Two Time Constants for the Binding of Proteins to DNA from Micromechanical Data

Matthew S. Turner
Center for Studies in Physics and Biology, Rockefeller University, 1230 York Avenue, New York, New York 10021, U.S.A. and Department of Physics, Warwick University, Coventry, CV4 7AL, United Kingdom

ABSTRACT Recent experimental advances allow the direct measurement of the force/extension behavior for DNA in the presence of strongly binding proteins. Such experiments reveal information about the cooperative mechanism of protein binding. We have studied the irreversible binding of such proteins to DNA using a simple simulation and present a method for estimating quantitative rate constants for the nucleation and growth of linear domains of proteins bound to DNA. Such rate constants also give information about the relative energetics of the two binding processes. We discuss our results in the context of recent data for the DNA-recA-ATPγs system, for which the nucleation time is 4.7×10^4 min per recA binding site and the total growth rate of each domain is 1400 recA/min.

INTRODUCTION

It is now possible to carry out precise micromanipulation of single molecules of protein or DNA molecules. These experiments have involved the use of both laser tweezers (Chu, 1991; Kellermayer et al., 1997; Hegner and Bustamante, 1998; Hegner et al., 1999; Shivashankar et al., 1998; Shivashankar and Libchaber, 1998) and the Atomic Force Microscope (Rief et al., 1997; Shivashankar and Libchaber, 1997; Leger et al., 1998; Shivashankar et al., 1999) or combinations of both techniques. The force/extension data for such molecules can yield information about the macromolecule's entropic stiffness, length, and inter- and intramolecular binding (Leger et al., 1998; Marko and Siggia, 1995). In the current work, we will be interested in the effect of proteins binding to a single DNA molecule. Many proteins are known to bind strongly to DNA (von Hippel and McGhee, 1972). One such is recA, which binds to both single- and double-stranded DNA in the presence of either ATP or chemically modified ATPγs (Stasiak et al., 1981; Egelman and Stasiak, 1986; Pugh and Cox, 1987, 1988; Shaner et al., 1987; Brenner et al., 1988; Pugh et al., 1989). ATP hydrolysis by RecA has been shown to render DNA strand exchange unidirectional (MacFarland et al., 1997). This suggests that suppressing ATP hydrolysis may have various physical consequences.

The biological role of recA is to control the organization of DNA during recombination (Alberts et al., 1994; Roca and Cox, 1990, 1997). Such DNA-protein complexes can have significantly different physical properties from bare DNA. For example, recA binds into the major groove of double stranded DNA causing partial local unwinding and extension of its contour length from 3.4 to 5.1 Å per base pair (Stasiak et al., 1981). Each recA molecule occupies a site 3 base pairs in extent (Di Capua et al., 1982).

The binding of recA to DNA is known to be a stochastic, reaction-limited process that depends on the length of the DNA substrate (Pugh and Cox, 1987) and is highly cooperative (Shaner et al., 1987). By measuring the rate of ATP hydrolysis, which depends on the extent of recA coverage on the DNA, the mean binding rate was previously estimated to be 300 recA/min (Pugh and Cox, 1988). There was no mechanical tension on the DNA in these experiments. In the presence of applied forces of 5 pN, Shivashankar et al. (1998) showed this could rise to 900 recA/min and Leger et al. (1998) showed that, over a range of applied forces from 15 to 100 pN, it varied between 300 and 3000 recA/min, respectively. These rates contain no information about chain-length dependence and are not directly comparable with the two separate nucleation and growth rates that are the focus of the present work. However, the mean rate of coverage may be obtained by simply dividing the number of recA sites at 100% coverage by the time taken to achieve full coverage. This rate is of the order of 900 recA/min for the data we use (Shivashankar et al., 1999), corresponding to DNA under a continuous force of 6 pN.

Applied forces have previously been shown to enhance the rate of recA binding (Leger et al., 1998). It is known that the applied force acts to do work on the DNA, thereby reducing the energy barrier to be overcome by recA during binding. A very crude estimate of the reduction in the energy barrier for binding of a single recA molecule to DNA in the presence of an applied force would involve constructing an energy from the product of the force (here 5–6 pN) and the extension of the DNA contour per binding site (here 5 Å) to give a reduction of approximately k_BT. This is consistent with the observed difference in the mean rate by a factor of about 3 between Pugh and Cox (1988) and Shivashankar et al. (1999). Similar micromanipulation experiments could, in principle, be carried out in the absence of continuously applied forces.

Received for publication 13 January 1999 and in final form 10 November 1999.

Address reprint requests to Matthew S. Turner, Department of Physics, University of Warwick, Coventry CV4 7AL, U.K. E-mail: m.s.turner@warwick.ac.uk.

0006-3495/00/02/600/08 $2.00

MODEL AND OBJECTIVES

Here, we present a numerical scheme for extracting quantitative protein–DNA binding rates. Our method allows us to determine the rate of nucleation of separate protein domains on DNA and the rate of growth of these domains. All that is required is a time trace of the fraction of occupied binding sites. For proteins that locally unwind or extend DNA when they bind, this may be obtained by monitoring the length of the DNA molecule.

Although we believe that our approach is quite general, we have in mind a comparison with experimental data from one particular set of micromechanical atomic force microscopy experiments (Shivashankar et al., 1999). These authors claim to measure the maximum extended length L_{max} of a 49-kb double-stranded λ-DNA molecule as a function of time after the addition of excess recA and ATPγs. Because ATPγs cannot be hydrolyzed, the step that removes recA from DNA (Egelman and Stasiak, 1986), recA-ATPγs, can be thought of as binding irreversibly to DNA. In this context "maximum extended length" means the end-to-end distance of the DNA molecule under a large force, here 6 pN. This force puts the chain well into the nonlinear force/extension regime, where all the statistical segments of the chain are almost fully aligned. Nonetheless, it is not so large as to cause significant unwinding or unpairing of the bare double-stranded DNA. Under such large forces the difference between the actual contour length of the DNA chain and the measured value of L_{max} has been predicted to become asymptotically small (Marko and Siggia, 1995). To be precise, these authors show that, for large applied forces, the measured L_{max} and the real contour length differ by a factor $1/\sqrt{4fA/k_BT}$, where f is the force applied to the DNA and A its statistical segment length, typically a few hundred Ångstroms. Taking $f = 6$ pN and $A = 300$ Å, one finds the difference to be only 7%.

We will assume in what follows that the length fraction of DNA decorated with bound proteins ϕ is linearly related to L_{max} according to $L_{max} = \text{constant}\ (1 + \phi(a - 1))$ with constant a given by the ratio of the lengths of decorated to undecorated DNA per base pair. The effect of the differing rigidity of the decorated and undecorated DNA on L_{max} will be small when the applied forces are large (Marko and Siggia, 1995). Similarly, significant unwinding of bare λ-DNA has been shown to occur only at far higher forces, of the order of 50 pN (Leger et al., 1998).

Our work is based on a simple nucleation and growth model that has recently been proposed elsewhere (Shivashankar et al., 1999). The basic tenets of this model are as follows: Proteins may bind at any of a large number of sites along the DNA double helix. This number is so large that we may choose to view the DNA as a continuum of binding sites. Both the energetics and, therefore, the rate of binding of a protein at a given site may depend strongly on the configuration of the DNA at neighboring sites, i.e., whether or not proteins are already bound there. It is this that causes the binding process to be cooperative. We make the further assumption that there are, effectively, only two such rates, corresponding to nucleation and growth events, see Fig. 1. A nucleation event corresponds to a protein binding to a region of DNA that is distant from any proteins already bound whereas a growth event corresponds to binding to a site that is adjacent to one that is already occupied. The motivation for distinguishing just two rate constants can be related to the microscopic properties of the protein–DNA complex. At the end of a protein-decorated domain, the DNA experiences a sharp drop in the forces by which the proteins drive its unwinding and extension. Similarly, the unwinding persists for only a very short distance into the undecorated DNA. For a discussion of a related effect see, e.g., Rudnick and Bruinsma (1999). Hence, the energetics of binding of recA to DNA change from that corresponding to a growth event to a nucleation event over this length-scale. Direct interprotein interactions, including hydrogen bonds, Van der Waals forces, etc., will be similarly short-ranged.

We further assume that there is no directionality in the binding process. It has been shown that the binding of recA to single-stranded DNA in the presence of ATP occurs in a 5′ to 3′ direction (Register and Griffith, 1985). However, the binding process to double-stranded DNA is rather different, with little evidence for directionality (Shaner et al., 1987). Should the binding later be shown to be unidirectional, some modification of the present simulation would be necessary.

It is the goal of the present work to present a scheme for extracting the rates of both the nucleation and growth processes from experimental data for the time variation of the

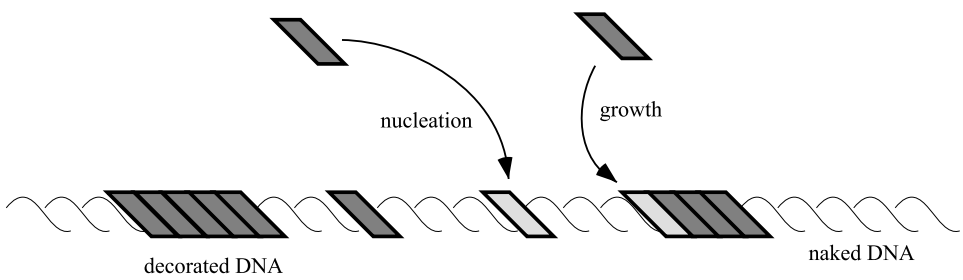

FIGURE 1 A schematic representation of the binding of proteins to a DNA molecule. For clarity, the DNA is shown artificially straight. Some proteins have already bound to the DNA at the time of this snapshot and are shown as grey boxes. Other proteins are shown in the process of binding to the lighter shaded sites via the nucleation and growth processes discussed in the text.

decorated length fraction $\phi(t)$. These time scales will give important information regarding the cooperative mechanism of binding of the protein to DNA, as well as the ratio of the energy scales, from the usual Arrhenius relation. It will prove necessary to carry out a simple numerical simulation of the nucleation and growth process for two reasons. As discussed below, we have reason to believe that the existing analytic mean field model may be unreliable in our regime of interest, and secondly, we can demonstrate that this model fails to distinguish between the two different rates, rather being sensitive only to a geometric mean of the two. We first briefly review this mean field model before going on to discuss the results of our simulation.

ANALYTIC TREATMENT

It is possible to write down coupled differential equations involving $\phi(t)$, the length fraction of DNA decorated with protein, and $n(t)$, the number of domains of bare DNA at time t after introduction of the binding protein. Within the analytic theory, end effects are neglected and the DNA is treated as if it were circular. Thus, the number of domains of naked DNA are equal to the number of domains of protein-decorated DNA.

The length fraction ϕ of decorated DNA increases during each growth and nucleation event. We make the assumption that the fraction of material deposited on the molecule during nucleation events is negligible. This assumption is equivalent to assuming that the maximum number of protein domains remains much less than the number of binding sites on the molecule N. We will explicitly determine later that this assumption is self-consistent in the regime of interest. Thus,

$$\dot{\phi} = n/\tau_{\text{grow}}, \tag{1}$$

where τ_{grow} is a time scale defining the rate of growth of domain ends. As usual, it is merely the inverse of the corresponding rate constant. Note that there is no factor of 2 present in this equation because the rate is defined per domain, rather than per identical domain end.

The second equation defines the rate of change of the number of domains of decorated DNA. There will be two contributions to this. The first must account for the creation, via nucleation events, of new domains on previously bare regions of DNA. This rate is proportional to the length fraction of naked DNA remaining, $1 - \phi$, and to the nucleation rate, $1/\tau_{\text{nuc}}$, which is the inverse of another, independent time scale τ_{nuc}. The second contribution must describe the loss of domains due to the amalgamation of two regions of decorated DNA into one when their growing ends meet. The mean field probability of two neighboring domain ends annihilating per unit time is proportional to both the growth rate $1/\tau_{\text{grow}}$ and the inverse of the mean undecorated length fraction between ends $(1 - \phi)/n$. This is the rate per domain, and so another factor of n is required to obtain the overall rate. Thus,

$$\dot{n} = (1 - \phi)/\tau_{\text{nuc}} - \frac{n^2}{(1 - \phi)\tau_{\text{grow}}}. \tag{2}$$

As described in more detail elsewhere (Shivashankar et al., 1999), the following are exact analytical solution to the simultaneous Eqs. 1 and 2, as can be checked by substitution.

$$\phi(t) = 1 - \exp[-t^2/(2\tau_{\text{grow}}\tau_{\text{nuc}})] \tag{3}$$

$$n(t) = vt \exp[-t^2/(2\tau_{\text{grow}}\tau_{\text{nuc}})] \tag{4}$$

Notice that they contain only a single characteristic time scale,

$$\tau = \sqrt{\tau_{\text{nuc}}\tau_{\text{grow}}}. \tag{5}$$

We will denote these the fast nucleation solutions because they are appropriate in the continuum regime,

$$1/N \ll \tau_{\text{nuc}}/\tau_{\text{grow}} \ll 1. \tag{6}$$

In this regime, the typical number of domains remains large ($\tau_{\text{nuc}}/\tau_{\text{grow}} \ll 1$) and the nucleation rate is sufficiently slow ($1/N \ll \tau_{\text{nuc}}/\tau_{\text{grow}}$) that changes in ϕ due to nucleation events need not be included in Eq. 1.

There are two reasons why we should expect this model to start to fail when $\tau_{\text{nuc}}/\tau_{\text{grow}} \gtrsim 1$ and the typical number of protein domains (at half coverage, say) does not greatly exceed unity. These are that the model explicitly treats n as a continuous variable and also that it fails to properly take account of end effects. These two effects, as well as the mean field treatment of the domain annihilation rate, will be sources of error at the quantitative level in this regime. In spite of the approximations inherent in the analytic treatment, it does give an excellent one-parameter fit to experimental data (Shivashankar et al., 1999). This is so, even in the regime where τ_{nuc} and τ_{grow} are comparable, which is strictly on the borderline of the models validity. However, in spite of this excellent qualitative agreement, we note that 1) the continuum results could be badly wrong at the quantitative level and 2) no information is obtained about the ratio of the two time scales τ_{nuc} and τ_{grow}. The numerical simulation described in the next section provides a scheme for quantitative extraction of both τ_{nuc} and τ_{grow}.

A second limit, that of slow nucleation, can also be treated analytically. This corresponds to $\tau_{\text{grow}}/\tau_{\text{nuc}} \gg 1$, in which limit the molecule waits some Poisson-distributed time, with average τ_{nuc}, for the first nucleation event to occur. After this, the domain quickly zippers to the ends of the DNA in a time $\simeq \tau_{\text{grow}}$. The prefactor here depends on the position of the first nucleation site. For a nucleation event, a fraction x along the DNA molecule at $t = 0$ the rate of growth of decorated length fraction can be written in

terms of Dirac delta functions (Dirac, 1982) and is given by

$$\dot{\phi} = \frac{1}{\tau_{\text{grow}}}\left[1 - \frac{1}{2}\,\delta\left(x - \frac{t}{2\tau_{\text{grow}}}\right) - \frac{1}{2}\,\delta\left(1 - x - \frac{t}{2\tau_{\text{grow}}}\right)\right]. \tag{7}$$

Here, the first term involving a delta function includes the increase in coverage due to growth of the left end of the domain. This term correctly includes the contribution of this end only for times $t < 2\tau_{\text{grow}}x$, before it has reached the end of the DNA molecule. The second delta function plays an exactly analogous role for the right end of the domain.

When integrated and suitably averaged over x, Eq. 7 yields

$$\phi(t) = t/\tau_{\text{grow}} - (t/\tau_{\text{grow}})^2/4, \tag{8}$$

which involves a single, new time scale depending only on τ_{grow}.

The regime of our present experimental interest happens to fall in the crossover regime between the fast and slow nucleation limits where τ_{nuc} and τ_{grow} are comparable. This motivates us to carry out a simple simulation of the nucleation and growth process, as described below.

Finally, we note that the constants τ_{nuc} and τ_{grow} are more mathematically natural units, but they depend implicitly on the length of the DNA. They can be related to corresponding microscopic (length-independent) time constants, denoted by an asterisk ($*$), by the relations

$$\tau_{\text{nuc}} = \tau^*_{\text{nuc}}/N, \tag{9}$$

where N is the number of binding sites per chain and $1/\tau^*_{\text{nuc}}$ is the rate of nucleation per site,

$$\tau_{\text{grow}} = \tau^*_{\text{grow}}N, \tag{10}$$

where $1/\tau^*_{\text{grow}}$ is the rate of growth of the domain in binding sites per unit time. Thus, $1/\tau^*_{\text{grow}}$ has the conventionally reported units of recA monomers per minute. The microscopic units are more physical and, unlike τ_{nuc} and τ_{grow}, do not depend implicitly on the chain length.

SIMULATION

In this section, we will describe the simple numerical scheme we used to model the nucleation and growth process. All times in our routine were recorded from the first nucleation event, which defined $t = 0$ for each run. In experiments, a lag time τ_{lag} may be observed before this event. We will discuss this feature in more detail in the next section, but ignore it at present.

Our scheme proceeds as follows. For each run, the first nucleation event occurs at position x, chosen randomly on [0, 1], and defines $t = 0$. The time step δt is chosen to be sufficiently small to reproduce the nearly continuous growth process and to ensure that only very rarely do we miss multiple nucleation events in the same step. At each time step the routine grows any existing domain ends a distance $\delta t/(2\tau_{\text{grow}})$ and a new nucleation event occurs with probability $\delta t(1 - \phi)/\tau_{\text{nuc}}$. If such a nucleation event does occur, it does so at a position chosen randomly on the unoccupied line fraction. Thus, at each time step, the routine records the position and length of each domain, and hence $\phi(t)$ and $n(t)$ the number of undecorated DNA domains. Each run proceeds until $\phi = 1$, at which time the data for the entire ith run $\{\phi_i(t), n_i(t)\}$ are stored and the next $i + 1$th run is started. A total of $i_{\text{max}} = 1000$ runs are performed. We then identify the appropriate statistical means as follows:

$$\bar{\phi}(t) = \frac{1}{i_{\text{max}}}\sum_{i=1}^{i_{\text{max}}} \phi_i(t), \tag{11}$$

$$\bar{n}(t) = \frac{1}{i_{\text{max}}}\sum_{i=1}^{i_{\text{max}}} n_i(t). \tag{12}$$

The expected bounds on ϕ and n for a single sample run from this ensemble are defined by the variances

$$\sigma^2_{\text{n}}(t) = \frac{1}{i_{\text{max}}}\sum_{i=1}^{i_{\text{max}}} (\phi_i(t) - \bar{\phi}(t))^2, \tag{13}$$

$$\sigma^2_{\text{n}}(t) = \frac{1}{i_{\text{max}}}\sum_{i=1}^{i_{\text{max}}} (n_i(t) - \bar{n}(t))^2. \tag{14}$$

These variances do not in any way represent statistical error bars for our routine, these are so small as to be practically invisible at this scale, but rather, the stochastic variation between runs. Furthermore, the variances σ^2_ϕ and σ^2_{n} are rather loosely constructed insofar as we have not made use of the constraint that both ϕ and n must be nonzero. We do not consider this to be a source of significant error and retain the above definitions for simplicity. However, this is the reason why the one standard deviation envelope apparently includes unphysical negative values of ϕ and n at early times in some runs, even though these values are never observed in the simulations. Our simulation data is now in an ideal format for comparison with the analytical results.

RESULTS

In this section, we will present our simulation data and will compare it with both the analytical solution and the single experimental data set of Shivashankar et al. (1999).

The analytic solutions are compared with our simulation results for $\bar{\phi}(t)$ in Fig. 2, *A–F* and for $\bar{n}(t)$ in Fig. 3, *A–C*. We observe that the simulations converge toward the appropriate analytic limits, but that there is significant departure from both when $\tau_{\text{nuc}}/\tau_{\text{grow}}$ is of order unity. The simulation results have all saturated after a maximum time $2\tau_{\text{grow}}$, corresponding to the extreme case in which the only nucleation event occurs very close to the end of the molecule.

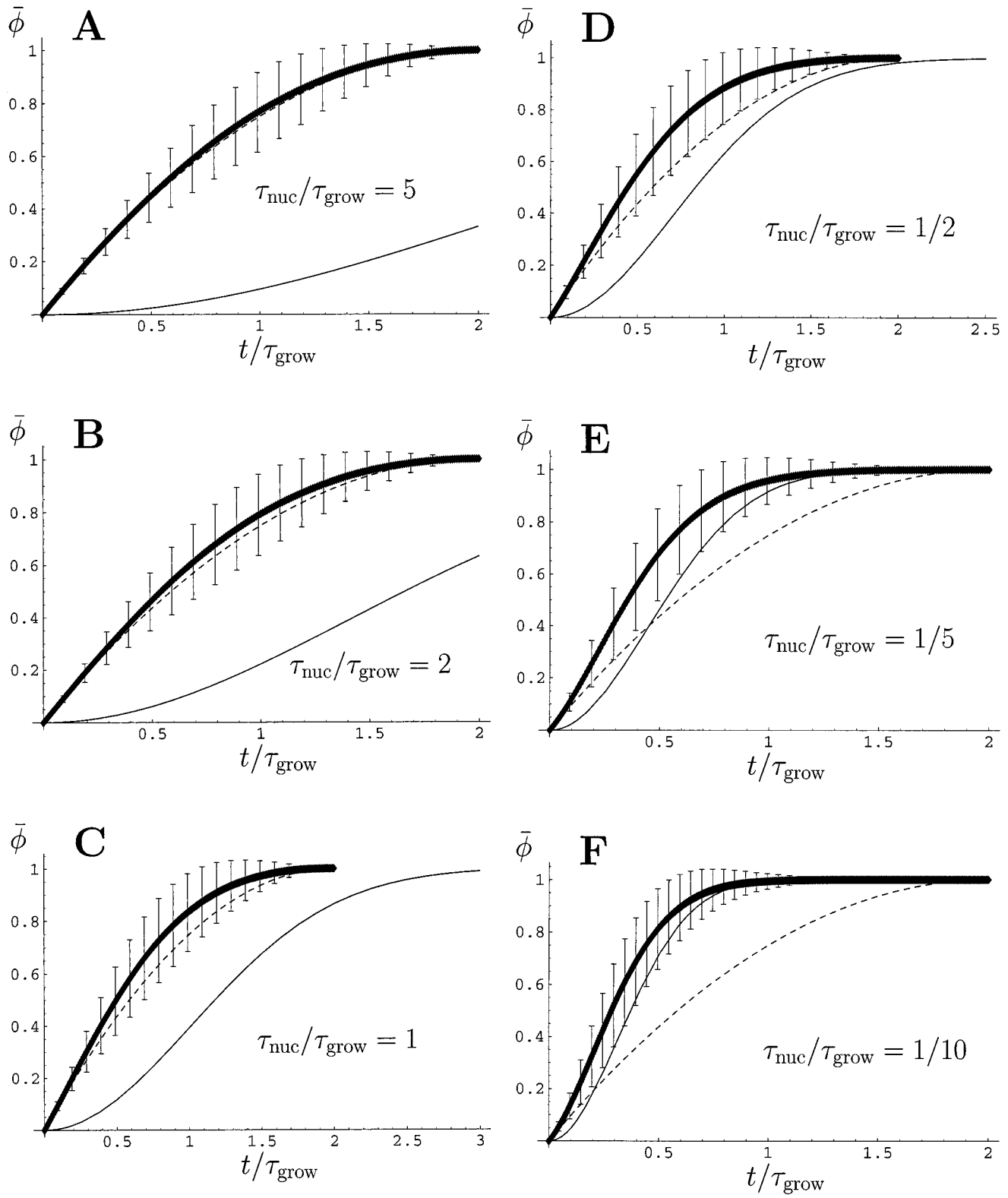

FIGURE 2 The average length fraction of the molecule decorated with protein $\bar{\phi}$ is plotted against time, measured in units of the growth time $\tau_{\rm grow}$, for $\tau_{\rm nuc}/\tau_{\rm grow}$ equal to (*A*) 5, (*B*) 2, (*C*) 1, (*D*) ½, (*E*) ⅕, (*F*) 1⁄10. The simulation results are shown as a thick line, and the bars represent one standard deviation bounds for the stochastic variation between runs. The thin solid line is the analytic result, Eq. 3, for the fast nucleation limit whereas the dashed line is the analytic result, Eq. 8, for the slow nucleation limit.

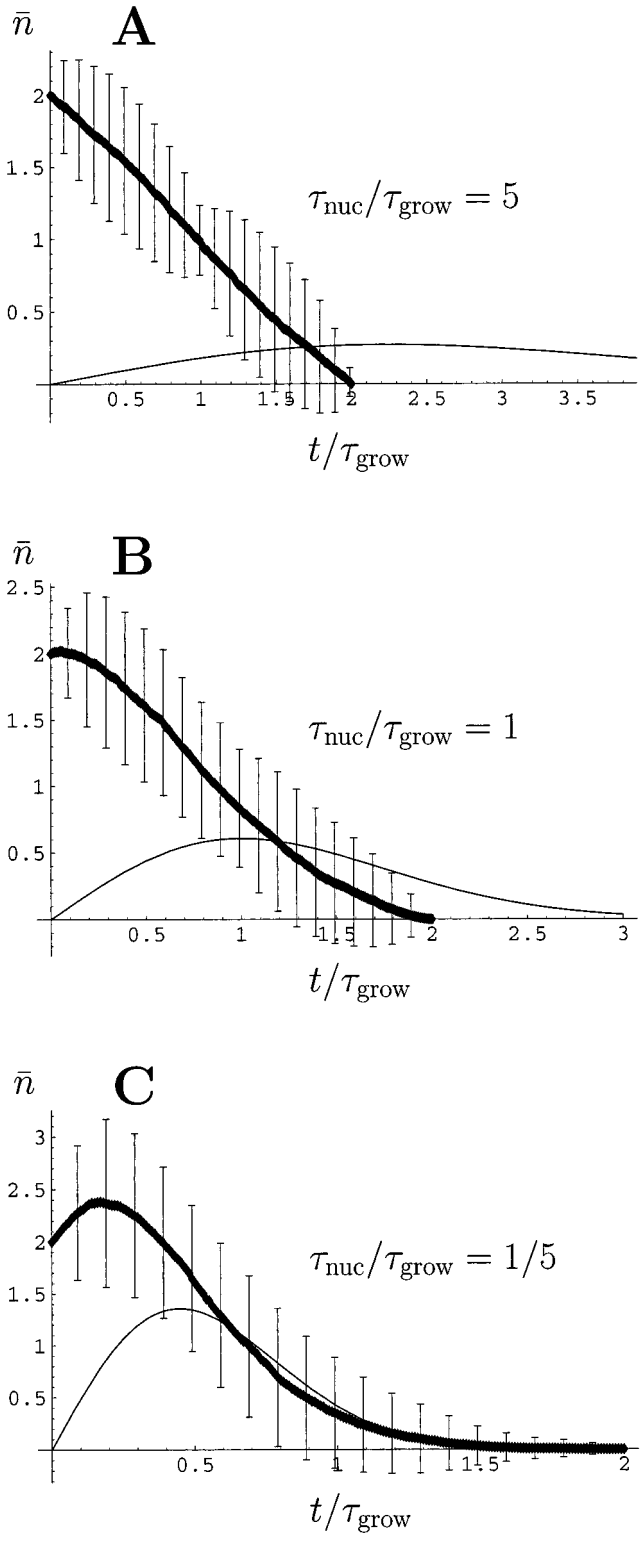

FIGURE 3 The average number of domains on the DNA that are undecorated with protein $\bar{n}$ is plotted against time, measured in units of the growth time τ_{grow}, for τ_{nuc}/τ_{grow} equal to (*A*) 5, (*B*) 1, (*C*) ⅕. The simulation results are shown as a thick line and the bars represent one standard deviation bounds for the stochastic variation between runs. The solid line is the analytic result, Eq. 3, for the fast nucleation limit.

We must be cautious in comparing our results with a single experimental data set. This is because our simulations predict the mean decorated length fraction $\bar{\phi}$, obtained by averaging over many runs. Thus, our results should be compared with data similarly averaged over many experiments. For this reason, the present comparison should be treated with caution; it is not a truly reliable estimate of the reaction times. However, a similar caution also applies to direct comparison of the data with the analytic solution(s) because these are also implicitly averaged.

We fit the experimental data for $[L_{max}(t) - L_{max}(0)]/L_{max}(0)$ directly to the rescaled simulation data $c\bar{\phi}(t + \tau_{lag})$ using a least squares method (See Fig. 4). The geometric scaling factor $c = a - 1$ relates the decorated length fraction to the change in maximum extended length. Both c and the lag time τ_{lag} are fit parameters. Furthermore, the simulation data is extended according to $\bar{\phi}(t < 0) = 0$ and $\bar{\phi}(t > 2\tau_{grow}) = 1$. There are two additional fit parameters, which may be thought of as τ_{nuc} and τ_{grow}, although we actually proceed as follows. The simulation data obtained for each value of τ_{nuc}/τ_{grow} is fitted to the experimental data using τ_{grow} as a single-fit parameter. Thus, the data set that gives the overall best fit identifies τ_{grow}, τ_{nuc}/τ_{grow}, and hence τ_{nuc}. The least squares residues for various values of τ_{nuc}/τ_{grow} are shown in Fig. 5. The best fit values obtained were $\tau_{lag} = 7.2$, $\tau_{nuc} = 2.9$, $\tau_{grow} = 11.5$ (all in minutes) and $c = 0.48$. These correspond to microscopic times of $\tau^*_{nuc} = 4.7 \times 10^4$ min per recA binding site and a total growth rate of each domain of $1/\tau^*_{grow} = 1400$ recA/min. Thus, the ratio $\tau_{nuc}/\tau_{grow} = ¼$ is somewhat less than unity. In spite of this, the quantitative agreement between the analytic solution in the fast nucleation limit and the simulation results is still not

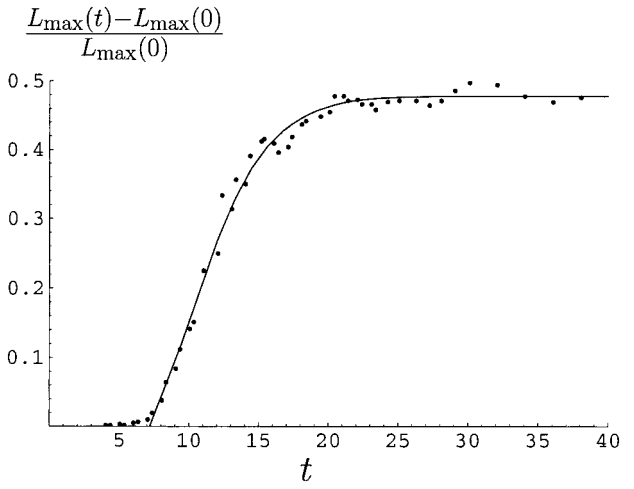

FIGURE 4 Best fit of the rescaled simulation results $c\bar{\phi}(t + \tau_{lag})$ to the experimental data for the rescaled maximum extended length of the DNA molecule $[L_{max}(t) - L_{max}(0)]/L_{max}(0)$ obtained by Shivashankar et al. (1999). The time axis is measured in minutes. The 4 fit parameters, with their best-fit values, were $\tau_{lag} = 7.2$, $\tau_{nuc} = 2.9$, $\tau_{grow} = 11.5$ (all in minutes) and the factor $c = 0.48$, by which $\bar{\phi}$ was rescaled.

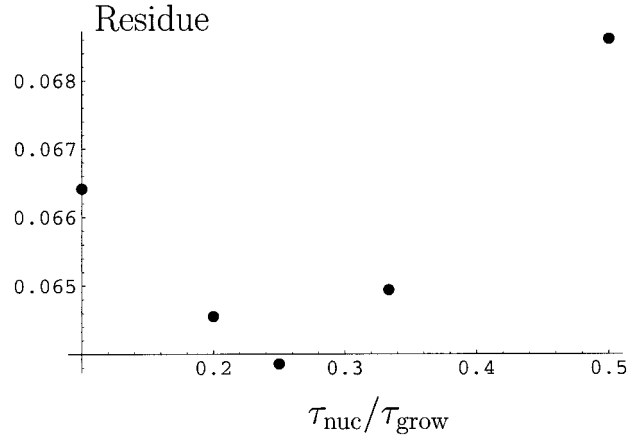

FIGURE 5 Plot of the least squares fit residue (*vertical axis*), in arbitrary units, obtained by fitting the experimental data (Shivashankar et al., 1999) to our simulation results for $\bar{\phi}(t)$. The comparison is made with simulation data obtained for several ratios of τ_{nuc}/τ_{grow}, as shown.

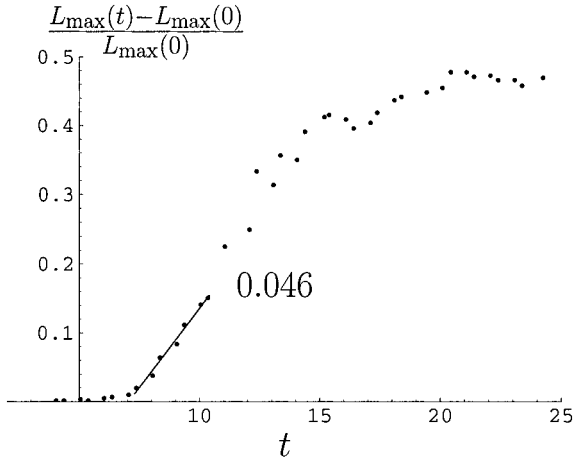

FIGURE 6 Plot of a linear regression fit for $[L_{max}(t) - L_{max}(0)]/L_{max}(0)$ over the 7 data points between $t = 7.4$ and $t = 10.4$ min, representing the early stages of coverage of the DNA by recA. The time axis is measured in minutes. The data reaches a plateau $[L_{max}(t) - L_{max}(0)]/L_{max}(0) = 0.48$ at late times. This corresponds to full coverage of the DNA by recA molecules. Very soon after coverage starts to increase, one expects a single domain to be present. The initial slope is 0.046 min^{-1}, which corresponds to a growth time $\tau_{grow} = 0.48/0.046 = 10.4$ min.

particularly good. It will be slightly poorer than that shown in Fig. 2 *E,* for which $\tau_{nuc}/\tau_{grow} = 1/5$. However, for comparison, we carry out a similar fit of Eq. 3 to the experimental data. The fit is 13% less good and yields the single time scale $\tau = \sqrt{\tau_{nuc}\tau_{grow}} = 5.0$ min with two additional fit parameters $\tau_{lag} = 5.7$ min and $c = 0.48$. As emphasized above, such a fit can yield no information about the ratio of the times τ_{nuc}/τ_{grow}. The ratio of the recA decorated to bare contour length per base pair, $a = 5.1/3.4 = 1.5$, agrees well with the estimate $1 + c = 1.48$ obtained by our fitting procedure.

The simulation results give an independent estimate of the nucleation time, which allows us, in principle, to determine whether the lag time τ_{lag} has a stochastic origin or whether it is an intrinsic mixing-related artifact. It is impossible to make a reliable statement regarding this issue from a single experimental data set, but we can say that it is not particularly unlikely that a lag time of $\tau_{lag} = 7.2$ might arise from a stochastic (Poisson) process with mean $\tau_{nuc} = 2.9$. One could make a more precise statement with data averaged over several experiments.

One additional source of information about the growth time can be identified from the data. To see this, note that the initial behavior is governed by the growth of the first domain according to $(d\phi/dt)|_{t\to 0} = 1/\tau_{grow}$. By fitting to the early data points, as shown in Fig. 6, we obtain an estimate of the growth time $\tau_{grow} = 10.4$ min, which is in reasonable agreement with the estimate of $\tau_{grow} = 11.5$ obtained by directly fitting to the simulation results.

Given an accurate estimate of the ratio τ_{nuc}/τ_{grow}, we would also be able to estimate the difference in energy barriers experienced by a protein binding at the end of a growing protein domain (a growth step) and one binding far from any previously bound proteins (a nucleation step). These will depend on the applied forces. Denoting the height of these energy barriers for the data set shown in Figs. 4 and 6 as E_g and E_n, a simple Arrhenius argument yields the rates for growth and nucleation per site as $1/(2\tau^*_{grow}) \propto \exp[-E_g/k_BT]$ and $1/(\tau^*_{nuc}) \propto \exp[-E_n/k_BT]$, respectively. The factor of 2 in the denominator of the growth rate arises from the assumption that there are two potential growth sites per domain, one at each end. Note that it is appropriate to use the microscopic time constants here, see Eqs. 9 and 10. These relations combine to produce

$$E_n - E_g = k_BT \log_e\left(\frac{\tau^*_{nuc}}{2\tau^*_{grow}}\right) = k_BT \log_e\left(\frac{\tau_{nuc}N^2}{2\tau_{grow}}\right), \quad (15)$$

where 49 kb λ-DNA possesses $N = 16{,}300$ recA binding sites. Using this, and our crude estimate $\tau_{nuc}/\tau_{grow} = 1/4$, we obtain $E_n - E_g = 17k_BT = 10$ kcal/M. This estimate is independent of the protein concentration in the reaction-limited regime. Separate estimates of E_n and E_g from τ_{nuc} and τ_{grow} would depend on the protein concentration and would require a more detailed model of the dynamics.

CONCLUSIONS

In conclusion, we have presented a technique for extracting both the characteristic nucleation and growth times for the formation of protein-decorated domains on a DNA molecule. We do this given only the time dependence of the DNA coverage.

Ideally, our technique requires data averaged over many identical experiments. However, for the single set of data available for the λ-DNA recA-ATPγs system, we crudely estimate that these times are $\tau_{nuc} = 2.9$ min, corresponding to 4.7×10^4 min per recA binding site, and $\tau_{grow} = 11.5$ min, corresponding to 1400 recA/min per domain. In principle, our technique also allows us to determine whether any experimental lag times are stochastic in origin or are mixing related artifacts.

The author thank Mario Feingold, G. V. Shivashankar, Albert Libchaber, and O. Krichevsky for numerous discussions and the initial inspiration for this work.

Support from the W. M. Keck foundation, the Burroughs Wellcome fund and the Royal Society is gratefully acknowledged.

REFERENCES

Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson. 1994. Molecular Biology of the Cell. Garland, New York.

Brenner, S. L., A. Zlotnick, and J. D. Griffith. 1988. RecA protein self-assembly—multiple discrete aggregation states. *J. Mol. Biol.* 204: 959–972.

Chu, S. 1991. Laser manipulation of atoms and particles. *Science.* 253: 861–866.

Di Capua, E., A. Engel, A. Stasiak, and T. Koller. 1982. Characterization of complexes between recA protein and duplex DNA by electron-microscopy. *J. Mol. Biol.* 157:87–103.

Dirac, P. A. M. 1982. The Principles of Quantum Mechanics. OUP, Oxford, U.K.

Egelman, E. H., and A. Stasiak. 1986. Structure of helical recA-DNA complexes formed in the presence of ATP-γ-s or ATP. *J. Mol. Biol.* 191:677–697.

Hegner, M., and C. Bustamante. 1998. Stretching individual recA-DNA filaments with optical tweezers. *Biophys. J.* 74:A150.

Hegner, M., S. B. Smith, and C. Bustamante. 1999. Polymerization and mechanical properties of single recA-DNA filaments. *Proc. Natl. Acad. Sci. USA.* 96:10109–10114.

Kellermayer, M. S. Z., S. B. Smith, H. L. Granzier, and C. Bustamante. 1997. Folding-unfolding transitions in single titin molecules characterized with laser tweezers. *Science.* 276:1112–1116.

Leger, J. F., J. Robert, L. Bourdieu, D. Chatenay, and J. F. Marko. 1998. RecA binding to a single double-stranded DNA molecule: a possible role of DNA conformational fluctuations. *Proc. Nat. Acad. Sci. USA.* 95: 12295–12299.

MacFarland, K. J., Q. Shan, R. B. Inman, and M. M. Cox. 1997. RecA as a motor protein—testing models for the role of ATP hydrolysis in DNA strand exchange. *J. Biol. Chem.* 272:17675–17685.

Marko, J. F., and E. D. Siggia. 1995. Stretching DNA. *Macromolecules.* 28:8759–8770.

Pugh, B. F., and M. M. Cox. 1987. Stable binding of recA protein to duplex DNA. *J. Biol. Chem.* 262:1326–1336.

Pugh, B. F., and M. M. Cox. 1988. General mechanism for recA protein binding to duplex DNA. *J. Mol. Biol.* 203:479–493.

Pugh, B. F., B. C. Shutte, and M. M. Cox. 1989. Extent of duplex DNA underwinding induced by recA protein binding in the presence of ATP. *J. Mol. Biol.* 205:487–492.

Register, J. C. III, and J. Griffith. 1985. The direction of recA protein assembly onto single stranded SNA is the same as the direction of strand assimilation during strand exchange. *J. Biol. Chem.* 260:12308–12312.

Rief, M., M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub. 1997. Reverse unfolding of individual titin immunoglobulin domains by AFM. *Science.* 276:1109–1111.

Roca, A. I., and M. M. Cox. 1990. The recA protein: structure and function. *Biochem. Mol. Biol.* 25:415–456.

Roca, A. I., and M. M. Cox. 1997. RecA protein: structure, function and role in recombinational DNA repair. *Prog. Nuc. Acid Res. Mol. Biol.* 56:129–223.

Rudnick, J., and R. Bruinsma. 1999. DNA–protein cooperative binding through variable-range elastic coupling. *Biophys J.* 76:1725–1733.

Shaner, S. L., J. Flory, and C. M. Radding. 1987. The distribution of *Escherichia coli* recA protein bound to duplex DNA with single stranded ends. *J. Biol. Chem.* 262:9220–9230.

Shivashankar, G. V., and A. Libchaber. 1997. Single DNA molecule grafting and manipulation using a combined atomic force microscope and an optical tweezer. *Appl. Phys. Lett.* 71:3727–3729.

Shivashankar, G. V., G. Stolovitzky, and A. Libchaber. 1998. Backscattering from a tethered bead as a probe of DNA flexibility. *Appl. Phys. Lett.* 73:291–293.

Shivashankar, G. V., and A. Libchaber. 1998. Single molecule observation of duplex DNA extension by recA. *Biophys. J.* 74:A242.

Shivashankar, G. V., M. Feingold, O. Krichevsky, and A. Libchaber. 1999. RecA polymerization on double-stranded DNA by using single molecule manipulation: the role of ATP hydrolysis. *Proc. Natl. Acad. Sci. USA.* 96:7916–7921.

Stasiak, A., E. Di Capua and T. Koller. 1981. Elongation of duplex DNA by recA protein. *J. Mol. Biol.* 151:557–563.

von Hippel, P. H., and J. D. McGhee .1972. DNA–protein interactions. *Annu. Rev. Biochem.* 41:231–300.