Kinetics of the Interaction between DNA and the Type IC Restriction Enzyme *Eco*R124II[†]

Jeremy J. Ramsden*[,‡] and Jürg Dreier[§]

Departments of Biophysical Chemistry and Microbiology, Biozentrum, CH-4056 Basel, Switzerland

Received September 11, 1995; Revised Manuscript Received December 8, 1995[⊗]

ABSTRACT: Optical waveguide mode spectroscopy was used to determine the binding constants characterizing the interaction of *Eco*R124II, a type IC restriction modification enzyme from *Salmonella typhimurium*, with DNA. The DNA is immobilized on the surface of an optical waveguide, and the enzyme is introduced in bulk solution flowing over the DNA under controlled hydrodynamic conditions. The binding kinetics of the protein to the DNA can be directly observed and the number of bound protein molecules per base pair determined to a high accuracy. Dissociation of the protein was measured by switching flowing protein to protein-free buffer. Binding to two different kinds of DNA, with and without the specific sequence recognized by *Eco*R124II, was investigated. Protein binding and dissociation ("nonspecific" binding), quantified by association and dissociation rate coefficients k_a and k_d, were the same for both types, but the DNA carrying the recognition site showed an additional process, "irreversible" association (i.e. dissociation was not observed on the time scale of the experiments) of the protein, quantified by a rate coefficient k_s. Some inferences regarding the mechanism of base pair searching are made from the measured k_a, k_d, and k_s values.

The binding of proteins to DNA lies at the heart of the regulation of life, from the earliest stages of development to the constant adjustment of protein expression to environmental demands. The accurate and reliable determination of the physicochemical binding parameters is a key step in understanding the mechanism of this regulation. In a binding process involving a DNA molecule thousands of base pairs long, a given protein binds to one specific sequence which may comprise at most a few tens of base pairs (bp). Whatever the details, the binding process must therefore comprise two distinct steps, which we may call respectively search and recognition: (i) general, low-affinity binding at randomly selected sites to bring the protein into contact with the DNA and (ii) high-affinity binding to the specific sequence (sometimes called the "recognition sequence"). In the absence of long range forces steering the protein to its recognition sequence, it must be presumed that the protein first associates with DNA at a randomly selected site. A major controversial issue is whether the protein then quickly dissociates from any sequence other than the recognition one, leaving it free to diffuse further in the bulk solution and try another sequence, or whether it diffuses ("tracks") along the DNA chain until it encounters the recognition sequence (Winter et al., 1981; Berg et al., 1982).

In this work, our aim is to measure the binding and dissociation kinetics more precisely and comprehensively than hitherto, in order to assess whether the pathway to the recognition sequence can be determined on the basis of kinetic data alone, and to use the values of the measured rate constants to gain insight into the molecular details of the process. Previous work has been based on techniques such as nitrocellulose filter binding (Riggs et al., 1970), centrifugation (Gilbert & Mueller-Hill, 1967), and gel retardation assays (Garner & Revzin, 1981; Fried & Crothers, 1981), electron microscopy (Dodson & Echols, 1991), footprinting (Galas & Schmitz, 1978; Brenowitz et al., 1986), cross-linking protein to DNA (Hockensmith et al., 1991), and methods which rely on a change of some spectroscopic property of either DNA or the protein upon complex formation (Lohman & Bujalowski, 1991). With these assays, it is often not convenient to observe dissociation of protein from the DNA. Dissociation rate coefficients may be calculated as the quotient of the directly measured association rate coefficient and the measured equilibrium coefficient, but care must be taken to ensure that, for the equilibrium measurement, the system really is at equilibrium. Equilibration times may be lengthy, and this is not always a trivial requirement.

To overcome some of these difficulties, various optical methods have recently appeared in the literature (Ramsden, 1994). The most popular of these so far seems to be surface plasmon resonance (Ramsden, 1994). However, surface plasmon resonance (SPR)[2] suffers from the disadvantage of only measuring one optical parameter, whereas to define the absolute number of bound molecules, two optical parameters are necessary (Ramsden, 1994). Moreover, most applications of SPR to biological systems appear to require that the binding reaction take place in a dextran matrix (e.g. the Pharmacia Biacore system). Due to the strong perturbations of the chemical potentials of reactants and products in such a matrix (Laurent, 1995; Lindner & Ralston, 1995), the

[†] Supported by the Commission for the Promotion of Scientific Research, Berne (J.J.R.), and the Swiss National Science Foundation (T. A. Bickle).

[‡] Department of Biophysical Chemistry.

[§] Department of Microbiology.

[⊗] Abstract published in *Advance ACS Abstracts,* March 1, 1996.

[1] This enzyme has previously been called *Sty*R124/3I or *Eco*R124/3I.

[2] Abbreviations: AdoMet, *S*-adenosylmethionine; ATP, adenosine triphosphate; OWMS, optical waveguide mode spectroscopy; SPR, surface plasmon resonance.

0006-2960/96/0435-3746$12.00/0

measured binding constants could be 1 or more orders of magnitude greater than the true values.

Hence, we have chosen optical waveguide mode spectroscopy (OWMS) in order to accurately measure the binding and dissociation of small numbers of molecules to the DNA without labeling or other perturbing influences. The principle of the method is that the binding of proteins to a conformationally flexible DNA layer immobilized at the surface of an optical waveguide without the use of dextran perturbs the evanescent field of light modes guided in the optical waveguide (Ramsden, 1993). The number of bound molecules is determined from the measurement of the effective refractive indices characterizing at least two modes. OWMS has the further advantage of being inherently 1 order of magnitude more sensitive than SPR (Lukosz, 1993).

Specifically, we have investigated the DNA binding of *Eco*R124II, a type IC restriction modification system (R/M) from *Salmonella typhimurium* (Meynell & Datta, 1966; Firman et al., 1983; Glover et al., 1983). Type I R/M systems show several enzymatic activities, namely N^6A-methylation, DNA cleavage, ATP hydrolysis, and possibly helicase activity [reviewed in Bickle (1993) and Gorbalenya and Koonin (1991)]. *Eco*R124II methylates its recognition sequence GAA(N_7)RTCG (N = any nucleotide, R = either purine) at specific adenines in both DNA strands (Price et al., 1987a). The methylation state of the recognition sequence determines what events follow binding of the protein to that sequence [reviewed in Bickle (1993)]; in the case of *Eco*R124II, nonmethylated DNA is cut randomly at distances up to several thousand base pairs away from the recognition site, provided that the cofactors *S*-adenosylmethionine (AdoMet), Mg^{2+}, and ATP are present. Hemimethylated DNA (which occurs after DNA replication) is methylated in the second strand; Mg^{2+} and AdoMet are required as cofactors. Fully methylated DNA is not a substrate, and the enzyme complex dissociates from it (Price et al., 1987b). *In vitro*, if no ATP is available, both DNA strands are methylated. In our experiments, DNA which either lacks any specific recognition site (nonspecific DNA) or has nonmethylated sites was used. ATP was omitted in order to exclude restriction and retain methylation activity only.

Before the protein binds to the recognition sequence and initiates activity, however, an even more elementary (relative to the enzymatic activity) process has to have occurred: unselective binding to any sequence, which releases the enzyme complex if it is *not* the specific recognition sequence. Understanding how the search efficiency can be optimized is a major goal of this work.

MATERIALS AND METHODS

Plasmids. pEKU60 was constructed by digestion of pBR322 (Watson, 1988) with *Hin*dIII and *Ava*I, followed by blunt-ending and religation of the 2.971-kilobase (kb) vector DNA. This procedure removed the single specific *Eco*R124II recognition site of pBR322 at position 1377−1390 (Kulik 1995). Both plasmids were propagated in the *Escherichia coli* K12 strain DH5α (Woodcock et al., 1989) and purified by the Qiagen plasmid purification protocol (Qiagen Inc.). They were phenol extracted and ethanol precipitated to remove any residual proteins (Maniatis et al., 1989). pBR322 has 4361 base pairs, and pEKU60 has 2971.

Protein Purification. Purification of *Eco*R124II was monitored by Western blot analysis (Towbin et al., 1979) of column fractions using antibodies raised against *Eco*R124I enzyme. *E. coli* WA921 cells (Wood, 1966) containing pUNG30 (Firman et al., 1985) (expressing *Eco*R124II) were grown to the stationary phase in Luria broth (LB) supplemented with 200 μg/mL ampicillin at 37 °C in a 20 L fermenter. All the following steps were carried out at 4 °C. A 20 g batch of the cell paste was resuspended in 30 mL of buffer A [20 mM Tris-HCl (pH 8.0), 50 mM NaCl, 5 mM $MgCl_2$, and 7 mM 2-mercaptoethanol] and broken by two passages through a French press cell at 20 000 psi. The extract was clarified by low-speed centrifugation (Sorvall SS34 rotor; 12 000 rpm; 20 min) and subsequent high-speed centrifugation (International Centrifuge A-237 rotor; 38 000 rpm; 2 h). The nucleic acids were precipitated by making the supernatant up to 0.4 M NaCl and adding neutralized polyethyleneimine up to 0.4%. After another low-speed centrifugation (as above), the proteins in the supernatant were precipitated in 70% saturated ammonium sulfate, redissolved in 20 mL of buffer B [20 mM Tris-HCl (pH 8.0), 50 mM NaCl, 0.5 mM EDTA, and 7 mM 2-mercaptoethanol], and dialyzed against the same buffer. The proteins were loaded onto a 70 mL DEAE-Sephacel (Pharmacia) column, washed with buffer B, and eluted with a 50 to 500 mM NaCl gradient. Fractions containing enzyme were pooled, dialyzed against buffer B, and loaded onto an 18 mL heparin-Sepharose (Pharmacia) column. After the column was washed with buffer B, elution was carried out with a 50 to 1000 mM NaCl gradient. The pooled enzyme-containing fractions were further purified and desalted with a Superose 6 HR 30/10 FPLC column (Pharmacia), applied to a 9 mL Affi-Gel Blue column (Bio-Rad), and eluted with a 50 to 1000 mM NaCl gradient. Enzyme-containing fractions were pooled, concentrated with Centrikon-30 devices (Amicon), and stored at −20 °C [modified from Bickle and Streiff (1983) and Gubler and Bickle (1991)].

Protein Characterization. The concentration was determined by dabsyl chloride amino acid analysis. In order to determine a suitable concentration range for the binding measurements, we estimated the *Eco*R124II concentration in the cytoplasm. From 20 L of culture, we obtained 7×10^8 cells/mL, from which we were able to extract 4.15 mg of purified protein. Taking a molecular mass of 385 kDa,[3] this corresponds to an average of 450 molecules per cell. However, in our culture, the protein is translated from 15−20 copies of the relevant genes (in pUNG30), whereas in wild type *E. coli*, there are only a few copies of these genes. If the protein is overexpressed 10-fold in the culture, we expect 40−50 molecules of *Eco*R124II in a wild type cell. Microscopic examination showed the cells to have an average volume of about 0.8 μm^3; the protein concentration is therefore about 100 nM. Hence, stock solutions (ca. 2 mg/mL) were diluted (immediately before each experiment) to a concentration of around 20 μg/mL, equivalent to 50−60 nM (depending on the actual stoichiometry of the complex[3]). The diffusion coefficient D of the protein was estimated as $(2.8 \pm 0.4) \times 10^{-7}$ cm^2/s using the IOS-1 instrument (see

[3] *Eco*R124II is composed of three types of subunits, M, S, and R, with molecular masses of 55, 43, and 116 kDa, respectively (Price et al., 1987b, 1989). The stoichiometry of *Eco*R124II is M_2SR_n, where n is probably 1 or 2. $n = 2$ yields a molecular mass $M_r = 385$ kDa; the mass of one protein molecule $m = 0.64 \times 10^{-12}$ μg.

Optical Assay below) as described elsewhere (J. J. Ramsden, in preparation).

Optical Assay. Planar $Si(Ti)O_2$ optical waveguides incorporating a grating coupler (grating constant $\Lambda = 433.33$ nm) were obtained from Artificial Sensing Instruments, Zürich, Switzerland (type 1400), and used without pretreatment. They were mounted in an IOS-1 integrated optics scanner (Artificial Sensing Instruments, Zürich), with which the effective refractive indices N_{TE} and N_{TM} of the waveguide could be measured to a precision of $\pm 2 \times 10^{-6}$; 10 adsorbed proteins per square micrometer would increase N by about 5×10^{-6}. The measurement aperture was defined by the diameter (0.8 mm) of the external beam (He−Ne laser, wavelength $\lambda = 632.82$ nm) used to excite the guided modes via the grating coupler (Tiefenthaler & Lukosz, 1989; Tiefenthaler, 1992; Ramsden, 1993). The temperature was stabilized at 25.2 ± 0.5 °C by enclosing the entire instrumentation in a thermostated cabinet and monitored continuously by a Pt-100 resistance thermometer embedded in the measuring head close to the grating.

Immobilization of DNA. Preliminary experiments showed that DNA did not adsorb onto untreated $Si(Ti)O_2$ waveguides, probably because both waveguides and DNA are strongly negatively charged at pH 8. Hence, a novel method recently reported by Lvov et al. (1993) was used to immobilize a layer of DNA onto the waveguide surface. First, a layer of the polycation polyallylamine was deposited onto the waveguide by placing a 150 μL drop of polyallylamine hydrochloride (Aldrich, "high molecular weight"), dissolved at a concentration of 0.01 M with respect to monomer in 2 M NaBr, over the grating coupler region of the waveguide and allowing it to stand for 20 min in a covered petri dish. This procedure results in a positively charged layer about 1 nm thick on the waveguide (Ramsden et al., 1995a,b). Following rinsing with TE buffer [10 mM Tris-HCl (pH 8.0) and 1 mM EDTA], the waveguide was placed in the IOS-1 instrument and a small cylindrical cuvette sealed over the grating region, the waveguide forming the bottom of the cuvette. The cuvette was first filled with TE buffer, and as soon as the base line signal was seen to be stable, the buffer was replaced by 200 μL of DNA solution freshly diluted to a concentration of 25 μg/mL. This solution was periodically (typically every 20−30 min) agitated by gently sucking and expelling 100 μL of the solution three times with a pipette. After ca. 3 h, no more DNA was being deposited, even after agitation. In general, we could deposit about 0.1 $\mu g/cm^2$; since the cuvette has a diameter of 1 cm, about 80 ng of DNA was deposited (i.e. our DNA solution essentially suffered no depletion and could be reused). The resulting DNA layer is about 3.3 nm thick (Lvov et al., 1993). From a detailed optical analysis of these polyelectrolyte layers, it has been concluded that the molecules have a looped configuration attached by a few random sites (Ramsden et al., 1995a,b). We believe that this is a favorable situation allowing much of the DNA to retain its native conformation, unlike biopolymers rigidly attached to smooth planar substrates for the sake of atomic force microscopy (Karrasch et al., 1993). The DNA was shown to be irreversibly bound by subsequently rinsing with TE, upon which there was no change in the quantity bound to the polyallylamine.

Protein Binding. Following DNA immobilization, the waveguide was removed from the IOS-1 instrument and equilibrated for 3−4 h in Mg200 buffer [10 mM Tris-HCl (pH 8.0), 10 mM $MgSO_4$, and 200 mM potassium glutamate] for 3−4 h. This composition was motivated by the idea that potassium glutamate is the principal osmolyte in an *E. coli* cell (Richey et al., 1987). We found that 200 mM potassium glutamate was an optimum concentration at which no nonspecifically bound *Eco*R124II was left after the washing step; higher salt concentrations resulted in no protein at all being bound, and at lower concentrations, even the nonspecific binding became irreversible. The waveguide was then replaced in the instrument, and a small flow-through cuvette of tubular cross section was sealed over the grating region, again such that the waveguide formed one wall of the cuvette. Solutions were drawn through the cuvette by a downstream peristaltic pump. After a base line was established with Mg200 buffer, protein freshly diluted with Mg200 to a concentration c_b of 22 μg/mL and mixed with freshly diluted 4.7 μM AdoMet flowed through the cuvette at a precisely monitored rate. In order to determine the rate of dissociation, the flow through the cuvette was later switched back to pure Mg200 buffer. All experiments were carried out in the absence of ATP in order to restrict the observable events to binding, recognition, and dissociation.

Data Evaluation. Prior to DNA deposition, the waveguide is constituted from three layers, the support S of optical glass (refractive index $n_S = 1.531\,51$), the high-refractive index waveguiding layer F [$Si(Ti)O_2$ + polyallylamine] of thickness d_F (≈180 nm) and refractive index n_F (≈1.8), and the cover medium (buffer) C of refractive index n_C. n_C and n_S are known,[4] but n_F and d_F differ slightly from waveguide to waveguide. These parameters were determined from the effective refractive indices of the zeroth order transverse electric and transverse magnetic modes, N_{TE} and N_{TM}, respectively, measured for the base line by solving the three-layer mode equations (Tiefenthaler & Lukosz, 1989).[5] The deposited DNA constitutes a fourth layer designated A (denoting adlayer); the surface loading Γ of the DNA is defined by two parameters, refractive index n_A and thickness d_A, according to de Feijter et al. (1978):

$$\Gamma = d_A(n_A - n_C)/(dn/dc) \qquad (1)$$

Herce, we need to measure two effective refractive indices, N_{TE} and N_{TM}. The adlayer parameters n_A and d_A are calculated from the measured N_{TE} and N_{TM} by solving the linearized four-layer mode equations (Tiefenthaler & Lukosz, 1989; Ramsden, 1993, 1994):

$$\frac{2\pi}{\lambda}(n_F^2 - N^2)^{1/2}\left(d_F + d_A\frac{n_A^2 - n_C^2}{n_F^2 - n_C^2}\left[\frac{(N/n_C)^2 + (N/n_A)^2 - 1}{(N/n_C)^2 + (N/n_F)^2 - 1}\right]^{\rho}\right) = \arctan\left[\left(\frac{n_F}{n_S}\right)^{2\rho}\left(\frac{N^2 - n_S^2}{n_F^2 - N^2}\right)^{1/2}\right] + \arctan\left[\left(\frac{n_F}{n_C}\right)^{2\rho}\left(\frac{N^2 - n_C^2}{n_F^2 - N^2}\right)^{1/2}\right] \qquad (2)$$

where $\rho = 0$ for N_{TE} and $\rho = 1$ for N_{TM}.

We determined the refractive increment dn/dc for our DNA using an LI3 Rayleigh interferometer (Carl Zeiss, Jena); the DNA was lyophilized and redissolved in the reference buffer.

[4] We measured n_C to a precision of $\pm 10^{-5}$ using an LI3 Rayleigh interferometer (Carl Zeiss, Jena). Since $\partial N/\partial n_C \approx 0.1$, this is accurate enough for the precision ($\pm 2 \times 10^{-6}$) with which we can currently measure N.

[5] The three-layer mode equations can be obtained from eq 2 by setting $d_A = 0$.

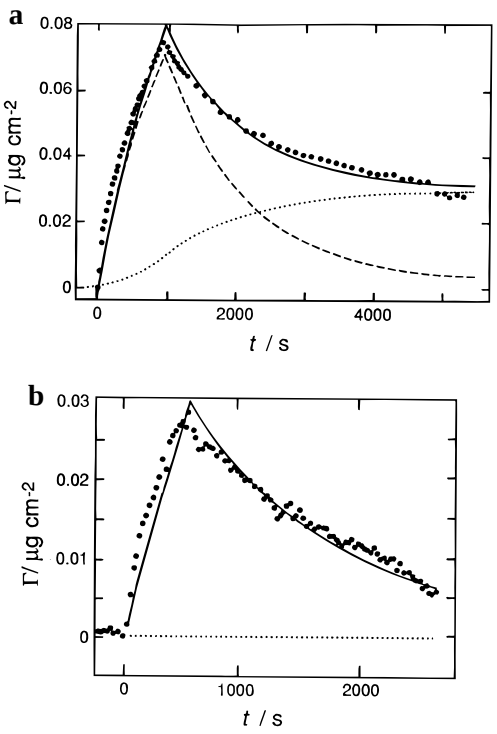

FIGURE 1: Typical results. ●, experimental points. Solid line, eq 3 fitted to the curve. (a) Binding to pBR322, $s = 1.5 \times 10^{-10}$ mol/cm^2, $c_b = 25$ μg/cm^3 (dashed line, calculated concentration of nonspecifically bound protein; dotted line, calculated protein which has found a recognition site). (b) Binding to pEKU60, $s = 2.4 \times 10^{-10}$ mol/cm^2, $c_b = 19$ μg/cm^3.

With this procedure, we cannot guarantee that no additional salt, which would inflate the apparent dn/dc, is present. We determined a value of 0.26 mL/g, compared with a typical protein value of 0.18 mL/g (Sober, 1970).

Following establishment of the new base line with buffer Mg200, n_F and d_F were calculated afresh as above,[6] and the mass of protein binding to the DNA to form a new adlayer A was calculated from eqs 1 and 2 using dn/dc = 0.18 mL/g.

Buffers and Other Reagents. The buffers were made up from analytical grade reagents. Before use, they were autoclaved and filtered through Millipore 200 nm filters. AdoMet (Sigma) was dissolved in potassium phosphate buffer (10 mM, pH 7.0) to approximately 100 mM. This solution was loaded onto a 0.5 mL BioRex (BioRad) column (mesh range of 100−200) and washed twice with 1 mL of the same phosphate buffer and twice with 1 mL of 250 mM acetic acid. AdoMet was eluted with 5 × 200 μL of 4 M acetic acid. Peak fractions were pooled, and the concentration was determined from the optical density at 260 nm, using $\epsilon = 15\,000$ M^{-1} cm^{-1}.

RESULTS

Figure 1 shows typical graphs of the kinetics of binding of *Eco*R124II to (a) DNA containing one copy of the recognition sequence GAA(N_7)RTCG (pBR322) and (b) DNA from which the recognition sequence had been excised (pEKU60). The association kinetics are identical, but whereas all the bound protein dissociates from pEKU60 during washing, a residual quantity, irreversibly bound on the time scale of the experiment, remains in the case of pBR322. This is assigned to protein bound to the recognition site.

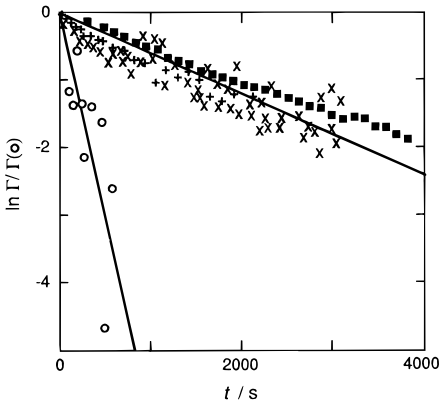

FIGURE 2: Dissociation at different buffer flow rates, plotted according to ln $\Gamma(t)/\Gamma(0)$ vs time (eq 14), where $\Gamma(0)$ corresponds to $t = 0$ (the start of washing), except that the curve for dissociation from pBR322 (incomplete dissociation) was plotted as $\ln[\Gamma(t) - \Gamma_s]/[\Gamma(0) - \Gamma_s]$ against time. The slopes of the curves are given in Table 2; the dissociation using up to and including flow rates of 60 μL/s all seemed to fall on about the same slope (drawn), viz. 6.0×10^{-4} s^{-1}: +, $F = 1.8$ μL/s (pEKU60); ×, $F = 60$ μL/s (pEKU60); ■, $F = 1.8$ μL/s (pBR322); ○, $F = 100$ μL/s (pEKU60).

These results concur with the concept, adumbrated in the introductory section, that, irrespective of whether the DNA contains a recognition sequence, the initial association of *Eco*R124II, characterized by a rate coefficient k_a, is "nonspecific". The bound mass Γ we actually measure is the sum of nonspecifically and specifically bound masses, i.e.

$$\Gamma = \Gamma_n + \Gamma_s \tag{3}$$

Figure 2 shows that lin−log plots of the washing phase always gave straight lines. Hence, dissociation is a Poisson process determined only by a rate coefficient k_d and the amount of protein molecules nonspecifically associated to the DNA. In the case of pBR322, encounter of and binding to the recognition sequence competes with dissociation; in our kinetic model, we characterize recognition by a rate coefficient k_s, which should be proportional to the ratio of base pairs comprising the recognition sequence to the total number of base pairs in the DNA molecule. The interpretation of k_s, which we determine experimentally, is one of the major tasks of this paper.

The rate dΓ_n/dt of nonspecific adsorption is given by (see Figure 3)

$$d\Gamma_n/dt = c_1 s k_a - \Gamma_n k_d - \Gamma_n(1 - \theta_s)k_s \tag{4}$$

The first term on the right-hand side represents the flux of *Eco*R124II from a zone just above the surface [where the concentration is $c_1 (\leq c_b)$] to the DNA, the second term its dissociation back into the solution, and the third term the process of recognition. The DNA is considered to be an array of one-dimensional molecules with a total density s of possible nonspecific binding sites. Since one base pair has an average mass of 650 g/mol, we have, in moles of base pairs per squared centimeter,

[6] Strictly speaking, each layer should be separately reckoned and the five- or six-layer mode equations (Ramsden, 1993) solved to determine the surface loading due to the ultimate adlayer, but the small error introduced by subsuming previously deposited layers into the F layer was checked to be less than the overall uncertainty in our measurements.

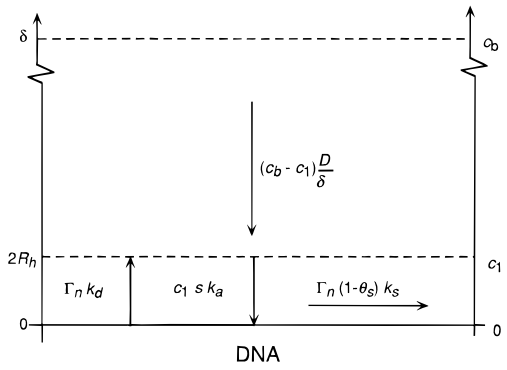

FIGURE 3: Reaction scheme described by eqs 4, 6, and 10.

$$s = \Gamma_{DNA}/650 \tag{5}$$

when Γ_{DNA} is measured in grams per squared centimeter. We apply the steady state approximation to the rate of change of c_1, setting

$$dc_1/dt = (c_b - c_1)D/\delta - c_1 s k_a + \Gamma_n k_d \tag{6}$$

equal to zero. Here, the first term on the right-hand side represents the flux of *Eco*R124II from the bulk solution through the diffusion boundary layer of thickness δ to the zone where the concentration is c_1. D is the diffusion coefficient of the protein. With $dc_1/dt \approx 0$, we obtain from eq 6

$$c_1 = \frac{c_b D/\delta + \Gamma_n k_d}{D/\delta + k_a s} \tag{7}$$

with which we can eliminate c_1 from eq 4. Since we have a well-defined hydrodynamic regime in our cuvette, we can use the expression of Levich (1962) for δ:

$$\delta = (3/2)^{2/3}(DxRA/F)^{1/3} \tag{8}$$

where x, R, and A are respectively the distance from the inlet to the measuring aperture (3.5 mm), the radius (1 mm), and the cross-sectional area (1.7 mm^2), and F is the flow rate.

Note that we do not explicitly consider saturation of the DNA by nonspecifically bound enzyme. This is because in our experiments the proportion of s occupied at any instant was very small. Even at the highest values of Γ measured, there were on average fewer than 10 protein molecules on each plasmid; it is unlikely that they interfere with one another. On the other hand, since there is only one recognition site per molecule of pBR322, we have to take its occupancy into account, which we do by defining θ_s, the average occupancy of the recognition site, by

$$\theta_s = \frac{\text{moles of irreversibly bound protein}}{\text{moles of recognition sites}} = \frac{\Gamma_s \times 4361}{sM_r} \tag{9}$$

The rate of specific binding to the recognition site is (see Figure 3)

$$d\Gamma_s/dt = \Gamma_n(1 - \theta_s)k_s \tag{10}$$

To determine the unknown parameters k_a, k_d, and k_s, eqs 4 and 10 were numerically integrated and fitted to the experimental data given by eq 3. Fitted parameters are given in Table 1.

Table 1: Kinetic Parameters for Protein−DNA Binding, Determined by Fitting the Experimental Data (Represented by eq 3) to eqs 4 and 10 as Described in the Text[a]

DNA	k_a (M^{-1} s^{-1})	k_d (s^{-1})	K_a (M^{-1})	k_s (s^{-1})
pEKU60	1.6×10^4	8×10^{-4}	2×10^7	0.0
pBR322	3.0×10^4	7.2×10^{-4}	4.2×10^7	2.3×10^{-4}

[a] A mass m of 0.64×10^{-12} μg was taken, i.e. corresponding to the stoichiometry M_2SR_2, but the choice of the other possible stoichiometry made an almost negligible difference to the fitted kinetic parameters. The nonspecific association equilibrium constant K_a is defined as $K_a = k_a/k_d$.

DISCUSSION

Equilibrium Occupation (Nonspecific). As expected, pBR322 and pEKU60 have similar nonspecific binding parameters. The value of K_a (defined as k_a/k_d) $\approx 3 \times 10^7$ M^{-1}. We might compare it with Taylor et al.'s (1992) specific binding constant of 10^8 M^{-1}, evaluated from gel retardation experiments with oligonucleotides containing the specific recognition sequence, although Winter et al. (1981) have shown that the length of the DNA significantly influences the binding. It should also be pointed out that Taylor et al. measured in the presence of chloride; Leirmo et al. (1987) and Richey et al. (1987) have found that glutamate slightly enhances binding compared with chloride.

The number of enzyme molecules simultaneously nonspecifically occupying one DNA molecule can be determined by noting that the steady solution (as $t \rightarrow \infty$) of eq 4 with $k_s = 0$ is

$$\Gamma_{eqm} = c_b s k_a/k_d \tag{11}$$

from which the equilibrium number of molecules ν per base pair is

$$\nu_{eqm} = \frac{c_b k_a}{M_r k_d} \tag{12}$$

From the data in Table 1, at a bulk concentration of 19 μg/mL, we find $\nu_{eqm} = 1$ molecule per 1010 base pairs or 3 molecules per molecule of pEKU60. This is sufficiently small to justify our neglect of mutual interference between protein molecules adsorbed on the DNA.

Stoichiometry of Recognition. We define this stoichiometry S as the final number of molecules of irreversibly bound *Eco*R124II per recognition site, i.e.

$$S = \frac{\Gamma_{s,fin} \times 4361}{sM_r} \tag{13}$$

For the result in Figure 1a, for example, $\Gamma_s = 0.026$ μg/cm^2, from which $S = 2.1$, taking $M_r = 385$ kDa.[3] A value of $S = 2$ is corroborated by our other experiments [the alternative composition M_2SR ($M_r = 296$ kDa) would give $S = 2.7$]. In comparison, Taylor et al. (1992) determined $S = 1$, but this was for oligonucleotides.

The main source of uncertainty in our value for S is the value of dn/dc for DNA. As pointed out in Materials and Methods, the presence of unknown amounts of salt in the lyophilized material could have led to a significant overestimation of the amount of adsorbed DNA. Taking the well-established typical value for protein dn/dc, 0.18 mL/g, as an estimate for DNA, we obtain $S = 1.4$. Until a more reliable

Table 2: Measured and Calculated Dissociation Rates from Nonspecific DNA (pEKU60), except Where Indicated[a]

F (μL s^{-1})	δ (μm)	s (pmol cm^{-2})	$k_d/(sk_a\delta/D + 1)$ (s^{-1})	measured slope (s^{-1})
1.8	91	245	-8×10^{-4}	-5.0×10^{-4}
20	41	300	-8×10^{-4}	-6.0×10^{-4}
60	28	300	-8×10^{-4}	-7.2×10^{-4}
100	24	300	-8×10^{-4}	-6.2×10^{-3}
pBR322				
1.8	91	140	-7.2×10^{-4}	-4.9×10^{-4}

[a] The rates were calculated using rate coefficients from Table 1. δ was calculated from eq 8.

measurement of dn/dc is available, we would like to be rather cautious regarding the stoichiometry. Fortunately, the kinetic parameters we have determined are robust, and their reliability does not depend upon the value of dn/dc.

Anatomy of Dissociation. Taking rebinding into account, eq 4 with $k_s = 0$ is solved with boundary conditions $c_b = 0$ at $t = 0$ to yield

$$\ln \frac{\Gamma_n}{\Gamma_0} = \frac{-k_d t}{sk_a\delta/D + 1} \tag{14}$$

where Γ_0 is the amount of protein bound to the DNA just before the initiation of dissociation. Table 2 gives the slopes of dissociation curves measured at different buffer flow rates F and plotted in Figure 2. According to eq 8, δ decreases with increasing F. If $k_d \gg sk_a\delta/D$, however, as we predict from our experimental data (Table 1), there should be no change in the dissociation rate over the range of F we investigated, which was true except at the highest flow rate, 100 μL/s, at which dissociation was almost 10 times faster. This may, however, be an artifact of our pumping arrangements; we used a commercial peristaltic pump in which the rollers run in a circular cavity, and the flow rate is not strictly uniform. The highest F involved running the pump as fast as possible, and these fluctuations were exacerbated by the mechanical instability of the pump, resulting in a significantly lower signal/noise ratio.

Significance of the Rate Coefficients. The basic question we seek to answer is as follows: how can *E. coli* ensure that it will survive following infection by phage? If the infected cell is to survive (taking phage λ as a model system), then the phage DNA must be destroyed before it has been replicated, i.e. within 15 min [a lytic cycle takes about 45 min (Friedman & Gottesman, 1983) and the phage burst follows the lysogenic induction after about 30 min (Bailone, 1979)]. From the point of view of the bacterial population as a whole, however, it is sufficient to ensure that the infected cell does not lyse and release assembled viruses, which allows it somewhat longer, around 45 min (Friedman & Gottesman, 1983), to act. We formulate a more specific question: how many *Eco*R124II molecules does the cell need to ensure that the infecting virus is thwarted? We attempt to answer it using the kinetic parameters we have determined for the interaction of *Eco*R124II with our model DNA.

As mentioned in the introductory section, it is not known how the protein finds its recognition site. Winter et al. (1981) inferred from investigations on the *lac* repressor−operator system that the protein binds to the DNA at a randomly selected site after diffusing through the cytoplasm and slides along (i.e. diffuses in one dimension) until it finds its specific binding site, benefiting from the reduction of dimensionality from three to one [Adam & Delbrück, 1968; but see also McCloskey and Poo (1986)]. The opposite viewpoint is that the protein does all its diffusing in the three-dimensional bulk solution; searching consists of binding to one randomly chosen site followed by dissociation if that site is not recognized and further diffusion in the bulk until the next random encounter with the DNA.

We accept that diffusion via the bulk is too inefficient to be the mechanism.[7] On the other hand, the presence of obstacles (especially other proteins) must surely hinder a one-dimensional random walk along a DNA molecule. A more realistic model for essentially one-dimensional tracking along the DNA molecule is dissociation followed by diffusion in the vicinity of the DNA; in solution, diffusive motion engenders strong correlations between consecutive sites of attachment, especially when these sites are arranged in a line (Luthi et al., in preparation). This is true even without explicitly taking into account the solvent molecules, which would further enhance the effect (Rabinowitch, 1937). We therefore reformulate the binding problem as follows; the bulk should be divided into two zones: the region near the DNA, well within the diffusion boundary layer, in which binding attempts will be correlated and the more distant regions in which binding attempts will be uncorrelated. Dissociation is defined as removal of the protein from the near-surface zone to the bulk zone from which association attempts are no longer correlated. During the period of association τ [which is, according to our results (Figure 2), Poisson-distributed with mean $\tau = 1/k_d$], the protein essentially executes a random walk along the DNA; microscopically, it probably does dissociate from the DNA and diffuse a little way along to the next site sampled,[8] but at the submolecular level, there is probably a continuum of degrees of association between associated and dissociated, involving various forces (van der Waals, electrostatic, hydrogen bonds, hydration, etc.) and in the sense of consecutively sampled sites being spatially correlated, is still associated. We denote the rate of moving from one site to another as u sites per second. The mean number of sites visited before dissociation, $\langle n_d \rangle$, is then simply

$$\langle n_d \rangle = u/k_d \tag{15}$$

but the number n_v of *distinct* sites visited is smaller, since the walk is undirected and much back-tracking occurs. Montroll (1969) has shown that, for a chain consisting of S sites, $S^2/6$ steps will on average be needed to reach a single recognition site. On pBR322, this event, characterized by k_s, occurs with a frequency of 2.3×10^{-4} s^{-1}. We have defined k_s as the conditional probability that an associated protein molecule will encounter and bind to a recognition site, i.e. it will have taken $4361^2/6 = 3 \times 10^6$ steps, from which it follows that the time for one step is $1/(k_s \times 3 \times 10^6) \approx 1.5$ ms $= 1/u$. Hence, $\langle n_d \rangle$, the average number of steps before dissociation, is 9×10^5 (from eq 15), taking about 20 min.

[7] Unless some long range force directs the protein toward the binding site. However, the only such force which would appear to be a candidate is electrostatic, and at the high salt concentrations encountered in the cell and in our experiments, they are highly effectively screened at ranges beyond a few nanometers.

[8] Note also the discussion of Winter et al. (1981).

We can now use these numbers to answer the question posed at the beginning of this subsection. We are aware that there must be many other processes in the cell running in parallel with restriction, and which may influence it. However, not least because we presently know so little about these other processes, we neglect them. The phase λ DNA contains 15 irregularly spaced recognition sites for *Eco*R124II; i.e. it is divided into segments bounded at each end by a recognition site. We wish to know how many steps are required to reach either end. This problem is simply the classical ruin problem (Feller, 1968). The expected duration D_z of a random walk on a line of length a with absorbing barriers at 0 and a for a walker starting at a position z along the line ($0 \leq z \leq a$) is

$$D_z = z(a - z) \tag{16}$$

and therefore, on average for a walker starting at any point z on the interval (0, a), by integrating eq 16 from 0 to a and dividing by a, we obtain for the mean number $\langle n_r \rangle$ of steps for a walker to encounter the recognition site:

$$\langle n_r \rangle = a^2/6 \tag{17}$$

The mean interval between recognition sites is 3233 base pairs; setting $a = 3233$ in eq 17, we obtain $\langle\langle n_r \rangle\rangle = 1.74 \times 10^6$ steps, the double brackets indicating averaging over a random starting point within an interval between two recognition sites and over the different intervals. From our above value for $1/u = 1.5$ ms, a single restriction enzyme would require, on average, about 45 min to find and cut the plasmid. This is clearly too long. The time may be shortened by allowing several enzymes, say a number ν, to search simultaneously. Although for large ν the same ground is repeatedly covered by different walkers, and the increase in the number of sites searched is subproportional to ν (Larralde et al., 1992), for small ν, we can take the increase to be linearly proportional, i.e. 15 copies of the enzyme searching simultaneously would reduce $\langle\langle n_r \rangle\rangle$ to 3 min, which would be acceptable. It also agrees with an estimated time of 3−5 min for restriction of λ DNA in an *E. coli* host [carrying the type IA restriction and modification system *Eco*KI (Weinfeld & Paigen, 1964; Schell & Glover, 1966)]. It further shows that, if our estimation of 40−50 *Eco*R124II complexes per cell is correct, the host defense of the bacteria against phage can quickly be saturated if enough phages invade the cell at the same time. There are several reports that increasing multiplicity of infection (moi) leads to an increasing number of phages escaping host restriction (Paigen & Weinfeld, 1963; Weinfeld & Paigen, 1964; Uetake et al., 1964); in fact, the probability of successful infection already starts to increase at moi = 2. Furthermore, a first infection of *E. coli* K by phage λ can protect a second, subsequent infection from host restriction (Weinfeld & Paigen, 1964).

We have approximated quite severely in our liberal use of averages, rather than convoluting the distribution of n_r with the interval lengths of the plasmid. This is partly because explicit formulae for the distribution of the number of steps required to reach the end are very cumbersome (Feller, 1968) and partly because the time available for searching, before the infection takes hold, is only approximately known and may depend on cell size, ambient temperature, etc. Furthermore, we know very little about the diffusion of *Eco*R124II molecules in the cytoplasm of *E. coli*. But it is gratifying to see that our estimate of the numbers required, about 15 per plasmid, accords well with the 40 or 50 which appear to be present in the cell as a whole.

ACKNOWLEDGMENT

We thank P. Jenö for protein concentration determinations and E. Kulik for having provided the plasmid pEKU60.

REFERENCES

Adam, G., & Delbrück, M. (1968) in *Structural chemistry and molecular biology* (Rich, A., & Davidson, N., Eds.) Freeman, San Francisco.

Bailone, A., Levine, A., & Devoret, R. (1979) *J. Mol. Biol. 131*, 553−572.

Berg, O. G., Winter, R. B., & von Hippel, P. H. (1982) *TIBS 7*, 52−55.

Bickle, T. A. (1993) in *Nucleases* (Linn, S. M., Lloyd, R. S., & Roberts, R. J., Eds.) pp 89−109, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Bickle, T. A., & Streiff, M. (1983) in *Techniques in the life sciences* (Favell, R. A., Eds.) pp 1−8, Elsevier, Amsterdam.

Brenowitz, M., Senar, D. F., Shea, M. A., & Ackers, G. K. (1986) *Methods Enzymol. 130*, 132−181.

de Feijter, J. A., Benjamins, J., & Veer, F. A. (1978) *Biopolymers 17*, 1759−1772.

Dodson, M., & Echols, H. (1991) *Methods Enzymol. 208*, 168−196.

Feller, W. (1968) *An Introduction to Probability Theory and its Applications, I*, 3rd ed., Wiley, New York.

Firman, K., Creasey, W. A., Watson, G., Price, C., & Glover, S. W. (1983) *Mol. Gen. Genet. 191*, 145−153.

Firman, K., Price, C., & Glover, S. W. (1985) *Plasmid 14*, 224−234.

Fried, M., & Crothers, D. M. (1981) *Nucleic Acids Res. 9*, 6505−6525.

Friedman, D. I., & Gottesman, M. (1983) in *Lambda II* (Hendrix, R. W., Roberts, J. W., Stahl, F. W., & Weisberg, R. A., Eds.) pp 21−51, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Galas, D. J., & Schmitz, A. (1978) *Nucleic Acids Res. 5*, 3157−3170.

Garner, M. M., & Revzin, A. (1981) *Nucleic Acids Res. 9*, 3047−3060.

Gilbert, W., & Müller-Hill, B. (1967) *Proc. Natl. Acad. Sci. U.S.A. 58*, 2415−2421.

Glover, S. W., Firman, K., Watson, G., Price, C., & Donaldson, S. (1983) *Mol. Gen. Genet. 190*, 65−69.

Gorbalenya, A. E., & Koonin, E. V. (1991) *FEBS Lett. 291*, 277−281.

Gubler, M., & Bickle, T. A. (1991) *EMBO J. 10*, 951−957.

Hockensmith, J. W., Kubasek, W. L., Vorachek, W. R., Evertsz, E. M., & von Hippel, P. H. (1991) *Methods Enzymol. 208*, 211−236.

Karrasch, S., Dolder, M., Schabert, M., Ramsden, J. J., & Engel, A. (1993) *Biophys. J. 65*, 2437−2446.

Kulik, E. (1995) Ph.D. Thesis, Biozentrum, Basel, Switzerland.

Larralde, H., Trunfio, P., Havlin, S., Stanley, H. E., & Weiss, G. H. (1992) *Nature 355*, 423−426.

Laurent, T. C. (1995) *Biophys. Chem. 57*, 7−14.

Leirmo, S., Harrison, C., Cayley, D. S., Burgess, R. R., & Record, M. T., Jr. (1987) *Biochemistry 26*, 2095−2101.

Levich, V. G. (1962) *Physicochemical Hydrodynamics*, Prentice Hall, Englewood Cliffs, NJ.

Lindner, R., & Ralston, G. (1995) *Biophys. Chem. 57*, 15−25.

Lohman, T. M., & Bujalowski, W. (1991) *Methods Enzymol. 208*, 258−290.

Lukosz, W. (1993) *Biosens. Bioelectron. 6*, 215−225.

Lvov, Y. M., Decher, G., & Sukhorov, G. (1993) *Macromolecules 26*, 5396−5399.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1989) *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
McCloskey, M. A., & Poo, M. (1986) *J. Cell Biol. 102*, 88−96.
Meynell, E., & Datta, N. (1966) *Genet. Res. 7*, 134−140.
Montroll, E. W. (1969) *J. Math. Phys. 10*, 753.
Paigen, K., & Weinfeld, J. (1963) *Virology 19*, 565−572.
Price, C., Shepherd, C. W., & Bickle, T. A. (1987a) *EMBO J. 6*, 1493−1497.
Price, C., Pripfl, T., & Bickle, T. A. (1987b) *Eur. J. Biochem. 167*, 111−115.
Price, C., Lingner, J., Bickle, T. A., Firman, K., & Glover, S. W. (1989) *J. Mol. Biol. 205*, 115−125.
Rabinowitch, E. (1937) *Trans. Faraday Soc. 33*, 1225−1233.
Ramsden, J. J. (1993) *J. Statist. Phys. 73*, 853−877.
Ramsden, J. J. (1994) *Q. Rev. Biophys. 27*, 41−105.
Ramsden, J. J., Lvov, Y. M., & Decher, G. (1995a) *Thin Solid Films 254*, 246−251.
Ramsden, J. J., Lvov, Y. M., & Decher, G. (1995b) *Thin Solid Films 261*, 343−344.
Richey, B., Cayley, D. S., Mossing, M. C., Kolka, C., Anderson, C. F., Farrar, T. C., & Record, M. T., Jr. (1987) *J. Biol. Chem. 262*, 7157−7164.
Riggs, A. D., Suzuki, H., & Bourgeois, S. (1970) *J. Mol. Biol. 48*, 67−83.
Schell, J., & Glover, S. W. (1966) *J. Gen. Microbiol. 45*, 71−83.
Sober, H. A. (1970) *Handbook of Biochemistry*, 2nd ed., Chemical Rubber Co., Cleveland, OH.
Taylor, I., Patel, J., Firman, K., & Kneale, G. (1992) *Nucleic Acids Res. 20*, 179−186.
Tiefenthaler, K. (1992) *Adv. Biosens. 2*, 261−289.
Tiefenthaler, K., & Lukosz, W. (1989) *J. Opt. Soc. Am. B 6*, 209.
Towbin, H., Staehelin, T., & Gordon, J. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 4350−4353.
Uetake, H., Toyama, S., & Hagiwara, S. (1964) *Virology 22*, 202−213.
Watson, N. (1988) *Gene 70*, 399−403.
Weinfeld, H., & Paigen, K. (1964) *Virology 24*, 71−83.
Winter, R. B., Berg, O. G., & von Hippel, P. H. (1981) *Biochemistry 20*, 6961−6977.
Wood, W. B. (1966) *J. Mol. Biol. 16*, 118−133.
Woodcock, D. M., Crowther, P. J., Doherty, J., Jefferson, S., De Cruz, E., Noyer-Weidner, M., Smith, S. S., Michael, M. Z., & Graham, M. W. (1989) *Nucleic Acids Res. 17*, 3469−3478.

BI952158F