Kinetic Models of Translocation, Head-On Collision, and DNA Cleavage by Type I Restriction Endonucleases†

Mark D. Szczelkun*

Department of Biochemistry, University of Bristol, Bristol BS8 1TD, U.K.

Received September 21, 2001; Revised Manuscript Received November 21, 2001

ABSTRACT: Digestion of linear DNA by type I restriction endonucleases is generally activated following the head-on collision of two translocating enzymes. However, the resulting distributions of cleavage loci along the DNA vary with different enzymes; in some cases, cleavage is located in a discrete region midway between a pair of recognition sites while in other cases cleavage is broadly distributed and occurs at nearly every intervening locus. Statistical models for DNA translocation, collision, and cleavage are described that can account for these observations and that are generally applicable to other DNA-based motor proteins. If translocation is processive (stepping forward is significantly more likely than DNA dissociation), then the linear distribution of an ensemble of proteins can be described simply using a Poisson relationship. The pattern of cleavage sites resulting from collision between two processive type I enzymes over a distance d can then be described by a binomial distribution with a standard deviation $0.5 \cdot d^{1/2}$. Alternatively, if translocation is nonprocessive (stepping forward or dissociating become equally likely events), the linear distribution is described by a continuum of populated states and is thus extended. Comparisons of model data to the kinetics of DNA translocation and cleavage discount the nonprocessive model. Instead, the observed differences between enzymes are due to asynchronous events that occur upon collision. Therefore, type I restriction enzymes can be described as having processive DNA translocation but, in some cases, nonprocessive DNA cleavage.

Central to many cellular events, translocation is the processive, unidirectional motion of proteins on DNA. It requires chemical energy to be converted into mechanical work, and consequently these proteins can be analyzed as motor enzymes using both classical and single-molecule kinetics (*1*, *2*). Fundamental to these techniques is an appreciation of the statistical distribution functions that describe the probabilities associated with motor enzyme dynamics. Unlike their macroscopic equivalents, microscopic DNA-based motors are not expected to move monotonically. Instead, the spatial and temporal distribution along DNA of an ensemble of proteins will be representative of the underlying distribution of time intervals for each kinetic step. Events that desynchronize motion, such as dissociation, stalling, back-stepping, or collisions with protein "roadblocks", will alter the distribution accordingly. In cases where the motor enzyme must also process the DNA, the downstream kinetics will also reflect the distribution during motion. In this paper, the validity of these statements is examined using kinetic models for DNA translocation and cleavage by type I restriction endonucleases.

The type I restriction-modification enzymes are large oligomeric assemblies (>400 kDa) which carry out both methylation and nuclease reactions within the same complex (*1*, *3*). The endonucleases are encoded by three genes: *hsdS* (responsible for DNA recognition), *hsdM* (DNA methylation), and *hsdR* (DNA and ATP hydrolysis), with a subunit composition $R_2M_2S_1$. They recognize specific asymmetric sequences (e.g., GAANNNNNNRTCG for *Eco*R124I, where N is any base and R is a purine). However, sequence-specific recognition does not lead to cleavage within the site, but to nonspecific cleavage from 50 to >10 000 bp distant from the site (*1*, *3*, *4*). This long-range communication is via one-dimensional DNA translocation (Figure 1A), dependent upon Mg^{2+} ions, ATP, and *S*-adenosylmethionine (*5*, *6*). Initially, an endonuclease associates with both its recognition site (via the HsdS and M subunits) and its adjacent nonspecific DNA (via the HsdR subunits). As ATP is hydrolyzed by the HsdR subunits, two loops of nonspecific DNA are pulled past the enzyme from opposite sides of the asymmetric recognition site (Figure 1A). Translocation rates for each HsdR of ∼400 bp/s (*Eco*R124I in vitro) and ∼100 bp/s (*Eco*KI in vivo) have been measured (*6*, *7*). The HsdR amino acid sequences have homology motifs characteristic of superfamily 2 DNA helicases (*3*). These motifs are essential for ATP hydrolysis, DNA translocation, and subsequent DNA cleavage (*8*, *9*, *10*). However, a helicase-like strand separation activity has never been demonstrated per se (*1*).

What causes DNA cleavage at distant sites? The simple answer is that DNA is cut wherever a pause in translocation occurs (*5*, *11*). This was first proposed to account for the distribution of cleavage sites produced on the phage T7 genome by *Eco*KI (*12*). As two neighboring enzymes translocate, they will converge and collide, on average,

† This work was supported by Welcome Trust Research Career Development Fellowship 053856/Z/98.

* Correspondence should be addressed to this author at the Department of Biochemistry, University of Bristol, Bristol BS8 1TD, U.K. Tel: +44 117 928 7439, FAX: +44 117 928 8274, Email: mark.szczelkun@bristol.ac.uk.

10.1021/bi011824b CCC: $22.00
Published on Web 01/09/2002

halfway between the two sites (Figure 1B). Accordingly, the series of discrete DNA fragments observed after digestion of T7 DNA corresponded to cleavage loci midway between each pair of *Eco*KI sites. Such distributions were also observed following digestion of T7 DNA by *Eco*KI in vivo, suggesting that similar translocation/collision events occur in the cell (*7*). Furthermore, type I enzymes cannot cleave single-site linear DNA at stoichiometric concentrations of enzyme and substrate (*5*)—head-on collision with a second translocating enzyme or a stationary roadblock is required, although specific protein−protein contacts are not critical (*11*). Therefore, it could be stated that DNA cleavage by type I enzymes is dependent upon the preceding translocation and collision processes.

In contrast to the discrete *Eco*KI profiles, the cleavage loci distributions generated in vitro by the *Eco*R124I endonuclease were markedly more distributive (*4*)—although maximal cleavage was observed midway between sites at short inter-site spacings (400−800 bp), at intermediate spacings (800−1000 bp) the distribution broadened, while at longer spacings (>1000 bp) the distribution was diffuse and cleavage was observed at nearly every intervening locus (Figure 1B). In the latter case, a maximum midway between the sites was not discernible. Even allowing for Lorentzian diffusion of DNA fragments during gel electrophoresis (*13*), these distributions are surprisingly broad. The kinetics of *Eco*R124I translocation are suggestive of an efficient process over distances <4370 bp, where dissociation, stalling, and back-stepping are exceedingly rare events (*6*). How can an apparently orderly translocation process give rise to such a broad distribution of cleavage sites? Does the rate and location of DNA cleavage provide a truly accurate reflection of the preceding DNA translocation? These questions are answered by investigating the effect of asynchronous events during translocation and collision upon the kinetics of DNA cleavage. The kinetic models are generally applicable to all DNA translocases, describing both motion on DNA and collision events with other mobile enzymes or stationary roadblocks.

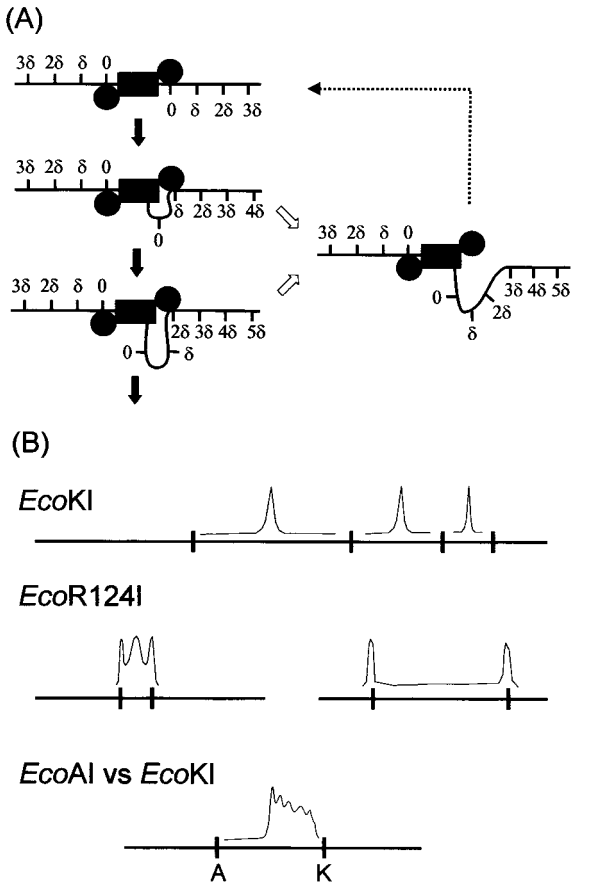
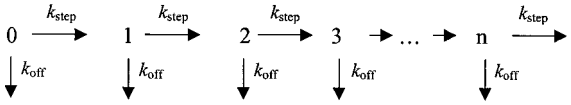


FIGURE 1: Schematic of DNA translocation and cleavage by type I restriction enzymes. (A) DNA is represented by a black line, and the endonuclease as a black rectangle (HsdM and HsdS subunits) and two black circles (HsdR subunits). Each HsdR starts at position 0 on opposite sides of the recognition site and executes ATP-dependent translocation steps in opposite directions of step size δ—motion on one side only is shown for clarity. With time, an ensemble of enzymes distribute between positions 0, δ, 2δ, 3δ, ..., etc., on each side. Because the HsdS subunit remains bound at its recognition site indefinitely (*3*), two expanding loops of DNA form. If an HsdR were to release the DNA at any stage (white arrows), then it could restart at position 0 on the same side of the DNA molecule via "resetting" (dotted arrow) (*6*). (B) Distribution of cleavage loci between pairs of type I restriction sites. DNA is represented as a black line with the recognition sites as tick marks and the extent and position of cleavage as fine black lines. For *Eco*KI, cleavage of phage T7 DNA produced discrete DNA fragments resulting from collisions midway between each pair of sites (*7*, *12*). For *Eco*R124I, cleavage was always observed at the recognition sites whereas discrete fragments were only observed at the shortest spacing (*4*). For a mixture of *Eco*AI and *Eco*KI, cleavage was located asymmetrically toward the *Eco*KI site with a distinct periodicity (*11*).

METHODS AND RESULTS

A General Kinetic Model for DNA Translocation. There are a number of papers that deal comprehensively with the kinetic and thermodynamic theory of nucleotide hydrolysis and protein motion on DNA (e.g., *14*, *15*). Instead, this paper focuses on the probability functions that describe the distributions resulting from multiple catalytic cycles, i.e., processive DNA translocation. To address the relationship between translocation, collision, and cleavage by type I restriction endonucleases, a general model for motion on DNA will be defined first. Suitable models have been described for other processive sequences of reactions (*16*) and successfully applied to DNA (*17*, *18*) and RNA (*19*) helicases, and to Exonuclease III (*20*). As in these cases, protein motion is considered simply as unidirectional stepwise movement over a regular lattice (Figure 1A). A protein is bound to the lattice at position 0, and translocation is initiated by addition of ATP. At each stage thereafter, the protein may either move forward one step or release the track—a process governed by just two rate constants. Although type I enzymes move bi-directionally (*3*, *6*, *11*), the kinetics of only one HsdR need be considered—this is valid as they act separately and can be treated as isolated kinetic entities. The kinetic scheme can be represented as follows:



(1)

Equation 1 is characterized by the following features:

1. Each particle moves one step (δ) according to k_{step}, a rate-limiting constant that encompasses many microscopic events, e.g., ATP binding, hydrolysis, phosphate release, motion on DNA, etc. Note that, in terms of a bona fide

helicase mechanism, the *kinetic* step size (δ) could relate to a *translocation* step size (distance physically moved forward) or an *unwinding* step size (number of bp unzipped). The validity of using simple macromolecular kinetic terms has been justified by studies of both type I enzymes and other helicases (*6*, *17*−*19*).

2. Motion is unidirectional, and back-steps are not considered—each step is thus considered as an irreversible unimolecular reaction.

3. k_{step} is identical at each step. While such uniformity is certainly not true of motors such as RNA polymerases that have complex kinetic relationships between dwell times and template sequence (*21*), both helicases (*17*, *19*) and exonuclease III (*20*) have been shown to have essentially equivalent translocation rates at each step (see below). Inevitably, there will be DNA sequences that introduce a structural barrier to regular progression (e.g., AT-tracts), but these are not considered here.

4. The enzyme can also release the DNA at any point with the rate constant k_{off}. Again, this is considered independent of sequence. In eq 1, when the enzyme releases its track, it cannot rebind and undertake further steps. The resetting model for type I enzymes, where rebinding can occur (Figure 1A), will be considered below.

5. Each step is statistically independent—there is no memory of the antecedent steps. Whether this is true is a matter of debate, as there are a growing number of examples from single molecule studies which suggest that successive enzyme turnovers are not independent (e.g., *22*).

The general solution for the fractional probability of occupying position n on a linear lattice at time t, $P_n(t)$, is given by (derivation in Supporting Information):

$$P_n(t) = \frac{(k_{step} \cdot t)}{n!} \exp[-(k_{step} + k_{off}) \cdot t] \qquad (2)$$

where similar equations have been used as diagnostics of helicase activity (*17*−*19*), fitting the empirical data relied on the presence of lag phases that had a linear relationship to the number of steps. For any distance, the length of the lag will be longest when k_{step} (and k_{off}) is of similar magnitude at each step (*16*). Thus, the kinetics of strand displacement suggest strongly that helicase translocation rates are independent of sequence. The *Eco*R124I translocation kinetics were determined using a similar lag−distance relationship (*6*), indicating that type I enzymes also move using a uniform, Poisson mechanism.

Distributions of Nonprocessive Motor Proteins. Processivity is a descriptor of translocation efficiency and is the probability (P_T) of translocating 1 step forward rather than dissociating from the lattice. In terms of eqs 1−2, it can be defined as $P_T = k_{step}/(k_{step} + k_{off})$. When $P_T < 1$, there is a finite chance that at any point during motion the enzyme will dissociate from the DNA rather than stepping forward. This has been seen with many helicases (e.g., *17*), but may be due in part to the absence from the in vitro reactions of vital "coupling factors" (*15*). The linear distribution of a nonprocessive motor protein has a *normal* profile, as expected for a Poisson enzyme, but the area under the curve tends from unity toward zero (Figure 2 for $P_T = 0.5$). Since rebinding of dissociated enzymes is not considered, the chance of reaching a distant site falls as a function of $(P_T)^n$.

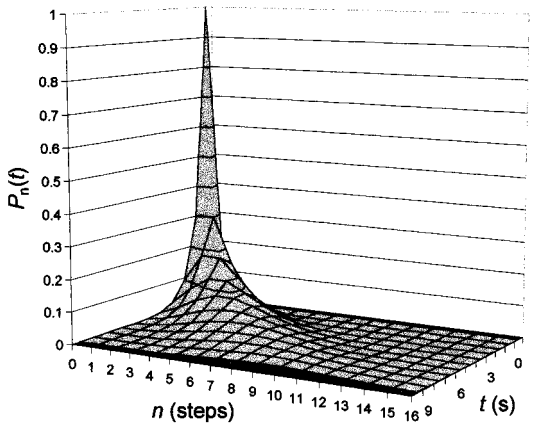


FIGURE 2: Probability distribution for a nonprocessive enzyme. Translocation was simulated using eq 2 for $n = 0-16$, $t = 0-10$ s, and $k_{step} = k_{off} = 1/s$ ($P_T = 0.5$). $P_n(t)$ represents the probability of taking n steps in time t.

Even if k_{step} is significantly faster than k_{off}, the chance of moving long distances is reduced. For example, if k_{step} were 1000-fold faster than k_{off} ($P_T = 0.999$), the probability of making at least 100 steps would be 0.90, but the probability of making at least 1000 steps would be only 0.36.

In contrast to many helicases, type I enzymes are self-contained motors with no need for coupling factors (*1*). They can translocate over distances >4000 bp (*6*, *7*), and cleavage resulting from collision between translocating enzymes can occur when the sites are >18 000 bp apart (*12*). Both translocation and cleavage go to 100% completion in every case. Either type I enzymes release their tracks *very* rarely ($P_T \approx 1$) or dissociated HsdR subunits can restart translocation by resetting (Figure 1A). *Eco*R124I can form an $R_1M_2S_1$ complex (*23*) which, although unable to cut DNA (*5*), is capable of nonprocessive translocation (*6*). The length of time taken for this complex to travel a given distance increased in a nonlinear manner, but every enzyme eventually reached the target site. These results suggest that dissociation is occurring but that the architecture of the type I complex is such that resetting is not a deleterious rate-limiting process. The chance of having dissociated from the DNA increases according to $1 - (P_T)^n$, and so the number of cycles of translocation−dissociation will also increase with distance. This produces a linear continuum of populated states along the DNA. In distribution terms, this is manifested as a burst amplitude in $P_n(t)$ that is proportional to $(P_T)^n$, followed by a very slow nonexponential decay in $P_n(t)$ (Figure 3). Correspondingly, these distributions lead to slow nonexponential translocation kinetics (*6*). Nonetheless, given enough time, an enzyme can still explore distant sites despite a high k_{off}:k_{step} ratio because resetting always returns the dissociated HsdR subunit to the same DNA molecule. The extended distribution that resetting produces could explain the positions of the *Eco*R124I cleavage loci (*4*), except that the translocation kinetics of the cleavage-proficient $R_2M_2S_1$ complexes are markedly different.

Another kinetically significant event that could produce asynchrony in an otherwise processive population of enzymes would be a rate-limiting step prior to the start of motion. If the subsequent translocation steps were "infinitely" fast in comparison, then the spatial and temporal distribution of the enzyme ensemble would become more distributive, dictated

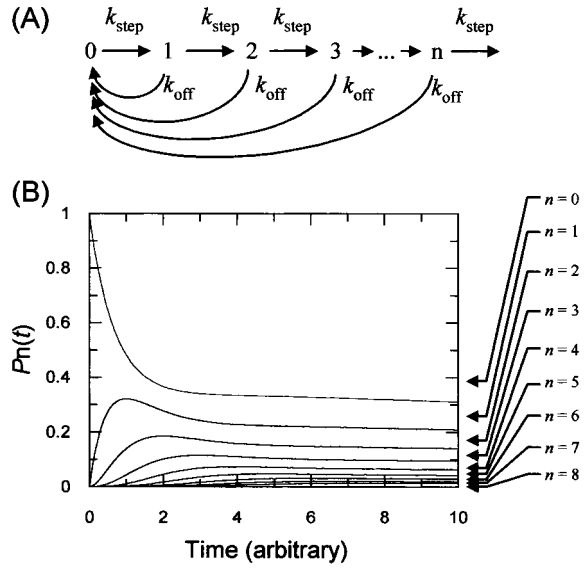


FIGURE 3: Affect of resetting on the distribution of translocating type I enzymes. (A) Schematic model of DNA translocation, dissociation, and resetting. Starting at position 0, an enzyme steps forward with a unimolecular rate constant k_{step} to position 1. Thereafter, the enzyme may either step forward again, or dissociate with the rate k_{off}. Upon release, the enzyme immediately restarts the process at position 0, regardless of where the track was released. This resetting process is defined as "infinitely" fast (*6*). (B) The model in (A) was simulated for $n = 8$ steps by numerical integration of a series of differential equations describing each state using Scientist (MicroMath Software, Salt Lake City, UT). The probability $P_n(t)$ of finding an enzyme at position n after time t is shown; $k_{step} = 1/s$ and $k_{off} = 0.5/s$ ($P_T = 0.67$). The initial burst at each locus represents those enzymes that travelled to/from that site without dissociation; the amplitude decreases as n increases.

solely by the rate-limiting initiation step. In this extreme case, all cleavage would occur at the type I sites as the chance of two opposing enzymes on the same DNA molecule being released from their "stall" sites at precisely the same time would be very small. Similarly, if rare stall events occurred during motion with half-lives significantly longer than the translocation steps, then the enzyme ensemble would also become desynchronized beyond the stall sites. However, the kinetics of triplex displacement used to measure translocation by *Eco*R124I showed linear and exponential relationships, which indicated clearly that neither of these events occur (see *6* for more details). Moreover, a slow initiation step would affect the distribution of enzymes independent of the spacing between pairs of sites, while the *Eco*R124I cleavage profiles become measurably more distributive as the distance between sites is increased (*4*). While stall events are certainly significant to polymerases (*21*) and slow initiation steps have been observed for some helicases (*18*), the kinetics of translocation by *Eco*R124I are indicative of a highly synchronized process in sharp contrast to the apparently unsynchronized cleavage reactions.

Distributions of Processive Motor Proteins. As k_{off} becomes vanishingly small relative to k_{step} and $P_T \rightarrow 1$, an enzyme can travel longer distances before releasing the DNA. For example, the RecBCD helicase is highly processive and is capable of translocating tens of thousands of base pairs without dissociation (*24*, *25*). In this case, k_{step} must be at least 33 000-fold faster than k_{off} ($P_T = 0.99997$). When $k_{off} \ll k_{step}$, eq 2 can be simplified to

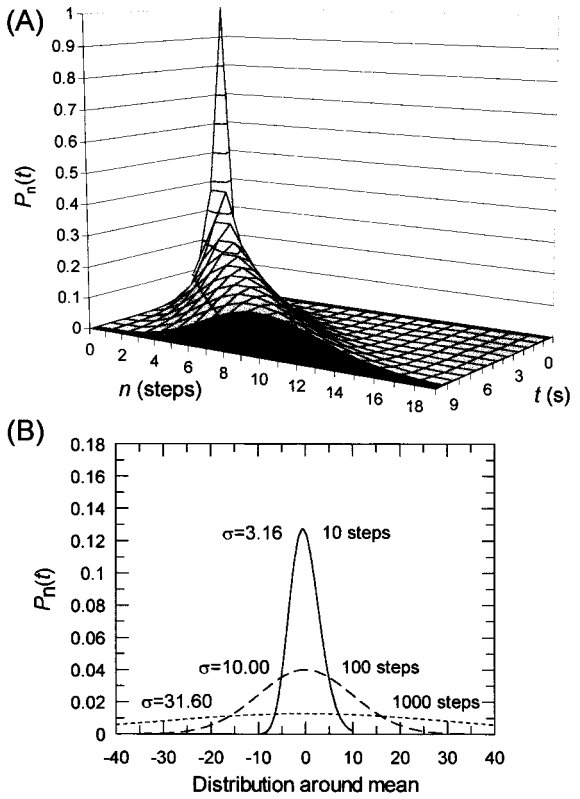


FIGURE 4: Probability distributions for a processive enzyme. (A) Translocation was simulated using eq 3 for $n = 0-19$, $t = 0-10$ s, and $k_{step} = 1/s$ ($P_T = 1$). $P_n(t)$ represents the probability of taking n steps in time t. (B) Spatial distributions after 10, 100, and 1000 s assuming the same kinetic parameters as in (A). Profiles are centered at $x = 0$ about the mean number of steps ($\mu = k_{step} \cdot t$) where $\mu = 10$ steps after 10 s (solid line), 100 steps after 100 s (dashed line), and 1000 steps after 1000 s (dotted line). Standard deviations [$\sigma = (k_{step} \cdot t)^{1/2}$] are shown next to the appropriate curves.

$$P_n(t) = \frac{(k_{step} \cdot t)}{n!} \exp(-k_{step} \cdot t) \qquad (3)$$

An example of the temporal and spatial distribution described by this Poisson function is shown in Figure 4A. The standard deviation is described by $\sigma = (k_{step} \cdot t)^{1/2}$ [or $\sigma = (\mu)^{1/2}$ where μ = mean number of steps at t]. From the distributions at $\mu = 10$, 100, and 1000 steps (Figure 4B), one can clearly see that the spatial distribution during translocation is relatively compressed. This distribution matches that expected given the kinetics of triplex displacement observed with the $R_2M_2S_1$ complex of *Eco*R124I (see *6*). Translocation by *Eco*R124I must therefore be processive. So how can the distributions in Figure 4B account for the alternative cleavage patterns seen with *Eco*R124I and *Eco*KI?

A General Kinetic Model for Head-On Collision between Processive Translocating Enzymes. Cleavage of linear DNA is stimulated by collision of two type I enzymes (*3*, *12*). What pattern of collision loci results given the profiles described for processive enzymes by eq 3? To derive a general solution for the probability of collision, a lattice model similar to eq 1 was used, except that two enzymes start at opposite ends of the lattice (Figure 5A). The *left* HsdR starts at position x and the *right* HsdR at y. The initial configuration is given by the coordinates (x,y) and the distance between the sites as $d = y - x$. The subunits move toward each other at a rate determined by k_1 (*left*) and k_2 (*right*); the relative probability

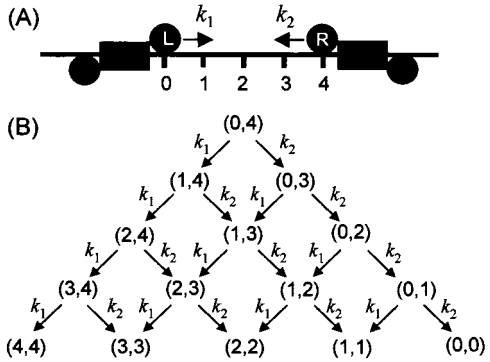


FIGURE 5: Binomial probability machine to describe collision events. (A) Example of schematic model showing two enzymes on linear DNA (as in Figure 1), 4 steps apart ($d = 4$). The HsdR subunits on the *left* (L) and *right* (R) travel toward each other with the rate constants k_1 and k_2, respectively. (B) The starting configuration in (A) is indicated by the coordinates (0,4). All the possible configurations that can result from convergence of the HsdR subunits are then arranged in a Pascal's triangle. Only one enzyme may move at each step, and no back-steps are allowed. For example, if the *left* HsdR moved first, the resulting position is indicated by (1,4). Thereafter, either the *left* or the *right* subunit may move to form (2,4) or (1,3), respectively, and so on. Collision occurs when the subunits occupy the same locus, e.g., (2,2).

that the *left* HsdR will advance one step is given by $k_1/(k_1 + k_2)$, the probability of the *right* by $k_2/(k_1 + k_2)$. If one considers a suitably short time interval (dt), then only one enzyme can move at any given point, and a collision scheme can be defined (Figure 5B). Stepping continues until the enzymes occupy the same locus; collision occurs when $x = y$. A general solution that describes the fractional probability of collision at $x = y = n$ at time t, $P_c n(t)$, can be derived for any value of d as (derivations in Supporting Information):

$$P_c n(t) = \frac{d!\cdot {k_1}^n\cdot {k_2}^{d-n}}{n!\cdot(d-n)!\cdot(k_1+k_2)^d} - \left(\frac{d!}{n!\cdot(d-n)!}\right)\cdot\sum_{i=0}^{d-1}\frac{{k_1}^n\cdot {k_2}^{d-n}\cdot t^i}{i!\cdot(k_1+k_2)^{d-i}}\cdot\exp[-(k_1+k_2)\cdot t] \quad (4)$$

This solution will apply to any pair of motor enzymes that move along a lattice and collide head-on. When collision is between two identical enzymes ($k_1 = k_2 = k_{step}$), the general solution reduces to

$$P_c n(t) = \frac{d!}{n!\cdot(d-n)!\cdot 2^d} - \left(\frac{d!}{n!\cdot(d-n)!}\right)\cdot\sum_{i=0}^{d-1}\frac{{k_{step}}^d\cdot t^i}{i!\cdot(2\cdot k_{step})^{d-i}}\cdot\exp(-2\cdot k_{step}\cdot t) \quad (5)$$

Figure 6A shows an example of eq 5 simulated for $d = 100$ steps. As $t \rightarrow \infty$, eq 5 simplifies to a binomial/Gaussian function with a standard deviation $\sigma = 0.5\cdot d^{1/2}$. From the endpoint profiles at $d = 20$, 200, and 2000 steps (Figure 6B), one can clearly see that the spatial distribution upon collision is even narrower than during translocation (Figure 4B). If $\delta = 1$ bp and the DNA was cut precisely where collision occurred, then the cleavage and collision profiles would overlay. This relationship can easily account for the *Eco*KI results (*7*, *12*, Figure 1B). It has been suggested that DNA looping by *Eco*KI—where two adjacent enzymes interact and loop-out the intervening DNA—coordinates translocation of the distal HsdR subunits such that collision always occurs *exactly* midway between the sites (*7*, *26*). There is no need to invoke this given the characteristics of the profiles in Figure 6B.

How can the above models explain the much broader patterns observed using *Eco*R124I (*4*, Figure 1B)? One way to expand the distributions is to assume $\delta > 1$ bp. Indeed, although step sizes of 1 bp have been estimated for some helicases (*27*), larger steps sizes of 3−5 bp (*17*, *19*) and even 23 bp (*28*) have also been measured. In such cases, the distribution upon collision would become wider but would also be discontinuous (the periodicity set by δ). For example, with $\delta = 10$ bp, collision over 2000 bp ($d = 200$ steps) would distribute with $\sigma = \pm 22.4$ steps (Figure 6B), corresponding to a cleavage profile with an *apparent* σ of ±224 bp (*apparent* because the distribution is discontinuous and thus not a true Gaussian). However, even allowing for band broadening during electrophoresis (*13*), this is still not as diffuse as the *Eco*R124I profiles. To be significant, δ would have to be $\gg$10 bp. But this would result in a clear periodicity in the cleavage profile which has never been observed when the pair of enzymes are identical. Another factor which one could invoke is that the translocation start site is randomized (these loci have never been mapped for type I enzymes). However, this variable would not change as a function of d and thus can be discounted as a factor in the observed *Eco*R124I profiles. A third possibility is that translocation consumes so much ATP that its concentration drops below the K_m for binding; motion and subsequently collision could then become asynchronous. However, the high concentrations of ATP used in most in vitro assays (2−4 mM) probably discount this as well.

Distinguishing between Dissociation during Motion or upon Collision: At What Stage Are Type I Enzymes Processive? Upon collision between *Eco*R124I enzymes, the HsdR subunits cooperate to cleave both DNA strands in a sequential process with equal rates for each step (*5*, *29*). If one assumes $\delta = 1$ bp and the translocation velocity is 400 bp/s (*6*), then k_{step} must be ~24 000-fold faster than the first cleavage step ($k_{cut} = \sim$1/min; *5*). As translocation appears highly processive for at least 4300 bp (*6*), one can also assume that k_{step} is ~24 000-fold faster than k_{off} (thus, $k_{off} = \sim$1/min). In this case, although dissociation happens only *very* rarely during motion ($P_T > 0.99995$), k_{cut} and k_{off} are of similar magnitude, and, upon collision, and there is a finite chance that one or other HsdR will release the DNA before cleavage can occur. (Furthermore, the increased stall force experienced by both proteins in the collision complex may artificially increase k_{off} relative to k_{cut}.) This can be defined in probability terms as a *processivity of cleavage*: $P_C = k_{cut}/(k_{cut} + k_{off})$. If dissociation occurs, the free HsdR could restart translocation at its origin by resetting (Figure 1A), while its distal partner from the collision complex could restart directly at the collision locus. From eq 5, the pair could then recollide, on average, halfway between the restart positions. Again, there is a finite chance that one or other HsdR will dissociate before cutting the DNA. If cleavage were never to occur, and given enough time, the pair would eventually explore every possible collision locus across d. Other

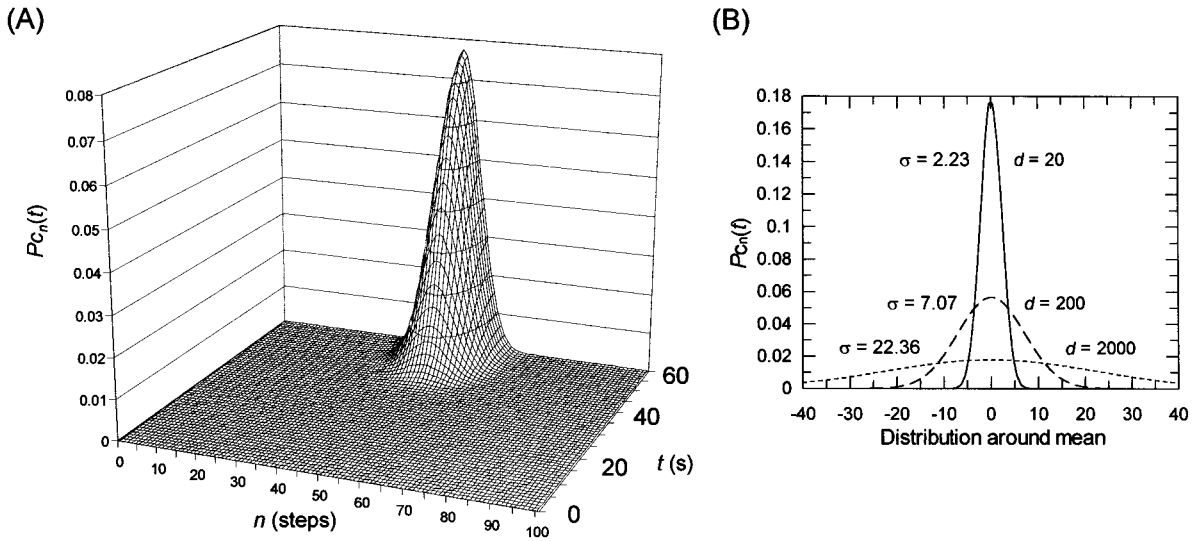


FIGURE 6: Probability distributions of collision loci between two converging motor proteins. (A) Collision was simulated using eq 5 for $d = 100$, $t = 0-60$ s, and $k_{step} = 1$/s ($P_T = 1$). $Pc_n(t)$ represents the probability of two enzymes colliding at position n in time t. (B) Spatial distributions of collision loci as $t \rightarrow \infty$ when $k_{step} = 1$/s and $d = 20$ (solid line), 200 (dashed line), and 2000 steps (dotted line). The profiles are centered at $x = 0$ about their mean collision sites ($\mu = 0.5 \cdot d$ when $k_1 = k_2$). Standard deviations ($\sigma = 0.5 \cdot d^{1/2}$) are shown next to the appropriate curves.

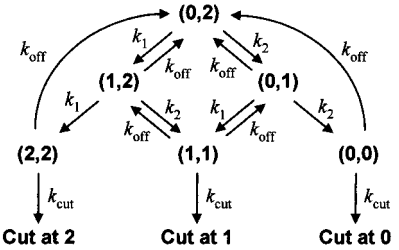


FIGURE 7: Probability machine to describe translocation, dissociation, resetting, collision, and cleavage by type I enzymes. The basic scheme is as described in Figure 5 except that dissociation (k_{off}) can occur during translocation, and dissociation or cleavage (k_{cut}) upon collision. As before, the position of each enzyme is described using (x,y) coordinates. If an HsdR dissociates, it returns to its origin and restarts; e.g., if the subunit at $x = 1$ in (1,1) falls off the DNA, the coordinate resets to (0,1). The resetting step is modeled as infinitely fast (*6*). Upon collision, either the DNA is cut or dissociation occurs, governed by the k_{cut}:k_{off} ratio. The enzymes continue to cycle until all the DNA has been cut at either position 0, 1, or 2. The model can be simulated for any value of d (Supporting Information); bearing in mind that when an HsdR dissociates, it does not move back just one step, but resets to its origin.

phenomenological models can be envisaged which produce similar asynchrony, such as a back-and-forth redistribution as the force generated by one motor causes the other to slip backward before cleavage can occur. However, the key point is that the asynchronous event occurs at collision/cleavage and *not* during translocation. Dissociation will become a significant event for any processive DNA translocase that stalls for a period similar to $\tau = 1/k_{off}$.

How could dissociation at collision affect the kinetics and distribution of cleavage? This was explored by simulating probability models as described in Figure 7 for $d = 2$, 4, 6, and 8 steps. When $k_{off} = 0$ ($P_T = P_C = 1.00$), cleavage rates are nearly identical for each spacing, and the loci distribute as the expected binomial profiles (Figure 8A). Alternatively, if there is a moderate chance of dissociation during motion ($P_T = 0.95$), then the probability of dissociation upon collision is greater because of the extended dwell time ($P_C = 0.50$). The resulting cleavage rates still overlay, but the distribution of cleavage sites becomes increasingly platykurtic as d lengthens (Figure 8B). When the chances of releasing DNA both during motion and upon collision are high ($P_T = 0.50$, $P_C = 0.05$), then the cleavage rates no longer overlay and decrease nonexponentially as d lengthens (Figure 8C). The cleavage distributions are even more platykurtic, to the extent that there is little or no variation across the probability space [as $P_T \rightarrow 0$, $P_cn(t)$ becomes identical at each n locus].

Measurement of the rates of linear DNA cleavage by *Eco*R124I showed kinetics that are independent of d between 800 and 3500 bp (*5*). If both P_T and P_C were $\ll 1$ (Figure 8C), then the model could account for the measured distribution of cleavage loci (*4*) but could *not* account for the kinetics (*5*). When the translocation steps are fast (*6*), resetting events could be masked by the rate-limiting cleavage step (*16*). But for $d \gg 100$ bp, there are at least several orders of magnitude more translocation states at which dissociation could occur compared to collision states, and dissociation−resetting produces extended nonexponential translocation kinetics (Figure 3) that in turn lead to nonexponential cleavage kinetics. However, if $P_T \approx 1$ but $P_C < 1$ (Figure 8B), both the kinetics of translocation and the kinetics and spatial distribution of cleavage match the *Eco*R124I observations (*4*−*6*). This can also explain cleavage sites that appear close to the *Eco*R124I sequences (*4*). For instance, if a dissociated HsdR resets to its origin but fails to restart, re-collision will then occur at this stall site.

DISCUSSION

From the models discussed above, simple statistical descriptions of type I enzyme activity can account for all the observed kinetics and distributions of DNA translocation and cleavage. The models can be readily adapted to circular DNA substrates, and the relationship between translocation, twisting, and cleavage of plasmids by type I enzymes is being investigated (S. McClelland and M. D. Szczelkun, unpublished data). The conclusions from linear DNA are that

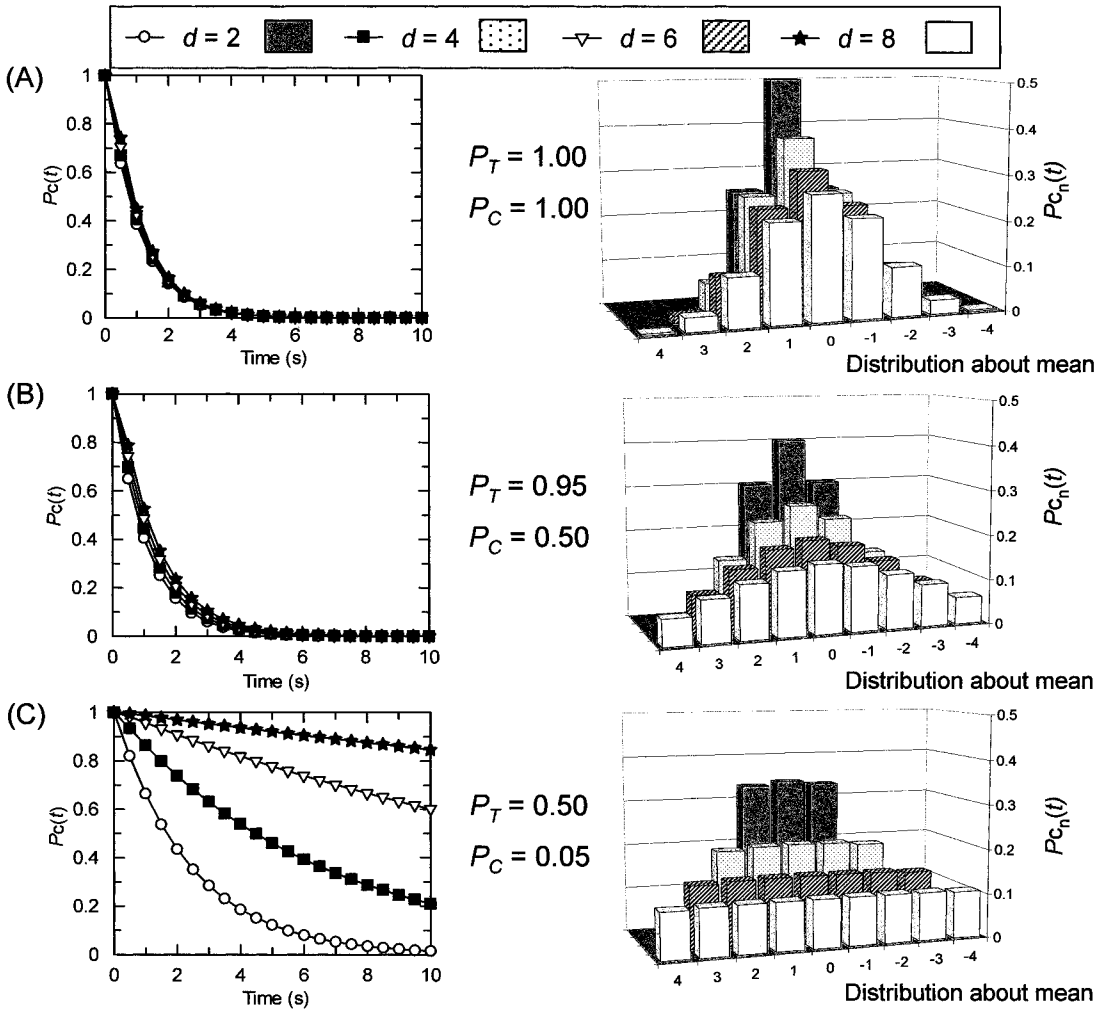


FIGURE 8: Dissociation during translocation and collision and its effect on cleavage rates and distributions. Reactions were simulated by numerical integration in Scientist using equations (see Scientist models in Supporting Information). Reactions with $d = 2$, 4, 6, or 8 steps (see key) were simulated for three processivity conditions: (A) $P_T = P_C = 1$ ($k_1 = k_2 = 20$/s, $k_{off} = 0$, $k_{cut} = 1$/s); (B) $P_T = 0.95$, $P_C = 0.50$ ($k_1 = k_2 = 20$/s, $k_{off} = 1$/s, $k_{cut} = 1$/s); and (C) $P_T = 0.50$, $P_C = 0.05$ ($k_1 = k_2 = 20$/s, $k_{off} = 20$/s, $k_{cut} = 1$/s). DNA cleavage is shown as a probability $Pc(t)$ (the fractional sum of all uncut states), and the final distribution of cleavage sites $Pc_n(t)$, as a series of bar-charts centered at $x = 0$ about $\mu = 0.5 \cdot d$.

translocation by *Eco*R124I is processive, whereas the collision and/or cleavage events are not. This produces uniform cleavage kinetics but widely distributed cleavage loci (*4*, *5*). Conversely, the models suggest that *Eco*KI rarely releases linear DNA at any stage, leading to discrete cleavage patterns (*7*, *12*). A key prediction from this is that if *Eco*R124I and *Eco*KI were to collide, the *Eco*R124I HsdR subunit would fall off more readily due to its lower P_C value. Notwithstanding any difference in k_{step} (*6*, *7*), the resulting cleavage profile would be shifted toward the *Eco*R124I site. This is exactly what has been seen (*11*)—combining *Eco*KI and *Eco*R124I on linear DNA carrying one copy of each site produced a cleavage pattern located primarily in a short region near the *Eco*R124I sequence. Similarly, combining *Eco*R124I and *Eco*AI resulted in cleavage predominantly at the *Eco*R124I site, suggesting that *Eco*AI also has a higher P_C value than *Eco*R124I. When *Eco*KI and *Eco*AI were combined, cleavage was observed throughout the region starting halfway between the sites and ending at the *Eco*KI site, with the fragments appearing as a regular ladder with spacings of ~100−150 bp (Figure 1B). These results can be explained if both enzymes have a similar k_{step} while the P_C of *Eco*AI is 1 and that of *Eco*KI is slightly less than 1. In this case, the first collision would be centered at ~d/2. Cleavage could occur but, equally, some *Eco*KI subunits would release the DNA too early. Re-collision events would then occur toward the *Eco*KI site at ~3d/4, ~7d/8, ~15d/16, ..., etc. (At each point, the number of enzymes involved, and thus the amount of cleavage, also decreases.) On the DNA used by Janscak et al. (*11*), this exponential series of collision loci would result in a regular ladder of cleavage products exactly as observed (Figure 9 in Supporting Information). These conclusions could be corroborated further by examining the translocation and cleavage kinetics of *Eco*AI on a suitable set of linear substrates.

There are many enzymes that translocate DNA (*1*), and, because of the temporal and spatial overlap of competitive cellular processes, collisions with stationary proteins or other motors will be a common feature (e.g., *30*). Therefore, the equations described in this study should be more widely applicable, particularly with the recent appreciation of the important protein−protein collisions that produce potentially catastrophic blockages in replication fork progression (*31*). The closely related type III restriction enzymes are proposed to act via a similar tracking−collision model to type I enzymes (*3*). However, unequivocal evidence for motion is

lacking (*1*), and new assays need to be developed before applying kinetic models.

Several general questions about the models remain: are the values for k_{step} truly independent of local DNA sequence; and, does every enzyme in an ensemble behave in exactly the same way at each step? The first question has been answered using cognate kinetic models to analyze strand separation by DNA and RNA helicases (*17*−*19*). In these cases, the linear relationship between duplex length and lag time could only have been observed if k_{step} was the same at each step (*16*). The second question has been addressed by single-molecule analysis of DNA unwinding by the RecBCD helicase (*25*). These studies have indicated that translocation rates vary far more than expected from eq 3 and Figure 4. Since the mean rates compared well with values derived from bulk solution measurements (*24*), the technique seems sound. Instead, it appears that individual enzymes have very different k_{step} values, a phenomenon usually hidden in bulk solution measurements. However, this can be tested using a "collision assay" based on eq 5. For instance, if two RecBCD enzymes translocated toward each other from opposite blunt ends of a linear DNA, then collision would occur, on average, midway along the substrate. The distribution of the collision sites would be diagnostic of the preceding translocation events: if every enzymes moved with the same k_{step}, the distribution would have a $\sigma = 0.5 \cdot d^{1/2}$ where d is the length of the DNA; but, if k_{step} varied from enzyme-to-enzyme, the distribution would become measurably broader. On a 4363 bp linear DNA at least, the half-length ssDNA fragments resulting from RecBCD translocation, cleavage, and collision at saturating concentrations of enzyme were distributed in a relatively narrow band (*32*). This would appear to be indicative of synchronous translocation, in contrast with the suggestions from the single-molecule studies. However, the cleavage studies were not carried out as a function of DNA length, and complete analysis using both classical and single-molecule techniques would be required to explore fully these properties.

Another hurdle to overcome in the analysis of DNA translocation is the complexity of the kinetic models. For instance, simulating motion for $n > 1000$ bp using eqs 2−5 involves using large denominator and numerator terms. For example, when $n = 1000$ steps, $n! = \sim 0.4 \times 10^{2568}$. Most computer packages simply cannot calculate these numbers. Those that can, using floating point arithmetic (e.g., Maple VI, Waterloo Maple, Ontario, Canada), sometimes introduce exceedingly small errors into the numerator and denominator terms that produce nonetheless significant deviations in the calculated fit. Furthermore, when summation terms involve thousands of iterations, the speed of calculation can become limited by microprocessor power (although this can be improved by simplifying the terms using gamma functions). Therefore, alongside development of new techniques to follow complex events, there is also a need to develop software that can manage the equally complex kinetic output.

ACKNOWLEDGMENT

I thank G. Davies, D. Dryden, and H. Gutfreund for discussions.

SUPPORTING INFORMATION AVAILABLE

The derivations of eqs 2−5, the Scientist models used in Figure 8, and Figure 9 are included. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Szczelkun, M. D. (2000) *Essays Biochem. 35*, 131−143.
2. Ishijima, A., and Yanagida, T. (2001) *Trends Biol. Sci. 26*, 438−444.
3. Rao, D. N., Saha, S., and Krishnamurthy, V. (2000) *Prog. Nucleic Acid Res. Mol. Biol. 64*, 1−63.
4. Szczelkun, M. D., Janscak, P., Firman, K., and Halford, S. E. (1997) *J. Mol. Biol. 271*, 112−123.
5. Szczelkun, M. D., Dillingham, M. S., Janscak, P., Firman, K., and Halford, S. E. (1996) *EMBO J. 15*, 6335−6347.
6. Firman, K., and Szczelkun, M. D. (2000) *EMBO J. 19*, 2094−2102.
7. Garcia, L. R., and Molineux, I. J. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 12430−12435.
8. Webb, J. L., King, G., Ternent, D., Titheradge, A. J., and Murray, N. E. (1996) *EMBO J. 15*, 2003−2009.
9. Davies, G. P., Powell, L. M., Webb, J. L., Cooper, L. P., and Murray, N. E. (1998) *Nucleic Acids Res. 26*, 4828−4836.
10. Davies, G. P., Kemp, P., Molineux, I. J., and Murray, N. E. (1999) *J. Mol. Biol. 292*, 787−796.
11. Janscak, P., MacWilliams, M. P., Sandmeier, U., Nagaraja, V., and Bickle, T. A. (1999) *EMBO J. 18*, 2638−2647.
12. Studier, F. W., and Bandyopadhyay, P. K. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4677−4681.
13. Shadle, S. E., Allen, D. F., Guo, H., Pogozelski, W. K., Bashkin, J. S., and Tullius, T. D. (1997) *Nucleic Acids Res. 25*, 850−860.
14. Young, M. C., Kuhl, S. B., and von Hippel, P. H. (1994) *J. Mol. Biol. 235*, 1436−1446.
15. von Hippel, P. H., and Delagoutte, E. (2001) *Cell 104*, 177−190.
16. Gutfreund, H. (1995) *Kinetics for the Life Sciences*, Cambridge University Press, Cambridge, U.K.
17. Ali, J. A., and Lohman, T. M. (1997) *Science 275*, 377−380.
18. Cheng, W., Hsieh, J., Brendza, K. M., and Lohman, T. M. (2001) *J. Mol. Biol. 310*, 327−350.
19. Jankowsky, E., Gross, C. H., Shuman, S., and Pyle, A. M. (2000) *Nature 403*, 447−451.
20. Linxweiler, W., and Horz, W. (1982) *Nucleic Acids Res. 10*, 4845−4859.
21. Uptain, S. M., Kane, C. M., and Chamberlin, M. J. (1997) *Annu. Rev. Biochem. 66*, 117−172.
22. Lu, H. P., Xun, L., and Xie, X. S. (1998) *Science 282*, 1877−1882.
23. Janscak, P., Dryden, D. T., and Firman, K. (1998) *Nucleic Acids Res. 26*, 4439−4445.
24. Roman, L. J., Eggleston, A. K., and Kowalczykowski, S. C. (1992) *J. Biol. Chem. 267*, 4207−4201.
25. Bianco, P. R., Brewer, L. R., Corzett, M., Balhorn, R., Yeh, Y., Kowalczykowski, S. C., and Baskin, R. J. (2001) *Nature 409*, 374−378.
26. Berge, T., Ellis, D. J., Dryden, D. T., Edwardson, J. M., and Henderson, R. M. (2000) *Biophys. J. 79*, 479−484.
27. Dillingham, M. S., Wigley, D. B., and Webb, M. R. (2000) *Biochemistry 39*, 205−212.
28. Bianco, P. R., and Kowalczykowski, S. C. (2000) *Nature 405*, 368−372.
29. Janscak, P., Abadjieva, A., and Firman, K. (1996) *J. Mol. Biol. 257*, 977−991.
30. Liu, B., and Alberts, B. M. (1995) *Science 267*, 1131−1137.
31. Dillingham, M. S., and Kowalczykowski, S. C. (2001) *Mol. Cells 8*, 734−736.
32. Dixon, D. A., and Kowalczykowski, S. C. (1993) *Cell 73*, 87−96.

BI011824B