**SUPPORTING MATERIAL**

Consider a circular DNA molecule consisting of $N$ base pairs, and assume that $B$ "car" molecules of equal length $n$ are parked at random on it. Thus each of the $N$ positions in which any car can be parked, is equally probable for that car, except the cars do not overlap. So the first car can be parked in $N$ different positions, the next car in $N - (2n - 1)$ different positions and so forth. Let $M_{per}(B, N)$ denote the number of different ways $B$ cars of length $n$ can be parked on a loop of length $N$. The letter $M$ stands for *multiplicity* and the subscript *per* for *periodic,* as in periodic boundary condition of a lattice, the lattice of N sites being the correct mathematical term for the way we think of the circular DNA molecule in the present context. The boundary condition is of importance and the periodic one is the simplest one. We shall see it admits a closed mathematical result that cannot be obtained with the boundary conditions for a linear DNA molecule with ends, except when $n = 1$. The linear molecule will be considered later.

For parking on a periodic lattice as just described, we may ask what is the probability $p_{per}(g; B, N)$ that $g$ given, consecutive lattice sites are unoccupied by cars? The answer is that it equals the probability that all of the $B$ cars are parked on the complement to those $g$ lattice sites, i.e., on a lattice with $N$-$g$ sites and two ends that are *closed* in the sense that the cars must be parked entirely within the $N$-$g$ sites of this linear lattice. Let $M_{clsd}(B, N - g)$ denote the number of ways this can be done. Because all arrangements of the parked cars are equally probable, the probability that we seek is then equal to the ratio $M_{clsd}(B, N - g) / M_{per}(B, N)$ .

***Evaluation of $M_{clsd}(B, N')$:*** Let $N'$ denote the number of sites in a lattice with a closed boundary condition. Below, we shall set $N' = N$-$g$. For $B$ given parking spaces, $B$ cars can be parked in $B!$ different ways, unless we choose to not distinguish between the cars. If we choose not to distinguish the cars, they obey the statistics of identical objects, and $B$ cars can be parked in one and only one way in $B$ non-overlapping parking spaces: One car in each space. If we exchange two parked cars, this does not count as a different way to park the cars because they are identical. Since molecules are indistinguishable, we choose to treat the cars as identical. Ultimately, this choice does not matter for results because the extra factor $B!$ which occurs if cars are treated as distinguishable, cancels in the ratios formed to obtain probabilities

Focus now on the spaces occupied by the cars. Any given arrangement of them on a lattice with $N'$ sites and a closed boundary condition is entirely defined by giving the sizes of the $(B - 1)$ gaps between the cars, plus the sizes of the two gaps between the two ends of the lattice and the two cars closest to the ends. Enumerating these gaps from one end of the lattice and denoting the gap sizes by $g_1, g_2, \ldots, g_{B+1}$, we have

$$M_{clsd}(B,N') = \sum_{g_1,\ldots,g_{B+1}=0}^{N'-nB} \delta_{\left(g_1+\ldots+g_{B+1},\, N'-nB\right)} \qquad \text{(Equation S1)}$$

where the Kronecker delta-function ensures that only gaps satisfying $g_1 + \ldots + g_{B+1} = N' - nB$ contribute to the sum, and all combinations of gaps that satisfy this condition, contribute equally, with one count. (Kronecker's delta-function is a function of two integers, say $j$ and $k$, with value $\delta_{j,k} = 1$ for $j=k$, and otherwise value zero.) This sum can be computed in a straightforward manner by replacing the delta function on the integers with its Fourier transform on the interval $[0, 2\pi]$, and applying Cauchy's Contour Integration Theorem to the resulting integral. The sum can also be done by observing that it equals the number of different ways that $N'-nB$ identical tokens (unoccupied lattice sites) and $B$ other identical tokens (parking spaces) can be ordered sequentially, when one does not distinguish between different tokens from the same category. The answer is simply a binomial coefficient,

$$M_{clsd}(B,N) = \binom{N'-nB+B}{B} = \frac{(N'-nB+B)!}{B!(N'-nB)!}. \qquad \text{(Equation S2)}$$

***Evaluation of $M_{per}(B, N)$:*** To calculate $M_{per}(B, N)$, pick *one* particular lattice site and focus on it. Any one will do, as they are all equivalent. This site is either covered by a parked car, or it is not. If covered by a car, it is either covered by the left-most unit of the car, or by the next-to-left-most unit of the car, etc. So there are $n$ mutually exclusive ways for a car to cover this lattice site. For each one of these ways, the other $(B - 1)$ cars are parked on the other $(N-n)$ sites of the lattice. Hence, there are $nM_{clsd}(B - 1, N - n)$ different ways that $B$ cars may be parked in such a manner that a car covers this particular lattice site. Add to this number the number of ways that the same $B$ cars may be parked *without* covering this particular lattice site, $M_{clsd}(B, N - 1)$. Then we have enumerated all mutually exclusive ways, in which $B$ cars can be parked on a periodic lattice of $N$ sites,

$$M_{per}(B, N) = nM_{clsd}(B-1, N-n) + M_{clsd}(B, N-1). \qquad \text{(Equation S3)}$$

***The exact probability:*** A particular set of $g$ consecutive lattice sites is then left uncovered by cars with probability

$$p_{per}(g;B,N)=\frac{M_{clsd}(B,N-g)}{M_{per}(B,N)}=\frac{M_{clsd}(B,N-g)}{M_{clsd}(B,N-1)+nM_{clsd}(B-1,N-n)}=\frac{(N-nB)!(N-(n-1)B-g)!}{(N-nB-g)!(N-(n-1)B-1)!N}$$

(Equation S4)

This result is exact.

***Asymptotic expression for large lattices:*** Typically, $N$ will be of order 1000 or greater, $n \leq 10$, and $B$ will at most be of order $N$, ($B = N/n$ means saturation of the lattice with cars, and that is practically impossible, the probability for it to happen is vanishingly small). Thus $N - nB >> 1$, and consequently the four factorials in the last expression in Equation (S4) can be approximated exceedingly well with Stirling's formula,

$$K!=\sqrt{2\pi}K^{(K+1/2)}e^{-K}\left(1+1\big/12K+O\left(1\big/K^2\right)\right), \qquad \text{(Equation S5)}$$

since all four factorials are taken of numbers of order $(N - nB)$ or larger. The term $1/(12K)$ in Stirling's formula correctly describes the error committed by leaving out this and higher-order terms in the formula even for $K = 1, 2, \ldots$. We use the formula only for larger values of $K$, so we leave out this term and still have an extremely good approximation,

$$\frac{(N-nB)!}{(N-nB-g)!}=(N-nB)^g\exp\left[-g-(N-nB-g+1/2)\ln\left(1-\frac{g}{N-nB}\right)\right] \qquad \text{(Equation S6)}$$

If furthermore $g/(N - nB) << 1$, one can Taylor-expand the logarithm with respect to this small quantity, and keep only the first term, or the first two terms, and maintain a very good approximation. Doing the latter, i.e., expanding to order $g^2/(N - nB)$ and ignoring terms of order $g^3/(N - nB)^2$, one finds

.

$$\frac{(N-nB)!}{(N-nB-g)!}=N^g(1-nv)^g\exp\left[-\frac{g(g-1)}{2(1-nv)N}\right]. \qquad \text{(Equation S7)}$$

Similarly, using Stirling's formula and expanding to the same order, one finds

$$\frac{(N-(n-1)B-g)!}{(N-(n-1)B-1)!N} = \frac{1}{N^g(1-(n-1)v)^{g-1}}\exp\left[\frac{g(g-1)}{2(1-(n-1)v)N}\right]. \qquad \text{(Equation S8)}$$

Taking the product of these two results, and observing that the term $g(g-1)/N$ in the exponent is only relevant for $g$ of order $N^{1/2} >> 1$, hence $g - 1 \approx g$, one finds from Equation (S4) that

.

$$p_{per}(g;B,N) = \frac{(1-nv)^g}{(1-(n-1)v)^{g-1}}\exp\left[\frac{g^2 v}{2(1-nv)(1-(n-1)v)N}\right]. \qquad \text{(Equation S9)}$$

The approximation involved ignores terms of order $g^3/N^2$ compared to terms of order one, i.e., it ignores terms of order $N^{-1/2}$ or smaller.

This result simplifies considerably if we assume that $nB/N = nv << 1$, as is usually the case because saturation, $nv \approx 1$, is practically impossible. In this case $nv$ is sufficiently small for the following approximations to the natural logarithm, irrespective of the size of $g$:

$$(1-nv)^g = \exp[g\ln(1-nv)] = \exp[-g(nv+1/2(nv)^2+...)] \qquad \text{(Equation S10)}$$

$$(1-(n-1)v)^{g-1} = \exp[(g-1)\ln(1-(n-1)v] = \exp[-(g-1)((n-1)v+1/2((n-1)v)^2+...)] \qquad \text{(Equation S11)}$$

and consequently

$$\frac{(1-nv)^g}{(1-(n-1)v)^{g-1}} \cong \exp\left[-(g+n-1)v-\left(\frac{1}{2}(n-1)^2+g(2n-1)\right)v^2+O(v^3)\right]. \qquad \text{(Equation S12)}$$

Plotting $\ln(p_{per})$ versus $v$, the graph is a straight line with slope $-[g + n - 1]$ at small values of $v$, then bends up at larger values, as we also find experimentally. Figure 6 shows our experimental data with linear fits of slope $-(g + n - 1)$ using $n =$

4, the derived gap sizes, $g$, of 66.4 for ATP hydrolysis and 293 base pairs for DNA cleavage, and equation S12 to first order in $v$.

When the last, exponential factor in Equation (S9) cannot be neglected, it too must be expanded in $v$ for $v$ small. That done, Equation (S9) becomes, for small values of $v$,

$$p_{per}(g;B,N) = \exp\left[-\left(g+n-1+\frac{g^2}{2N}\right)v-\left(\frac{1}{2}(n-1)^2+g(2n-1)+\frac{g^2(2n-1)}{2N}\right)v^2+O(v^3)\right]. \text{ (Equation S13)}$$

Clearly, when the gap-size $g$ is large, the probability that no dye molecule attaches in the gap is non-vanishing only for low concentrations of dye, specifically, for $gv = O(1)$ according to Equations (S12) and (S13). Thus we see that for large $g$, the result simplifies to a simple Gaussian function of $v$, Equation (S12), or even to a simple exponential, if the term quadratic in $v$ can be neglected. But we also see that unless $g^2v/2N << 1$, we need the extension of the result provided by the terms proportional to $g^2/2N$ in Equation (S13). We need these terms even if they are negligible compared to the terms proportional to $g$ that they add to, because they affect the overall magnitude of $p_{per}$ as long as they are not negligible compared to 1.

How good are our approximations? We compare equation 1 (same as equation S9) with equation 3, equation S12 and equation S13. For $1 < g < 100$ and $1 < n < 100$, the difference between equations 1 and 3 is completely negligible for all values of $nB/N$. The exponential forms, equations S12 and S13 are good for $nB/N$ <0.2 and deviate significantly above 0.2. An example comparison of these equations for calculating $p_{per}(g; B, N)$ is given in figure 7 for $N$ = 4000 base pairs, car $n$ = 4 base pairs and a loading bay gap of $g$ = 10 base pairs. Similar results were obtained for other combinations of parameters. We conclude that equation 3 is perfectly satisfactory for the calculation of $p_{per}(g; B, N)$.

***The role of boundary conditions: linear vs. circular DNA:*** For comparison of results, we replace the periodic boundary condition, corresponding to circular DNA, with closed boundary conditions, corresponding to linear DNA, and ask: What is the probability that $g$ particular sites are left unoccupied when $B$ cars of length $n$ are parked at random on a lattice of $N$ sites with closed boundary conditions? The answer depends on where those g sites are located relative to the ends of the lattice. Above, periodic boundary conditions ensured that there were no ends. So let us assume that the distance between those $g$ sites and the nearest end is $N'$, then denote the number of cars parked on those $N'$ lattice sites by $B'$, and denote the sought probability by $p_{clsd}(g; B, N, N')$. Then,

$$p_{clsd}(g;B,N,N') = M_{clsd}(B,N)^{-1} \sum_{B'=0}^{\min(B,\lfloor N'/n \rfloor)} M_{clsd}(B',N') M_{clsd}(B-B',N-N'-g) . \qquad \text{(Equation S14)}$$

This result is exact, but the summation over $B'$ cannot be carried out. The various approximations applied to $p_{per}$ above can also be applied to $p_{clsd}$, and in a large $N$ approximation where the effect of the ends of the DNA is subdominant and neglected, one finds no difference between $p_{per}$ and $p_{clsd}$.

***Poisson-distributed car number B:*** The theory above gives the probability of a gap of size $g$ or larger in the appropriate place on circular DNA for a *given* number $B$ of parked cars. But for a given concentration of cars in solution, the number of cars $B$ on DNA of length $N$ will differ from one DNA molecule to the next, with an expectation value $\langle B \rangle$ that is given by the concentration of dye. This variation in $B$ is of order $\sqrt{\langle B \rangle}$ in unsaturated situations, and smaller near saturation. So it is significant relative to $\langle B \rangle$ when this expectation value is not large. In this situation one can use Equation (S14) because $v$ is very small, and one can say that the relative frequency with which a given $B$-value occurs, is Poisson distributed,

$$P_B\left(\langle B \rangle\right) = c^{-\langle B \rangle} \frac{\langle B \rangle^B}{B!} .$$

For a given value of $\langle B \rangle$, the probability that a gap of length $g$ or longer occurs where required, is then

$$\left\langle p_{per}(g;N) \right\rangle = \sum_{B=0}^{\infty} P_B(\langle B \rangle) p_{per}(g;B,N) . \qquad \text{(Equation S15)}$$

The sum over $B$ can be carried out with the same approximation as used for $p_{per}$ in Equation (S13) by using

$$\left\langle \exp(X) \right\rangle = \exp\left[ \langle X \rangle + \frac{1}{2}\left( \langle X^2 \rangle - \langle X \rangle^2 \right) + O(X^3) \right] \qquad \text{(Equation S16)}$$

which is proven by taking the natural logarithm on both sides and Taylor-expanding the left-hand-side with respect to $X$ about $X = 0$. Thus

$$\left\langle p_{per}(g;N)\right\rangle = \exp\left[-(g+n-1)\frac{\langle B\rangle}{N} - \left((2n-1)g + \frac{1}{2}(n-1)^2 + \frac{(2n-1)g^2}{2N}\right)\frac{\langle B\rangle^2}{N^2}\right]. \quad \text{(Equation S17)}$$

Comparing Equations (S13) and (S17), we see that the latter has no term of order $g^2/2N$ in the coefficient to $\langle B\rangle$/N. This makes $\langle p_{per}(g; B, N)\rangle > p_{per}(g; \langle B\rangle, N)$, i.e., large gaps are made more probable by fluctuations in $B$. This is so because $p_{per}$ in Equation (S13) is a convex function of $B$.

***Activity footprinting:*** The calculated probability, $p_{per}(g; B, N)$ from equation 1 (or S9), is for a gap existing and encompassing the correct location on the DNA lattice to allow binding of the truck. An implicit assumption is that the cars bind essentially irreversibly to the lattice. If they did not, but trucks do bind irreversibly, all loading bays will eventually become occupied by trucks, because cars leave and arrive at random. So for each loading bay there is at all times a finite probability that it is vacant and subsequently is occupied by a truck.

If the gap is greater than or equal to the size of the loading bay, then a truck *can* bind to its specific DNA target sequence. To calculate the frequency with which a truck actually *does* binds to the DNA, one uses the dissociation constant, $K_d$, for the binding affinity of the truck for its target sequence, $K_d = [E]\,[S] / [ES]$ where [E] is the concentration of free enzyme at equilibrium, [S] is the concentration of free target sequences (i.e. the concentration of gaps completely encompassing the loading bay) and [ES] is the concentration of target sequences with the enzyme bound (i.e. the concentration of loading bays with a parked truck). Given that $[E]_{total} = [E]+[ES]$ and $[S]_{total} = [S]+[ES] = p_{per}(g; B, N)[DNA]_{total}$ then one can solve the quadratic equation, Equation S18, for the concentration [ES] in terms of $[E]_{total}$, $[DNA]_{total}$, $K_d$ and $p_{per}(g; B, N)$. We use Equation 3 for $p_{per}(g; B, N)$.

$$p_{per}(g;B,N)[\mathrm{DNA}]_{total}[\mathrm{E}]_{total} - [\mathrm{ES}]([\mathrm{E}]_{total} + p_{per}(g;B,N)[\mathrm{DNA}]_{total} + \mathrm{K_d}) + [\mathrm{ES}]^2 = 0 \quad \text{(Equation S18)}$$

Generally, $[E]_{total}$, $[DNA]_{total}$ and $K_d$ are known for a particular sequence-specific DNA binding system and $p_{per}(g; B, N)$ is related to the experimental variable for saturation of the lattice with non-specific parked cars, $B/N$.

The measured experimental data, enzyme activity in our experiments, is proportional to [ES]. Thus one can fit the experimental measurements of enzyme activity as a function of saturation of the lattice with parked cars and determine a value for the minimum gap size necessary for activity to be observed. In our experiments (31), $[E]_{total}$ is total EcoKI concentration of 67 nM, $[DNA]_{total}$ is total DNA concentration of 50 nM, and the $K_d$ is 2nM (17). [ES] is proportional to the experimental observable; ordinate = (rate constant) / (rate constant in absence of YOYO) = $k$[ES]. Since the protein and DNA concentrations are much larger than $K_d$, then $k$ should be nearly equal to 1 / $[DNA]_{total}$ = 0.02. The results of fitting the experimental data shown in figure 3 with these parameters are given in Table 1.

**Table 1.** Results of fitting experimental data to equation 3.

| Loading bay size, $g$ base pairs, for ATPase activity | Scaling factor, $k$ | Loading bay size, $g$ base pairs, for DNA cleavage activity | Scaling factor, $k$ |
|---|---|---|---|
| 66.4 +/- 11.8 | 0.0208 +/- 0.0010 | 293 +/- 45 | 0.0233 +/- 0.0011 |

**Figure legends for supporting information**

**Figure 6.** This graph shows ln(fractional activity) for ATP hydrolysis (full circles) and DNA cleavage (open circles) against saturation, $B/N = v$. The lines have slopes of $-(g + n - 1)$ and come from using equation S12 to first order in $v$.

**Figure 7.** Theoretical curves calculated using equations S9 (same as equation 1), equation 3, equation S12 and equation S13. The curve for equation 3 overlaps completely with that of equation S9 and is shown as a solid line. The curve using equation S12 superimposes completely with that using equation S13 and is shown as a dashed line. Also shown as a dotted line is the curve using equation S12 to only first order in $v$. In these examples, $N = 4000$, $n = 4$ and $g = 10$.

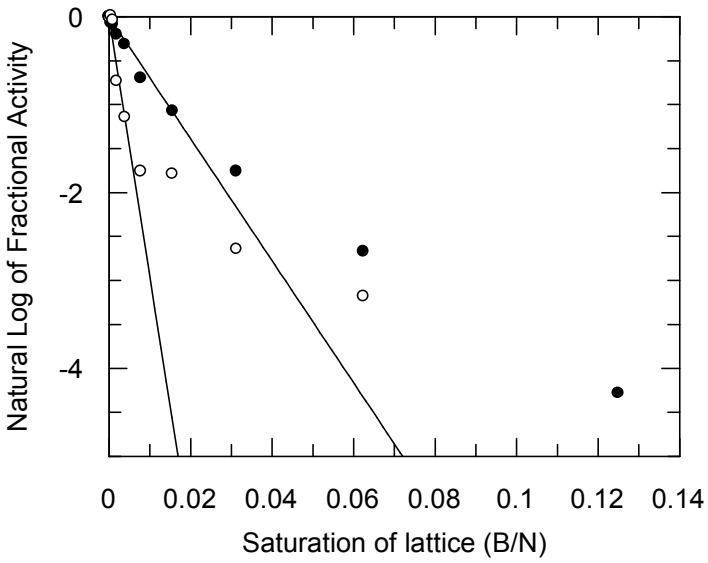
0
-2
-4
Natural Log of Fractional Activity
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
Saturation of lattice (B/N)

**Figure 6**

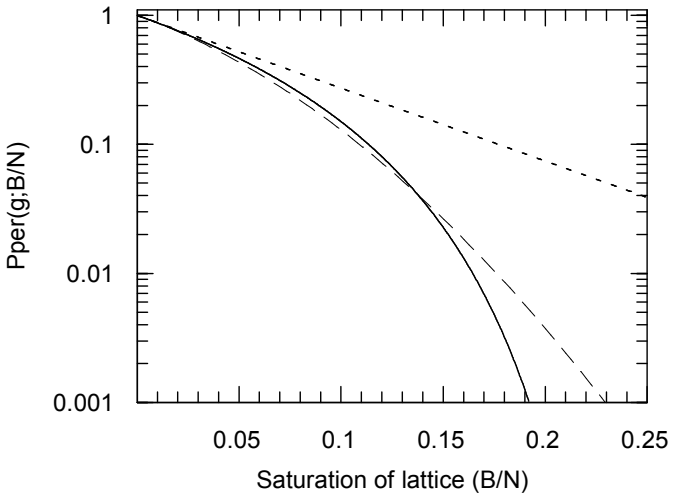
1
0.1
0.01
0.001
Pper(g;B/N)
0.05
0.1
0.15
0.2
0.25
Saturation of lattice (B/N)

**Figure 7**