiently low to moderate, but nonzero, DAPI stain intensities we therefore find a minimum binding time for a nonzero value of the volume exclusivity. This indicates that volume exclusion by chromatin may help to speed up the search process of regulatory proteins for active binding sites.

In contrast, when the binding site is restricted to regions of sufficiently high fluorescence intensity, the 70 to 80 curve, this effect is lost. As  $\bar{U}$  is increased the time to find the binding site dramatically increases. This result arises because the regions in which the binding sites are now located become substantially

more difficult to enter as the chromatin is made more volume excluding.

We next examined whether the spatial structure of the chromatin density was important in the observed decrease in median binding time as a function of the volume exclusivity. To test this question we randomly permuted the fluorescence intensity values among the voxels within the nucleus. This procedure preserved the distribution of intensity values, shown in Fig. 1*C*, while eliminating the underlying spatial structure of the intensity field. The field was ‘‘shuffled’’ once, and the statistics of the binding time within this new field, and the potential fields associated with it, were then studied. The 20 to 30 (S) curves in Fig. 3 *B* and *C* show the effect of the shuffling procedure on the binding time. For these simulations the binding site position was sampled from voxels with intensity values in the new intensity field between the twentieth to thirtieth percentile of all values. Note that within statistical error the binding time simply increases as the volume exclusivity is increased. That is, the previously observed decrease in binding time as a function of volume exclusivity depends on the spatial structure and correlations within the intensity field (and not just on the distribution of intensity values).

The reduction in binding time as the volume exclusivity initially increases from zero, as shown in Figs. 2*C* and 3 *B* and *C*, is the



**Fig. 3.** Statistics of the binding time as a function of the maximum value of the potential,  $\bar{U}$ , and binding site position. For each curve  $D = 10 \mu\text{m}^2/\text{s}$ . In all graphs, 20 to 30 denotes that for each simulation the binding site location was sampled from a uniform distribution among voxels with intensity values in the twentieth to thirtieth percentile of all intensity values. The labels 40 to 50 and 70 to 80 correspond to the fortieth to fiftieth percentiles and seventieth to eightieth percentiles respectively. An (S) at the end of a label, such as 20 to 30 (S), denotes that the intensity values were randomly shuffled among the voxels. That is, the values of  $I_i$  were randomly rearranged between the voxels within the nucleus. The volume exclusion potential was then generated from this new intensity field, and both were then used in simulations where the protein binding site was restricted to voxels within the twentieth to thirtieth percentiles of intensity values. In all graphs, for each value of  $\bar{U}$  128,000 simulations were run. (A) Survival time distribution,  $\text{Prob}[T > t]$ , for the binding time random variable,  $T$ , when using the 20 to 30 binding site distribution. The dashed lines above and below each curve indicate 95% confidence intervals. (B) The median binding times vs.  $\bar{U}$ . (C) The same graph, but with the y-axis rescaled to better show the curves that initially decrease to a minimum. Error bars corresponding to 95 percent confidence intervals are shown in (B) and (C), however, they are smaller than the marker size. Each graph shows that when the binding site is localized to regions of low, but nonzero intensity, the binding time initially decreases to a minimum and then increases as the volume exclusivity,  $\bar{U}$ , is increased. Note when the DAPI stain intensity values are shuffled in space, or the binding site is moved to regions of higher intensity values, this effect is lost.

principal result of this paper. The abolition of this effect by shuffling the voxel intensity values (see curves labeled “S” in Fig. 3 B and C) shows that this reduction in binding time is not merely a consequence of a reduction in the total volume that needs to be searched, but in fact depends on the spatial structure of chromatin, and indeed on some aspect of the spatial structure on a scale that can be resolved with the voxel size used here. Although our results do not reveal specifically what feature of the spatial structure of chromatin is responsible for this effect, we may speculate that the chromatin geometry partitions the three-dimensional intranuclear space into something like a network of one-dimensional channels. This would be significant because of the disparity between the time to find a target by diffusion in the one-dimensional case in comparison to the substantially greater amount of time that is required in higher dimensions (see *SI Text*). We emphasize, however, the speculative nature of this interpretation of our computational results.

## Conclusions

We developed a mathematical model to study how the time required for a protein to find a specific binding site varies as a function of volume exclusion by chromatin and binding site location. The model suggests that binding sites located within regions of small, but nonzero, DAPI stain intensity are found most quickly when chromatin has some, but not too much, volume exclusivity. Randomly shuffling the DAPI stain intensity values among the voxels within the nucleus caused this behavior to disappear. This suggests the macroscopic distribution of chromatin density within the nucleus may be arranged to help funnel proteins toward binding sites within regions of euchromatin (where we expect the chromatin density to be lower). In contrast, binding sites located within regions of high DAPI stain intensity are found most quickly when no volume exclusion is modeled. Because such regions are more likely to contain silenced heterochromatin, it may be less important to optimize chromatin distribution to make them accessible to binding proteins.

## Materials and Methods

As input, our model makes use of the structured illumination microscopy data of ref. 1, specifically supplementary movie 1 of ref. 1. From the movie, we were able to reconstruct a nuclear membrane surface and an intensity field for the DAPI stained DNA fluorescence within an individual mouse C2C12 cell nucleus. The movie was split into a collection of images, each corresponding to a slice plane perpendicular to the z-axis of the cell nucleus. Two types of fluorescent data were present in each frame: green “dots” from fluorescently labeled nuclear pores, and a magenta field representing DAPI stained DNA (chromatin).

- Schermelleh L, et al. (2008) Subdiffraction multicolor imaging of the nuclear periphery with 3D structured illumination microscopy. *Science* 320:1332–1336.
- Halford S (2009) An end to 40 years of mistakes in DNA-protein association kinetics? *Biochem Soc T* 37:343–348.
- Smoluchowski MV (1917) Mathematical theory of the kinetics of the coagulation of colloidal solutions. *Z Phys Chem* 92:129–168.
- Berg OG, Winter RB, von Hippel PH (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* 20:6929–6948.
- Li GW, Berg OG, Elf J (2009) Effects of macromolecular crowding and DNA looping on gene regulation kinetics. *Nat Phys* 5:294–297.
- Malherbe G, Holcman D (2007) The search kinetics of a target inside the cell nucleus. arXiv:0712.3467v1 [q-bio.BM].
- Slutsky M, Mirny LA (2004) Kinetics of protein–DNA interaction: Facilitated target location in sequence-dependent potential. *Biophys J* 87:4021–4035.
- Mirny L, et al. (2009) How a protein searches for its site on DNA: The mechanism of facilitated diffusion. *J Phys A-Math Theor* 42:1–23.
- Elf J, Li G, Xie XS (2007) Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* 316:1191–4.
- Bancaud A, et al. (2009) Molecular crowding affects diffusion and binding of nuclear proteins in heterochromatin and reveals the fractal organization of chromatin. *EMBO J* 28:3785–3798.
- Vargas DY, Raj A, Marras SAE, Kramer FR, Tyagi S (2005) Mechanism of mRNA transport in the nucleus. *Proc Natl Acad Sci USA* 102:17008–13.
- Berger MJ *Cubes AMR Embedded Boundary Meshing Program*.
- Isaacson SA, Peskin CS (2006) Incorporating diffusion in complex geometries into stochastic chemical kinetics simulations. *SIAM J Sci Comput* 28:47–74.
- Wang H, Peskin CS, Elston TC (2003) A robust numerical algorithm for studying biomolecular transport processes. *J Theor Biol* 221:491–511.

**Nuclear Membrane Reconstruction.** The MATLAB Imaging Toolbox was used to segment and then find the centroid of each nuclear pore. This generated a noisy point cloud representing the locations of the nuclear pores within the cell. Because (functional) nuclear pores are localized to the nuclear membrane, and are generally well distributed over the surface of the membrane and present in large numbers (order of thousands), the point cloud gave a good approximation to the membrane’s location. After removing extreme outliers, most likely fluorescent components of the nuclear pore that were present in other locations within the cell, the implementation of the eigencrust algorithm of ref. 24 distributed in ref. 25 was used to generate a watertight triangulated surface from the point cloud. The resulting surface was cleaned and postprocessed in MeshLab (26). Fig. 1B and *Movie S1* show the final triangulated nuclear membrane surface. The mesh is comprised of approximately 5,000 nodes and 9,700 triangles.

**DAPI DNA Intensity Field.** Let  $i = (i_1, i_2, i_3)$  label the three-dimensional voxel corresponding to pixel  $(i_1, i_2)$  in the  $i_3$ th image frame. The pixel plane is assumed to be centered within the voxel, perpendicular to the z axis. A three-dimensional DNA fluorescence intensity field,  $I_i$ , was created from the DAPI stain intensity values for each pixel. The intensity field was normalized to have values in  $[0, 1]$  by setting

$$I_i \leftarrow \frac{I_i}{\max_i I_i}. \quad [9]$$

Fig. 1A and *Movie S2* show a volume rendering of the resulting discrete DNA fluorescence intensity field,  $I_i$ . Note, the rendering is semitransparent to allow the viewer to see within the field. This causes the field to appear to be “clumpier” than it actually is, see the original imaging data, supplemental movie 1 of ref. 1.

**Embedded Boundary Cartesian Mesh Derived From Imaging Data.** The three-dimensional voxels centered about the pixels of the image slices define a natural Cartesian mesh. The discretization procedure of the *Mathematical Model* section requires the calculation of the portion of each voxel that is within the nucleus. Using the *Cubes (12)* software program the intersection of the reconstructed nuclear membrane surface with this Cartesian mesh was calculated.

**ACKNOWLEDGMENTS.** We thank Ravi Iyengar for suggesting the potential randomization procedure, the referees for their helpful suggestions, and Marsha Berger for providing access to, and help in using, the *Cubes (12)* meshing software. All three authors are supported by the Systems Biology Center New York (National Institutes of Health Grant P50GM071558). S.A.I. is also supported by National Science Foundation Grant DMS-0920886. All three-dimensional figures and movies were produced with the *Visit* visualization software (27). The numerical simulations presented in this manuscript made use of the Computational Center for Nanotechnology AMD Opteron cluster at Rensselaer Polytechnic Institute and the Scientific Computing and Visualization Katana cluster at Boston University.

- Isaacson SA (2009) The reaction-diffusion master equation as an asymptotic approximation of diffusion to a small target. *SIAM J Appl Math* 70:77–111.
- Isaacson SA, Isaacson D (2009) Reaction-diffusion master equation, diffusion-limited reactions, and singular potentials. *Phys Rev E* 80:1–9.
- Isaacson SA (2008) Relationship between the reaction-diffusion master equation and particle tracking models. *J Phys A-Math Theor* 41:1–15.
- Gillespie DT (1977) Exact stochastic simulation of coupled chemical-reactions. *J Phys Chem* 81:2340–2361.
- Bortz AB, Kalos MH, Lebowitz JL (1975) A new algorithm for Monte Carlo simulation of Ising spin systems. *J Comp Phys* 17:10–18.
- Price RM, Bonett DG (2001) Estimating the variance of the sample median. *J Stat Comput Sim* 68:295–305.
- Li G, Levitus M, Bustamante C, Widom J (2004) Rapid spontaneous accessibility of nucleosomal DNA. *Nat Struct Mol Biol* 12:46–53.
- Poirier MG, Bussiek M, Langowski J, Widom J (2008) Spontaneous access to DNA target sites in folded chromatin fibers. *J Mol Biol* 379:772–786.
- Sprague BL, et al. (2006) Analysis of binding at a single spatially localized cluster of binding sites by fluorescence recovery after photobleaching. *Biophys J* 91:1169–1191.
- Kolluri R, Shewchuk J, O’Brien J (2004) Spectral surface reconstruction from noisy point clouds. *SGP ’04: Proceedings of the 2004 Eurographics/ACM SIGGRAPH Symposium on Geometry processing* pp 11–21.
- Kolluri R *Eigencrust Software* Available in the MATLAB PointCloudToolbox at <http://bdtnc.lbl.gov/Fly-Net/bioimaging.jsp?v=pcml>.
- 3D CoForm Project *MeshLab* Available from <http://meshlab.sourceforge.net/>.
- Lawrence Livermore National Lab *Visit Visualization Software* Available at <https://wci.llnl.gov/codes/visit/home.html>.