Mayrand, S., & Pederson, T. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 2208-2212.
McCready, S. J., Godwin, J., Mason, D. W., Brazell, I. A., & Cook, P. R. (1980) *J. Cell Sci. 46*, 369-386.
Munro, H. N., & Fleck, A. (1965) *Methods Biochem. Anal. 14*, 113-176.
Nelkin, B. D., Pardoll, D. M., & Vogelstein, B. (1980) *Nucleic Acids Res. 8*, 5623-5632.
Old, R., Murry, K., & Roizes, G. (1975) *J. Mol. Biol. 92*, 331-339.
Pardoll, D. M., & Vogelstein, B. (1980) *Exp. Cell Res. 128*, 466-469.
Pardoll, D. M., Vogelstein, B., & Coffey, D. S. (1980) *Cell (Cambridge, Mass.) 19*, 527-536.
Razin, S. V., Mantieva, V. L., & Georgiev, G. P. (1978) *Nucleic Acids Res. 5*, 4737-4751.
Razin, S. V., Mantieva, V. L., & Georgiev, G. P. (1979) *Nucleic Acids Res. 7*, 1713-1735.
Rigby, P. W., Dieckmann, M., Rhodes, C., & Berg, P. (1977) *J. Mol. Biol. 113*, 237-251.
Scheer, U., Spring, H., & Trendelenburg, M. F. (1979) *Cell Nucl. 6*, 3-17.
Sharper, J. H., Pardoll, D. M., Kaufmann, S. H., Barrack, E. R., Vogelstein, B., & Coffey, D. S. (1979) *Adv. Enzyme Regul. 17*, 213-248.
Sinden, R. R., & Pettijohn, J. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 224-228.
Toole, J. J., Hastie, N. D., & Held, W. A. (1979) *Cell (Cambridge, Mass.) 17*, 441-448.
van Eekelen, C. A. G., & van Venrooiz, W. J. (1981) *J. Cell Biol. 88*, 554-563.
Villa-Komaroff, L., Efstratiadis, A., Broome, S., Lomedico, P., Tizard, R., Naber, S. P., Chick, W. L., & Gilbert, W. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 3727-3731.
Vogelstein, B., Pardoll, D. M., & Coffey, D. S. (1980) *Cell (Cambridge, Mass.) 22*, 79-85.
Wang, D., & Moore, S. (1978) *J. Biol. Chem. 253*, 7216-7219.
Warren, A. C., & Cook, P. R. (1978) *J. Cell Sci. 30*, 211-226.

Downloaded by UNIV MARYLAND COLG PARK on August 31, 2015 | http://pubs.acs.org
Publication Date: November 1, 1981 | doi: 10.1021/bi00527a028

Diffusion-Driven Mechanisms of Protein Translocation on Nucleic Acids. 1. Models and Theory[†]

Otto G. Berg,[‡] Robert B. Winter,[§] and Peter H. von Hippel*

ABSTRACT: Genome regulatory proteins (e.g., repressors or polymerases) that function by binding to specific chromosomal target base pair sequences (e.g., operators or promoters) can appear to arrive at their targets at faster than diffusion-controlled rates. These proteins also exhibit appreciable affinity for nonspecific DNA, and thus this apparently facilitated binding rate must be interpreted in terms of a two-step binding mechanism. The first step involves free diffusion to any nonspecific binding site on the DNA, and the second step comprises a series of protein translocation events that are also driven by thermal fluctuations. Because of nonspecific binding, the search process in the second step is of reduced dimensionality (or volume); this results in an accelerated apparent rate of target location. In this paper we define four types of processes that may be involved in these protein translocation events between DNA sites. These are (i) "macroscopic" dissociation-reassociation processes within the domain of the DNA molecule, (ii) "microscopic" dissociation-reassociation events between closely spaced sites in the DNA molecule, (iii) "intersegment transfer" (via "ring-closure") processes between different segments of the DNA molecule, and (iv) "sliding" along the DNA molecule. We present mathematical and physical descriptions of each of these processes, and the consequences of each for the overall rate of target location are worked out as a function of both the nonspecific binding affinity between protein and DNA and the length of the DNA molecule containing the target sequence. The theory is developed in terms of the *Escherichia coli lac* repressor-operator interaction since data for testing these approaches are available for this system [Barkley, M. (1981) *Biochemistry 20*, 3833; Winter, R. B., & von Hippel, P. H. (1981) *Biochemistry* (second paper of three in this issue); Winter, R. B., Berg, O. G., & von Hippel, P. H. (1981) *Biochemistry* (third paper of three in this issue)]. However, we emphasize that this approach is general for the analysis of mechanisms of biological target location involving facilitated transfer processes via nonspecific binding to the general system of which the target forms a small part.

1. Introduction

It is clear that in discharging many of their physiological functions (e.g., the processes of replication, transcription, translation, recombination, and repair) the proteins or protein complexes involved in various aspects of regulation of genome expression must translocate (move) along DNA or RNA molecules. Such translocation is generally unidirectional, proceeds at fairly well-defined rates, and requires the conversion of chemical to mechanical energy [for a recent summary, see Kornberg (1980)].

Simple protein-nucleic acid binding interactions are of two general types, and each may also involve various protein translocation mechanisms, though these are driven by diffusion processes (i.e., thermal fluctuations) only. These interactions include (i) the binding of regulatory proteins to one or a few specific target sites on the DNA genome (for example, the binding of repressors to specific operator sequences and the

[†] From the Institute of Molecular Biology and Department of Chemistry, University of Oregon, Eugene, Oregon 97403. *Received April 20, 1981.* This work was supported in part by U.S. Public Health Service Research Grant GM-15792 (to P.H.v.H.). O.G.B. also gratefully acknowledges partial support from the Swedish Natural Science Research Council. R.B.W. is pleased to acknowledge support as a predoctoral trainee on U.S. Public Health Service Training Grants GM-00715 and GM-07759.

[‡] Present address: Department of Theoretical Physics, Royal Institute of Technology, S-100 44 Stockholm, Sweden.
[§] Present address: Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, CO 80302.

0006-2960/81/0420-6929$01.25/0 © 1981 American Chemical Society

initial binding of RNA polymerases to closed promoters) and (ii) the (strandedness-specific, but sequence-nonspecific) binding of proteins to nucleic acid lattices at saturating, or close to saturating, levels (for example, the cooperative binding of helix-destabilizing proteins (HDP) to the single-stranded DNA structures involved as intermediates in DNA replication or recombination or of histones to double-stranded DNA after replication as a final step in chromatin formation).

The possible role of diffusion-based translocation mechanisms in establishing equilibrium protein arrangements on DNA lattices in binding processes of type ii is just beginning to be considered [e.g., see Epstein (1979); Kowalczykowski et al., 1980; Lohman, 1980; Lohman & Kowalczykowski, 1981]. In this and the following papers in this issue (Winter & von Hippel, 1981; Winter et al., 1981), we focus on the kinetics and equilibria of interactions of type i.

Proteins that function by binding to specific DNA target sites, such as *Escherichia coli lac* repressor binding to operator, can (in principle) reach their targets by simple three-dimensional diffusion. Two features of these systems suggest that the situation may be more complex. First, the concentration of target sites is often very low, for example, the *lac* operator occurs only once per *E. coli* chromosome, corresponding to an in vivo operator concentration of $\sim 2 \times 10^{-9}$ M. And second, the specific target sequences are buried among many non-target sites ($\sim 10^7$/genome) that share many of the structural features of the target and for which, as a consequence, genome binding proteins often display an appreciable non-base-pair sequence-specific affinity.

These two features conspire to make target location by direct (trial-and-error) three-dimensional diffusion very slow. Binding to the target sequence must be precise; in principle, even binding to an operator one base-pair out of register would result in a totally nonspecific (and physiologically ineffective) interaction. In addition nonspecific binding (and the subsequent multiple series of dissociations from nonspecific sites required to reach the target site) would slow the reaction still further.

In principle, as early appreciated by Adam & Delbrück (1968) and Richter & Eigen (1974), this nonspecific binding affinity can be converted from a kinetic liability to a kinetic asset if a two- (or more) step binding process is invoked. The first step must involve a diffusional encounter with an "extended" target, i.e., with the macromolecule (or organelle) of which the target sequence forms a part. The second (and subsequent) step must comprise some sort of transfer events in which nonspecific binding to the extended target holds the ligand to the target-containing structure and serves to reduce the dimensionality of (and thus speed up) the search process. Clearly this principle applies equally well to a protein searching (basically in one dimension; see below) for a target sequence along (within) a DNA molecule and to a membrane-adhering ligand searching in two dimensions on a membrane surface for a protein receptor site. What is required is some nonspecific affinity of the ligand for the general macromolecular structure within which the target is located plus a mechanism of facilitated transfer of the ligand while bound to this structure.

The problem first came to light experimentally at the quantitative molecular level when Riggs et al. (1970) showed, using filter binding methods at very low component concentrations ($\sim 10^{-12}$ M), that the observed second-order rate constant (k_a) for the binding of *E. coli lac* repressor to a *lac* operator site inserted into λ DNA was $\sim 10^{10}$ M^{-1} s^{-1}. When the Debye–Smoluchowski equation is used, a theoretical estimate of this rate constant for a diffusion-controlled reaction may be made:

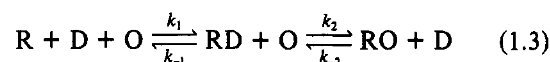
$$k_a = 4\pi\kappa f_{elec} b(D_R + D_O)N_0/1000 \qquad (1.1)$$

where κ is a (unitless) steric interaction factor, f_{elec} is a (unitless) electrostatic (attractive or repulsive) factor, b is the interaction radius (in cm), D_R and D_O are the free-volume diffusion constants for R and O (in cm^2/s) and N_0 is Avagadro's number. (As written, the units of k_a are M^{-1} s^{-1}.) Using reasonable estimates of the above parameters [e.g., see von Hippel (1979); a further discussion of these factors is presented in Winter et al. (1981)], we *calculate* that k_a for a one-step *diffusion-controlled* interaction of repressor (R) with operator (O)

$$R + O \underset{k_d}{\overset{k_a}{\rightleftharpoons}} RO \qquad (1.2)$$

should not exceed $\sim 10^7$ to $\sim 10^8$ M^{-1} s^{-1}; i.e., the calculated maximum value of k_a is 100–1000-fold *smaller* than the experimentally measured rate constant.

Since it is manifestly impossible for a process to be *faster* than diffusion controlled, this can only mean that the reaction is not properly represented by eq 1.2 and that a binding process involving at least two sequential steps must be invoked. This can be written (see section 2)



$$R + D + O \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} RD + O \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} RO + D \qquad (1.3)$$

where D represents any nonspecific (nonoperator) DNA binding site for repressor. The first step of eq 1.3 then represents a three-dimensional diffusion of repressor to *any* site on the DNA molecule, and the second step(s) represents (represent) a diffusion process of reduced dimensionality (and/or volume); the measured or calculated rate constants for the individual steps must then be such as to result in an *overall* transfer of repressor to operator (eq 1.2) with an observed second-order rate constant (k_a) that exceeds the value calculated by eq 1.1 by 2–3 orders of magnitude.

In this paper we address the general problem of such facilitated transfer processes by first defining discrete molecular models for the various ways in which proteins can (in principle) bind to and diffusionally translocate on a nucleic acid lattice. Then we write out the mathematical theory for the way in which each of these molecular processes, substituted for the second step of the two-step reaction scheme 1.3, might be expected to affect the overall observed association and dissociation rate constants for the repressor–operator complex as a function of the major experimental variables accessible in this system. These variables include primarily salt concentration (and salt type), size of the DNA fragment containing the operator (i.e., ratio of nonspecific to specific binding sites), and the number (per DNA fragment) and overall affinity for repressor of the operator binding site(s). In the second paper (Winter & von Hippel, 1981), we report the experimental measurement of equilibrium parameters for repressor binding to operator as a function of the above variables, and in the third paper (Winter et al., 1981), we summarize our kinetic measurements on the system as a function of the above variables, compare the experimental results with the predictions of the theory presented here, and consider in vivo implications.

2. Molecular Models for Translocation Mechanisms

Dissociation–Reassociation. As indicated above, the central problem for the protein is to identify a specific target sequence

Downloaded by UNIV MARYLAND COLG PARK on August 31, 2015 | http://pubs.acs.org
Publication Date: November 1, 1981 | doi: 10.1021/bi00527a028

Downloaded by UNIV MARYLAND COLG PARK on August 31, 2015 | http://pubs.acs.org
Publication Date: November 1, 1981 | doi: 10.1021/bi00527a028

among a vast excess of structurally similar nonspecific binding sites. In a random search, the protein would have to "test", on the average, a large fraction of all the nonspecific sites before the target site is located. This requires a large number of nonspecific dissociation–reassociation events. In the simplest representation, each such dissociation–reassociation event involves a full (*macroscopic*) dissociation of the protein from the nucleic acid, followed by random reassociation to a totally uncorrelated site (a fully random diffusional search). We note, however, that the linear arrangement of nonspecific sites in the DNA makes a *correlated* search possible. Thus we define a *microscopic* dissociation event which releases the protein to a point at which, though free to move, it is still very near the original site and can, with high probability and within a very short time, reassociate with the same or a nearby site. The theory of diffusion-controlled processes requires that the number of such microscopic (release from the chain) dissociations per macroscopic (release *and* transport away from the chain) dissociations be very large (Berg, 1978). This correlated search can be envisioned as a sort of "*hopping*" process and leads to the fact that during the time that a protein remains macroscopically bound it can actually test several nearby sites on the chain through repeated microscopic dissociations.

This definition of a microscopic dissociation event requires a precise distinction between bound and unbound states. The nonspecific binding of *lac* repressor to DNA is entirely electrostatic [i.e., it depends on charge–charge interactions between DNA phosphates and basic residues of the protein; see de-Haseth et al. (1977); Revzin & von Hippel, 1977; Winter & von Hippel, 1981)]. Thus, in terms of the approach of Record and Manning and their co-workers (Record et al., 1976, 1978; Manning, 1978), this association is entirely driven by the release of condensed counterions from the DNA. In these terms, a microscopic dissociation event can be defined as one in which the protein is removed just far enough from the DNA to permit counterion recondensation.

These protein dissociation–reassociation events comprise translocation modes that *must* exist in general and as a necessary consequence of the molecular structure of the system. They are depicted schematically in the upper two branches of Figure 1.

Intersegment Transfer. In addition to these fundamental translocation modes, additional facilitating mechanisms can also be envisioned. In branch 3 of Figure 1, we depict intersegment transfer—or direct transfer between DNA segments as proposed by von Hippel et al. (1975)—which postulates that the protein can be transiently "doubly bound" between two DNA segments of the same chain via a "ring-closure" event. This could occur as a consequence of random spatial fluctuations in the DNA chain that bring a second segment close enough to an already bound protein to establish such a doubly bound complex. When the segments again separate, the protein either stays on its original site or is carried off by the other segment. It can be assumed that the doubly bound complex is unfavorable and that when this complex dissociates the protein will, with equal probability, remain bound to either of the two segments. Such a scheme circumvents dissociation barriers and provides a potentially fast pathway for the sampling of different DNA sites.

We note that this process (branch 3 of Figure 1) comprises a random search which is totally analogous, in its consequences, to the uncorrelated dissociation–reassociation process (branch 1) described above (except that in the intersegment transfer process, because of the stiffness of the DNA, one-step transfer events between two sites that are close together along

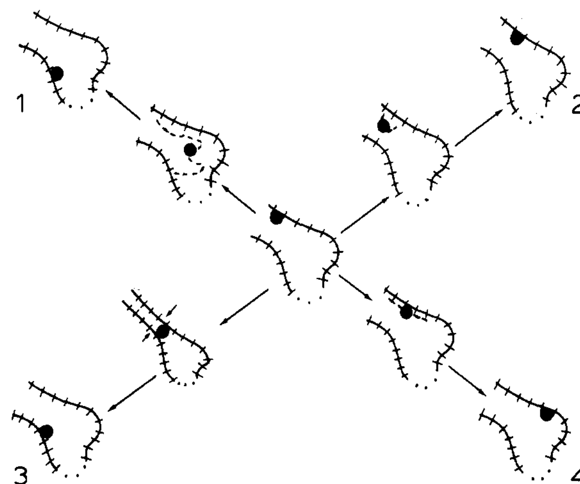


FIGURE 1: Possible processes contributing to the effective transfer rate constant (k_2). Starting from the center picture where the protein is bound to one nonspecific site, it can be transferred to another nonspecific site in the following four ways: Branches 1 and 2 represent the intradomain dissociation–reassociation reactions which are always present; thus branch 1 shows the uncorrelated (macroscopic) transfer reaction and branch 2 the correlated (microscopic) process (hopping). Branch 3 represents the intersegment transfer mechanism, which is driven by the spatial fluctuations of the DNA chain and requires a doubly bound intermediate. Branch 4 models the sliding mechanism, which requires that the protein slides linearly across nonspecific sites without intervening dissociation.

the DNA are discriminated against). Thus this translocation mechanism will be quantitatively effective only if it results in faster protein translocation than does the macroscopic dissociation–reassociation process. This intersegment transfer rate must ultimately be limited by the rate with which two segments approach each other, i.e., by the rate of segmental diffusion of the DNA chain (Berg, 1979).

The second DNA binding site that this model presupposes has recently been experimentally observed by O'Gorman et al. (1980) for *lac* repressor binding to short operator fragments; in the molecular model suggested by these workers, the operator fragments are envisioned as binding in parallel to opposite sides of the repressor. For larger fragments such double binding has not been observed [see Winter & von Hippel (1981)]. This could indicate that the doubly bound complex between longer chain segments (if it exists) is strongly destabilized, perhaps by steric or electrostatic repulsion between the portions of the DNA segments that protrude beyond the repressor. Obviously such a destabilization is required to facilitate the proposed intersegment transfer mechanism since a *stable* doubly bound complex would only serve as an effective trap to further *slow* the search for the specific target site.

Sliding. The other mechanism that has been proposed for facilitated translocation of genome-binding proteins on DNA is "sliding" (Riggs et al., 1970; Richter & Eigen, 1974; Berg & Blomberg, 1976). By this we mean transfer (during a nonspecific binding event) of the protein *along* the contour length of the DNA. In this process (as opposed to macroscopic dissociation and intersegment transfer by ring closure), the protein will sample *strongly correlated* sites since transfer is only between linearly contiguous binding positions on the DNA lattice. In contrast to "hopping", which also proceeds between strongly correlated sites, sliding occurs while the protein remains nonspecifically bound (compare branches 2 and 4 of Figure 1). The basic assumption is that the protein can slide along the DNA in a one-dimensional random walk while bound [for a mechanistic discussion of this process, see Winter et al. (1981)]. This sliding "search pattern" is then interrupted either by location of (and binding to) the specific target site

Downloaded by UNIV MARYLAND COLG PARK on August 31, 2015 | http://pubs.acs.org
Publication Date: November 1, 1981 | doi: 10.1021/bi00527a028

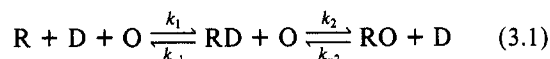
or by dissociation of the protein from the DNA molecule.

Schurr (1979) has calculated a theoretical upper limit for the sliding rate of the *lac* repressor based on purely hydrodynamic considerations. This upper limit turns out to be a one-dimensional diffusion rate of $\sim 4.5 \times 10^{-9}$ cm^2 s^{-1} (corresponding to a random walk rate along the DNA of $\sim 4 \times 10^6$ base pairs/s), which is substantially slower than the free diffusion rate of protein ($\sim 5 \times 10^{-7}$ cm^2 s^{-1}) primarily because the protein is viewed as "spiraling" along the DNA double helix, always "facing" the same way toward the sugar–phosphate backbone. This requires that the protein make a full rotation about the double helix for each 10 base pairs translocated, and the main solvent resistance opposing sliding turns out to be that directed against this rotational motion.

Since the nonspecific affinity of repressor for DNA is electrostatic in origin, resistance to motion could also develop due to local variations (along the DNA) of the electrostatic binding potential. Because the phosphates occur at specific sites on the DNA, rather than being uniformly "smeared" over the cylindrical molecule, there will exist potential barriers to sliding due to the discrete positioning of these groups. However, these "bumps" or barriers opposing sliding can be effectively diminished if the positive charges on the protein are placed somewhat "out-of-register" with the negative DNA phosphates. In principle, sliding could also be inhibited by local fluctuations in the counterion concentrations along the DNA. We note, on the average, that no *net* displacement of counterions is required for sliding, since those displaced from the DNA in "front" of the protein are replaced "behind". Furthermore, relative to the rate of sliding, these counterion rearrangements should be fast. Thus it is not inconceivable that the sliding rate will, in fact, be close to the upper limit calculated from solvent resistance alone. This point is considered further in Winter et al. (1981).

3. Two-Step Reaction Scheme

A protein that has some affinity for nonspecific DNA in addition to its affinity for the specific (target) site will almost certainly bind nonspecifically at first, due simply to the vast excess of nonspecific sites over specific ones. Thus, regardless of the existence of any facilitating transfer mechanism, the specific association process (eq 1.2) must be viewed as a two-step process (eq 1.3) with a nonspecific complex as intermediate. As developed in section 1, if we let R represent the protein component (repressor), D the nonspecific DNA sites, and O the specific (operator) site(s), we obtain the global reaction scheme



$$\mathrm{R} + \mathrm{D} + \mathrm{O} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{RD} + \mathrm{O} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \mathrm{RO} + \mathrm{D} \qquad (3.1)$$

As this reaction is written, the rate constants k_1, k_2, and k_{-2} are treated as bimolecular (although the second step is actually an intramolecular transfer and thus essentially monomolecular), and therefore k_2 and k_{-2} will be concentration dependent.[1] k_2 is the effective rate constant for transfer from nonspecific sites to the specific one(s). No particular facilitating mechanism is implied in eq 3.1, and as written, the transfer process could simply involve macroscopic dissociation–reassociation events (Table I). Since nonspecific sites are in large excess, their free concentration [D] is constant and equal to the total concentration, D_T, of such sites. Hence, at equilibrium, we have

$$\frac{[\mathrm{RD}]}{[\mathrm{R}]} = \frac{k_1 D_T}{k_{-1}} = D_T K_{RD} \qquad (3.2)$$

which defines the nonspecific binding constant K_{RD}. Similarly, the specific binding constant K_{RO} is defined as

$$\frac{[\mathrm{RO}]}{[\mathrm{R}][\mathrm{O}]} = \frac{k_2}{k_{-2}} K_{RD} = K_{RO} \qquad (3.3)$$

[1] All the symbols used recurrently in the main text are summarized alphabetically in Table I to facilitate cross reference and avoid confusion. The table also includes the number of the equation in which these symbols are first introduced or defined.

Table I: Glossary of Symbols Used Recurrently in the Main Text

symbol	definition	eq [a]
a	persistence length of DNA	4.5
b	reaction radius for nonspecific binding	4.2
c	free repressor concentration	4.11
D	free diffusion constant for the protein	4.2
D_1	diffusion constant for the one-dimensional sliding	5.17
D_c	local DNA concentration (base pairs) inside a domain	4.10
D_T	total DNA concentration (base pairs) in solution	3.2
k_1	nonspecific association rate constant	3.1, 4.19
k_{-1}	nonspecific dissociation rate constant	3.1, 4.18
k_2	effective forward transfer rate constant	3.1, B.10
k_{-2}	effective backward transfer rate constant	3.1
k_a	total association rate constant to specific site	3.5
k_a'	total association rate constant to secondary site	6.3
k_d	total dissociation rate constant from specific site	3.6
k_d'	total dissociation rate constant from secondary site	6.3
k_i	microscopic nonspecific association rate constant	4.1
k_{assoc}	intradomain nonspecific association rate constant	4.6
K_{RD}	nonspecific binding constant	3.2
K_{RO}	specific binding constant	3.3
l	length of a base pair	4.2
L	half of the contour length of the DNA chain	4.4
L_0	contour length between specific sites	6.19
M	total number of sites per DNA chain	3.5
N	number of specific sites per DNA chain	6.1
n_0	number of base pairs between specific sites	6.19
O_T	total concentration of DNA chains	3.5
P	probability that a DNA chain has one or more proteins specifically bound	6.5
P_r	intradomain reassociation probability	4.9, 4.17
r_g	radius of gyration of DNA chain	4.5
R_c	average intersegment distance in the DNA domain	4.4
R_T	total protein concentration	3.2
γ	ratio of nonspecific and specific binding constants	3.7
η	geometric factor for domain association	4.15
θ	fractional saturation of primary binding site	6.4
θ^*	fractional saturation of secondary binding site	6.4
λ	microscopic nonspecific dissociation rate constant	4.1
Λ	intradomain nonspecific dissociation rate constant	4.3
μ	ratio of total protein and DNA chain concentrations	6.4
ν	intersegment transfer rate constant	5.9
ψ	microscopic reassociation probability	4.2

[a] Number of the equation in which the symbol is first introduced or defined.

However, what is generally measured [see Winter et al. (1981)] is the concentration of repressor molecules bound to operator sites relative to the total concentration of those *not* bound there. That is

$$\frac{[\mathrm{RO}]}{[\mathrm{RD} + \mathrm{R}][\mathrm{O}]} = \frac{k_2}{k_{-2}}\left[\frac{K_{\mathrm{RD}}}{1 + D_{\mathrm{T}}K_{\mathrm{RD}}}\right] = K_{\mathrm{RO}}^{\mathrm{obsd}} \quad (3.4)$$

The effective specific association rate can be calculated from eq 3.1 by means of standard steady-state analysis or by calculating the dominant relaxation time in a kinetic analysis. If it is assumed that the specific complex is stable (i.e., is formed irreversibly on the time scale of the association, with $k_{-2} = 0$), both approaches show that

$$k_{\mathrm{a}} = \frac{k_2 D_{\mathrm{T}} K_{\mathrm{RD}}}{1 + D_{\mathrm{T}}K_{\mathrm{RD}} + k_2 O_{\mathrm{T}}/k_{-1}} = \left[(Mk_1)^{-1} + \left(\frac{k_2 D_{\mathrm{T}} K_{\mathrm{RD}}}{1 + D_{\mathrm{T}}K_{\mathrm{RD}}}\right)^{-1}\right]^{-1} \quad (3.5)$$

Here O_{T} is the total concentration of specific sites and M ($= D_{\mathrm{T}}/O_{\mathrm{T}}$) is the number of nonspecific sites per (one operator containing) DNA chain, such that Mk_1 is the association rate per chain. The second term in the square brackets is simply the preequilibrium result, i.e., the association rate one would obtain if the first step of eq 3.1 is always at equilibrium.

The whole reaction scheme (eq 3.1) breaks down when k_{-1} becomes much smaller than k_2O_{T}. In this case the repressors are confined to the first DNA chain they encounter. That is, on the time scale of the total association reaction, there is no exchange of proteins between chains, as tacitly assumed in the original reaction scheme. Instead it becomes appropriate to consider each DNA chain as a closed system. Then both backward rates (corresponding to k_{-1} and k_{-2} of the two-step scheme) are negligible, and the full time course of the total association reaction is readily calculated.

The dissociation process can be analyzed in a similar manner. The case where $k_1 = 0$ in reaction scheme 3.1 is relevant to the experimental situation. This situation implies that when a protein has dissociated from a DNA chain or, more exactly, has departed from the "domain" of the chain (see below), it will be lost in solution, either by dilution or by adsorption onto the large excess of "cold" operator-containing DNA that can be added to prevent reassociation [see Winter et al. (1981)]. The equation for the observed specific dissociation rate constant is then

$$k_{\mathrm{d}} = \frac{k_{-2}D_{\mathrm{T}}}{1 + k_2O_{\mathrm{T}}/k_{-1} + k_{-2}D_{\mathrm{T}}/k_{-1}} = \left[(k_{-2}D_{\mathrm{T}})^{-1} + \left(\frac{k_{-1}D_{\mathrm{T}}K_{\mathrm{RD}}}{O_{\mathrm{T}}K_{\mathrm{RO}} + D_{\mathrm{T}}K_{\mathrm{RD}}}\right)^{-1}\right]^{-1} \quad (3.6)$$

which has the same simple form as eq 3.5.

The specific binding constant can be related to the nonspecific one via eq 3.3. We introduce a molecular parameter γ which is always less than unity and which represents the stability of the nonspecific complex relative to that of the specific one; i.e.

$$\gamma \equiv K_{\mathrm{RD}}/K_{\mathrm{RO}} = k_{-2}/k_2 \quad (3.7)$$

Then the total dissociation rate (from eq 3.6) is

$$k_{\mathrm{d}} = \gamma M\left[(k_2O_{\mathrm{T}})^{-1} + \left(\frac{k_{-1}}{1 + \gamma M}\right)^{-1}\right]^{-1} \quad (3.8)$$

When $\gamma M \ll 1$ (i.e., when $MK_{\mathrm{RD}} \ll K_{\mathrm{RO}}$ and specific binding is dominant), the dissociation rate becomes

$$k_{\mathrm{d}} = K_{\mathrm{RO}}^{-1}[(Mk_1)^{-1} + (k_2D_{\mathrm{T}}K_{\mathrm{RD}})^{-1}]^{-1} \quad (3.9)$$

This situation applies in all the experiments described in the following papers in this issue (Winter & von Hippel, 1981; Winter et al., 1981).

4. Kinetics of Nonspecific Binding Events within and between DNA Domains

Microscopic Dissociations and Reassociations. Let us first consider an equilibrium situation with nonspecific DNA (of total concentration D_{T}) and protein molecules (of total concentration R_{T}) in solution. For simplicity we assume (as is generally experimentally true) that nonspecific binding sites are in great excess over protein, $D_{\mathrm{T}} \gg R_{\mathrm{T}}$. As pointed out by von Hippel et al. (1975), in the usual experimental situation the DNA chains exist in solution well separated into small "domains" containing one chain each, with most of the intervening solution "empty" of DNA. However, *at equilibrium*, the concentration of *free* protein [R] must be homogeneous throughout the solution, including the "insides" of the DNA domains. Consequently at equilibrium it is totally immaterial that the DNA distribution is inhomogeneous, and we can define a *microscopic* association rate constant (k_{i}) as well as a *microscopic* dissociation rate constant (λ) such that $k_{\mathrm{i}}[\mathrm{R}]D_{\mathrm{T}}$ equals the association flux at equilibrium and $\lambda[\mathrm{RD}]$ is equal to the dissociation flux at equilibrium [cf. Berg (1978) for a more thorough discussion of these microscopic rates). Detailed balance requires these fluxes to be equal, and the nonspecific binding constant (eq 3.2) can also be expressed as

$$K_{\mathrm{RD}} = k_{\mathrm{i}}/\lambda \quad (4.1)$$

Since rates are commonly measured by disturbing the equilibrium, these microscopic rates will not be observed unless they are so small that the diffusion effects are infinitely faster, in which case the distribution of protein remains homogeneous throughout the experiment. This is the so-called reaction-controlled case, which is probably not relevant for real protein–DNA association reactions, but is included as a possible limit on the results to follow.

As defined above, the microscopic dissociation rate constant λ describes a dissociation that merely *releases* the protein from the chain but leaves it directly adjacent to its former binding site. This definition of λ requires a precise distinction between the bound and unbound states. One possibility, utilized in section 2 to describe "hopping", is to let λ define a dissociation event that has only proceeded far enough to allow the counterions to recondense on the DNA chain. In this way the microscopic *association* rate constant (k_{i}) will contain essentially all of the activation free energy for the removal of the counterions on binding of the protein ligand. However, as long as nonspecific association is diffusion limited, the observed rates will be independent of the precise definition of the microscopic ones.

Most microscopic dissociations will be very short-lived and will be followed by almost immediate reassociation of the protein to the same binding site. Thus no observable change will have taken place. However, some of these dissociations will be sufficiently long-lived to allow some free diffusion and reassociation to a neighboring binding site on the chain. The nature and effects of such strongly correlated reassociation processes will be considered further in section 5.

In a small fraction of the microscopic dissociation events, diffusion will transport the protein to a point where it loses its correlation with the binding site (and chain segment) it has just left. However, the protein remains within the domain of

Downloaded by UNIV MARYLAND COLG PARK on August 31, 2015 | http://pubs.acs.org
Publication Date: November 1, 1981 | doi: 10.1021/bi00527a028

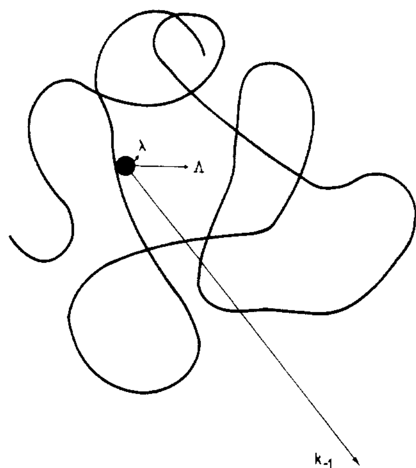


FIGURE 2: Schematic representation of the heirarchy of nonspecific dissociation rates. λ represents a *microscopic* dissociation event that releases the protein from the DNA chain to a distance just beyond the counterion atmosphere. Λ represents an *intradomain* dissociation event that carries the protein away from the original segment but leaves it within the domain. k_{-1} represents an *interdomain* dissociation event resulting in the loss of the protein from the domain altogether.

Downloaded by UNIV MARYLAND COLG PARK on August 31, 2015 | http://pubs.acs.org
Publication Date: November 1, 1981 | doi: 10.1021/bi00527a028

the same DNA chain, and consequently this is called an *intradomain* dissociation. Finally some fraction of these intradomain dissociations will take the protein out of the domain altogether. These are the *interdomain* dissociations given by the macroscopic rate k_{-1} in the general two-step scheme of section 3. Thus, on purely geometric grounds, one can distinguish three levels of nonspecific dissociations—microscopic, intradomain, and interdomain—as depicted schematically in Figure 2. Each of these will enter at an appropriate level in the total association scheme. Below we proceed by defining each set of corresponding rate constants.

Intradomain Dissociations and Reassociations. The occurrence of dissociation events that result in transfer of the protein between different segments of the same chain is controlled by the segment density within the DNA random coil and the microscopic dissociation rate constant, λ. In principle, such intradomain dissociation events will take the protein to a point at which it loses its spatial correlation with the binding site it has just left and is now afforded an equal probability of reassociating at *any* binding site on the same chain. Thus, it is appropriate to count as dissociated those proteins that reach an approximate midpoint between nearby, but uncorrelated, segments. [As used here, the term uncorrelated means that the segments are far apart as measured along the DNA contour but may be (transiently) spatially close.] Using cylindrical coordinates, one can calculate the probability ψ that the protein, once it has dissociated, does *not* reach the distance R_c without reassociation (Berg & Blomberg, 1977):

$$\psi = \frac{\ln (R_c/b)}{\ln (R_c/b) + 2\pi Dl/k_i} \quad (4.2)$$

where b is the reaction radius (here taken as the radius of the DNA double helix), D is the free diffusion constant for the protein, and l is the distance between binding sites (here taken as the length of a base pair). The dissociation rate Λ to distance R_c is the product of the microscopic rate λ and the probability of reaching this distance:

$$\Lambda = \lambda(1 - \psi) = \frac{\lambda(2\pi Dl/k_i)}{\ln (R_c/b) + 2\pi Dl/k_i} \quad (4.3)$$

R_c, which is a measure of the segment density in the DNA coil, can be defined by

$$2\pi L R_c^2 = \frac{4}{3}\pi r_g^3 \quad (4.4)$$

where $2L$ is the chain length and r_g is the radius of the DNA domain, taken here to be the radius of gyration of the DNA coil. For a long DNA chain consisting of many persistence lengths, r_g is determined by

$$r_g^2 = 2La/3 \quad (4.5)$$

where a is the persistence length. Thus

$$R_c = \frac{2}{3}(r_g a)^{1/2} \simeq 0.60(La^3)^{1/4}$$

In this way, R_c is defined as the average distance to the midpoint between one DNA segment and the closest uncorrelated neighboring segment. Consequently Λ as defined by eq 4.3 is the intradomain dissociation rate constant.

This result is consistent with the association rate constant per binding site, k_{assoc} for a *homogeneous* segment distribution of the same density (Berg & Blomberg, 1976, 1977):

$$k_{assoc} = \frac{2\pi Dl}{\ln (R_c/b) + 2\pi Dl/k_i} \ (\mathrm{cm^3\ s^{-1}}) \quad (4.6)$$

Thus the equilibrium constant from (4.1) can also be expressed as $K_{RD} = k_{assoc}/\Lambda$. If $k_i \gg 2\pi Dl$, k_{assoc} is dependent only on geometry and the free diffusion constant D. This is the so-called *diffusion-controlled* limit. In the *reaction-controlled* limit, $k_i \ll 2\pi Dl$, and k_{assoc} from (4.6) becomes equal to the microscopic rate k_i. Note that, for simplicity of notation, all bimolecular rate constants are given in units of cm^3 s^{-1}, so that the ratio $2\pi Dl/k_i$ is dimensionless.

The need to consider a dissociation distance at all arises from the essentially two-dimensional character of the diffusion described in cylindrical coordinates when the coordinate along the chain is immaterial. In three dimensions and spherical symmetry, the probability that a dissociating particle reaches a distance R without reassociation rapidly approaches a limiting value with increasing R, such that

$$(1 - \psi)_{3D} = \left[\frac{k}{4\pi bD}\left(1 - \frac{b}{R}\right) + 1\right]^{-1} \xrightarrow[R \gg b]{} \frac{4\pi bD}{4\pi bD + k} \quad (4.7)$$

where b is the reaction radius and k is a microscopic reaction rate. Hence, the common procedure of using an infinite dissociation distance in three dimensions is justified. As is obvious from eq 4.2, there is no such limiting value for the escape probability in two dimensions:

$$(1 - \psi)_{2D} = [[(k_i/2\pi Dl) \ln (R/b) + 1]^{-1}] \xrightarrow[R \to \infty]{} 0 \quad (4.8)$$

and it is essential to define a dissociation event properly.

Interdomain Dissociations and Reassociations. At large distances from the chain, the entire DNA coil serves as an essentially spherical target. Escape from this entire DNA domain can be defined as a third level of dissociation (see Figure 2). Obviously the corresponding rate constant, k_{-1}, can be expressed as the product of Λ and the escape probability from the domain:

$$k_{-1} = \Lambda(1 - P_r) \ (\mathrm{s^{-1}}) \quad (4.9)$$

where P_r is the probability that the protein will reassociate to the chain rather than leave the domain altogether. We can express the equivalent reassociation rate for a protein starting free somewhere inside the domain and binding onto the chain inside the domain as

$$k_{\text{assoc}}D_c = \frac{2D/R_c^2}{\ln (R_c/b) + 2\pi Dl/k_i} \ (\text{s}^{-1}) \qquad (4.10)$$

where $D_c = 1/\pi R_c^2 l$ is the local concentration of base pairs or nonspecific binding sites within the domain. For simplicity, the segment density is assumed homogeneous. This results in the diffusion equations

$$\frac{\partial c}{\partial t} = D\nabla^2 c - k_{\text{assoc}}D_c c \qquad 0 \le r < r_g$$
$$\frac{\partial c}{\partial t} = D\nabla^2 c \qquad r > r_g \qquad (4.11)$$

where $c(r,t)$ is the free repressor concentration at the distance r from the center of the domain.

The initial condition is defined as one repressor molecule placed somewhere within the domain. Assuming a homogeneous probability distribution, we have

$$c(r,0) = c_0 = (4\pi r_g^3/3)^{-1} \qquad 0 \le r < r_g$$
$$c(r,0) = 0 \qquad r > r_g \qquad (4.12)$$

To calculate the reassociation probability, it is sufficient to consider the time-integrated form of the diffusion equations:

$$-c_0 = D\nabla^2 \tilde{c} - k_{\text{assoc}}D_c\tilde{c} \qquad 0 \le r \le r_g$$
$$0 = D\nabla^2 \tilde{c} \qquad r > r_g \qquad (4.13)$$

where $\tilde{c}(r) \equiv \int_0^\infty c(r,t)\,dt$.

Note that this is equivalent to considering a stationary state or a Laplace transform in the limit where the Laplace variable is zero. With the appropriate continuity conditions at $r = r_g$, the solution is

$$\tilde{c}(r) = \frac{1}{Mk_{\text{assoc}}}\left[1 - \left(\frac{1}{\cosh (\eta r_g)}\right)\left(\frac{\sinh (\eta r)}{\eta r}\right)\right]$$
$$0 \le r < r_g \qquad (4.14)$$

where $M = D_c/c_0$ is the number of nonspecific sites per chain and η is defined by

$$(\eta r_g)^2 = \frac{3L/r_g}{\ln (R_c/b) + 2\pi Dl/k_i} \qquad (4.15)$$

The total reassociation probability is

$$P_r = k_{\text{assoc}}D_c\int_0^{r_g}\tilde{c}(r)4\pi r^2\,dr \qquad (4.16)$$

which gives the escape probability

$$1 - P_r = \frac{4\pi Dr_g}{Mk_{\text{assoc}}}\left[1 - \frac{\tanh (\eta r_g)}{\eta r_g}\right] = \frac{3}{(\eta r_g)^2}\left[1 - \frac{\tanh (\eta r_g)}{\eta r_g}\right] \qquad (4.17)$$

Thus from eq 4.9 the dissociation rate constant is

$$k_{-1} = \frac{\Lambda}{Mk_{\text{assoc}}}4\pi Dr_g\left[1 - \frac{\tanh (\eta r_g)}{\eta r_g}\right] = K_{\text{RD}}^{-1}\frac{4\pi Dr_g}{M}\left[1 - \frac{\tanh (\eta r_g)}{\eta r_g}\right] \qquad (4.18)$$

where the nonspecific binding constant K_{RD} has been introduced from eq 4.1, 4.3, and 4.6. The bimolecular association rate constant onto any binding site in the domain is given by (Berg & Blomberg, 1977)

$$k_1 = \frac{4\pi Dr_g}{M}\left[1 - \frac{\tanh (\eta r_g)}{\eta r_g}\right] \ (\text{cm}^3\ \text{s}^{-1}) \qquad (4.19)$$

The expression in square brackets is simply the probability that a protein that has reached the domain will also bind nonspecifically to the chain. Consequently, the nonspecific binding constant as originally defined in eq 3.2 is consistently given by the ratio of association and dissociation rate constants at all three levels:

$$K_{\text{RD}} = k_i/\lambda = k_{\text{assoc}}/\Lambda = k_1/k_{-1}$$

5. Predicted Association Rates

General Considerations. The effective transfer rate constant (k_2) remains to be determined before the expected association rate constant, k_a (from eq 3.5), can be calculated. In Figure 1 we have sketched the different transfer processes that may contribute to the overall rate. In the following sections, the effects of these transfer processes on the overall rate are considered individually; in Appendix A, we derive the full solution which applies when all these processes contribute simultaneously.

We assume throughout that the nonspecific binding constant, K_{RD}, is a known quantity. For most DNA binding proteins, K_{RD} has a strong electrostatic component; i.e., the overall binding affinity is very salt concentration dependent. For *lac* repressor, nonspecific binding appears to be almost exclusively electrostatic (deHaseth et al., 1977; Revzin & von Hippel, 1977) in that log K_{RD} extrapolated to 1 M salt is negative. In eq 4.1, K_{RD} ($\equiv k_i/\lambda$) is expressed as a ratio of microscopic rate constants. In the diffusion-controlled limit ($k_i \gg 2\pi Dl$), only this *ratio* appears in the final equations. Thus, the actual choice of k_i and λ is immaterial, and the experimental values of K_{RD} can be used together with the known geometries and the free diffusion constant (D) to predict the specific association rate. In the numerical predictions below (Figures 3–5), we shall assume that the diffusion-controlled limit is applicable; otherwise k_i and λ would have to be known separately. However, unless stated otherwise, this assumption has not been used in the theoretical expressions which follow.

From eq 3.5 we have

$$k_a = \left[(Mk_1)^{-1} + \left(\frac{k_2D_TK_{\text{Rd}}}{1 + D_TK_{\text{RD}}}\right)^{-1}\right]^{-1} \qquad (5.1)$$

The first term (Mk_1) given by eq 4.19 is the rate of the first nonspecific association. It is essentially constant, is determined only by the size of the DNA chain, and will serve as an upper limit to k_a. Keeping this in mind, it suffices to consider the preequilibrium result:

$$k_a = \frac{k_2D_TK_{\text{RD}}}{1 + D_TK_{\text{RD}}} \simeq \begin{cases} k_2D_TK_{\text{RD}} & D_TK_{\text{RD}} \ll 1 \quad (5.2a) \\ k_2 & D_TK_{\text{RD}} \gg 1 \quad (5.2b) \end{cases}$$

The upper limit in eq 5.2 is valid at high salt (weak nonspecific binding) and the lower limit at low salt (strong nonspecific binding). Only when this estimate approaches or exceeds Mk_1 is it necessary to invoke the upper limit for k_a as given by eq 4.19:

$$k_a \simeq Mk_1 = 4\pi Dr_g\left[1 - \frac{\tanh (\eta r_g)}{\eta r_g}\right] \qquad (5.3)$$

Association without Facilitating Mechanisms. In the absence of any facilitating mechanism, the transfer rate between

Downloaded by UNIV MARYLAND COLG PARK on August 31, 2015 | http://pubs.acs.org
Publication Date: November 1, 1981 | doi: 10.1021/bi00527a028

Downloaded by UNIV MARYLAND COLG PARK on August 31, 2015 | http://pubs.acs.org
Publication Date: November 1, 1981 | doi: 10.1021/bi00527a028

nonspecific sites, k_2D_T, can be interpreted simply as the rate of exchange of bound repressor between different segments of the DNA by macroscopic dissociation–reassociation events. Initially this neglects the possibility of intersite transfer by microscopic dissociation–reassociation processes, an effect which is always present and is of particular importance when there are no other facilitating mechanisms. In this approximation, the specific association rate can easily be calculated without recourse to the details of the two-step scheme. The mean time τ_1 for the *first* nonspecific association is τ_1 [= $1/(k_1D_T)$]. Dissociation takes place after a mean time, τ_{diss} (= $1/\Lambda$). The mean time, τ_{assoc}, which must pass before the next nonspecific association (onto the same chain or another) for a protein starting within the domain, can be calculated by using the methods of section 4. We find $\tau_{assoc} = 1/(k_{assoc}D_T)$.

In this random search, the protein will test nonspecific sites M times, on the average, before the specific target site is found.[2] Thus, the total mean time before specific association is

$$\tau_a = \tau_1 + M(\tau_{diss} + \tau_{assoc})$$

and the total specific association rate is

$$k_a = \frac{1}{\tau_a O_T} = \frac{\Lambda K_{RD}}{1 + D_T K_{RD} + \Lambda/(Mk_{-1})} \simeq \frac{\Lambda K_{RD}}{1 + D_T K_{RD}} \xrightarrow[D_T K_{RD} \ll 1]{} k_{assoc} = \frac{2\pi Dl}{\ln (R_c/b) + 2\pi Dl/k_i} \xrightarrow[2\pi Dl/k_i \ll 1]{} \frac{2\pi Dl}{\ln (R_c/b)} \text{ (cm}^3\text{ s}^{-1}) \quad (5.4)$$

The successive approximations are (i) that the term $\Lambda/(Mk_{-1}) \ll 1$ is always negligible, (ii) that the case $D_TK_{RD} \ll 1$ applies when the nonspecific binding is not competitive enough to slow down the association, with the result that k_a is simply the nonspecific association rate constant k_{assoc} from eq 4.6, and (iii) that in the limit $2\pi Dl/k_i \ll 1$, k_a represents, finally, the diffusion-controlled case. This corresponds to the "screening-controlled" case of Lohman et al. (1978), and the correspondence becomes more obvious when it is noted that a more complete derivation of Λ (Berg & Blomberg, 1978) replaces the logarithmic factor:

$$\ln (R_c/b) \rightarrow \int_b^{R_c} \rho^{-1} \exp[V(\rho)/(k_BT)] \, d\rho \quad (5.5)$$

and the microscopic reaction rate constant:

$$k_i \rightarrow k_i \exp[-V(\rho)/(k_BT)] \quad (5.6)$$

where $V(\rho)$ is the electrostatic interaction potential at distance ρ from the DNA chain axis. The integral in eq 5.5 may well

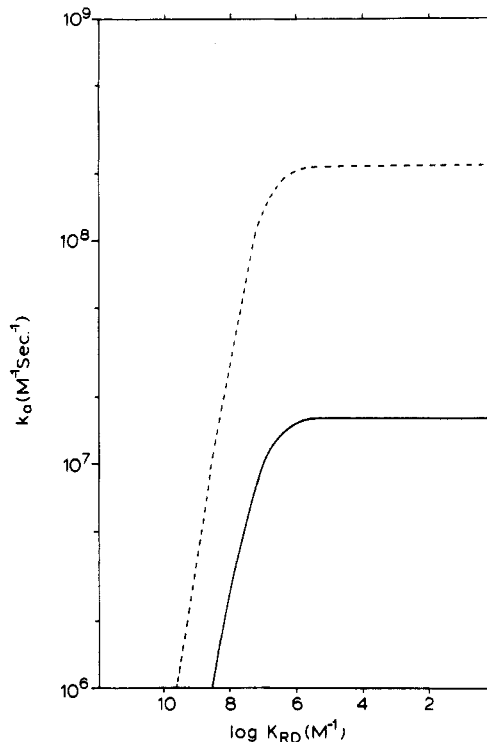


FIGURE 3: Total specific association rate without facilitating mechanisms as a function of the nonspecific binding constant. The solid line is calculated by using $k_a = k_{assoc}/(1 + D_TK_{RD})$ (from eq 5.4; hopping not included). The dashed line is calculated by using $k_a = 1.5\pi bD/(1 + D_TK_{RD})$ (from eq 5.8; hopping included). The breakpoint of the curves occurs at $K_{RD} = D_T^{-1}$, and a value of $D_T = 5 \times 10^{-8}$ M has been used throughout. The diffusion-controlled limit ($k_i \gg 2\pi Dl$) is assumed.

carry the low salt dependence that Lohman et al. (1978) predict from thermodynamic considerations. Obviously, one needs a much more detailed knowledge of the potential to corroborate this. The full expression for k_a, eq 5.4, allows corrections. At low salt concentrations, the competition effect can contribute (i.e., $D_TK_{RD} > 1$), and at high salt, the *nonspecific* association may become reaction controlled, if $k_i < 2\pi Dl$.

Equation 5.4, which also has been derived more rigorously in Appendix A, is identical with eq 3.5 when the effective transfer rate between nonspecific sites is taken as

$$k_2D_T = \Lambda \quad (5.7)$$

Some of the very short-lived microscopic dissociations may return the protein to a neighboring site instead of to the one it left. In contrast to "sliding" (Figure 1), where the protein remains bound, this process is envisioned as a "hopping" mechanism. Thus the protein dissociates from the chain but remains close and reassociates in a very short time. However, during this short time, it acquires an increased mobility and has a certain probability of reassociating with a neighboring site.

This type of (microscopic) dissociation–reassociation event takes the protein between strongly correlated sites. It has been separated from the uncorrelated dissociation–reassociation discussed above primarily for emphasis and mathematical convenience. It needs to be considered only because of the linear arrangement of nonspecific binding sites. In the more common binding situation with independent sites, each located on a different molecule, microscopic dissociation events will not contribute and are not normally discussed.

In Appendix A, this hopping has been included as a part of the general kinetic description. As long as the nonspecific

[2] The *average* (mean) number of nonspecific association–dissociation events which take place before the target site is located in a random search is simply equal to M, the total number of nonspecific sites per chain (actually $M - 1$). Since the probability of hitting the target site is $1/M$ for each binding event, the probability (P_i) of hitting it on the ith try (and *not* on any of the previous $i - 1$ tries) is $P_i = (1/M)[1 - (1/M)]^{i-1}$. Thus the mean number of nonspecific binding events prior to target location is

$$\sum_{i=1}^{\infty}(i-1)\frac{1}{M}\left(1-\frac{1}{M}\right)^{i-1} = M - 1$$

association is diffusion controlled, the number of different sites tested by hopping during the time that the protein remains "macroscopically associated" will be determined by geometric factors. This number can simply be multiplied by the result of eq 5.4 to obtain

$$k_a \simeq 1.5\pi bD/(1 + D_T K_{RD}) \quad (5.8)$$

from eq A27 of Appendix A. A simliar result has been derived in a standard steady-state analysis (O. G. Berg and M. Ehrenberg, unpublished results), thus confirming that the hopping is indeed a fundamental part of the diffusion process.

In Figure 3 we have plotted the estimated specific association rate as a function of the strength of the nonspecific binding, both with and without the inclusion of the microscopic hopping process. Hopping does not change the shape of the curve; it is simply elevated on the graph. The absolute levels should not be taken as more than order-of-magnitude estimates. Also, the plateaus may not be completely level due to counterion screening effects as defined by eq 5.5.

It should be noted that hopping is a purely geometric effect that follows straightforwardly from the diffusion equations. Consequently, it is not a facilitating mechanism as such since it is always present. However, the results above have been derived under some very idealized assumptions. First, the protein has been taken to be completely free, even in the immediate neighborhood of the chain where both hydrodynamic and electrostatic interactions should influence its motion. In addition, no steric factors have been included. Some of these neglected factors may cancel one another. It should also be remembered that over such short distances as the length of one base pair the motion of a DNA segment may actually be more rapid than is the free diffusion of the protein. To model this "hopping" process accurately, we would have to know the electrostatic potential, the charge distribution on the protein, the hydrodynamic interaction between the protein and the DNA, etc. At the present, such complete modeling is not feasible, and it suffices to point out here that the idealized "geometric" results above provide a first estimate of the importance of the microscopic dissociations. Equation 5.8 is a very reasonable representation of such a purely geometric effect since it gives an ordinary diffusion-controlled result. In effect, the target is extended from something smaller than a base pair $[\sim l/[2 \ln (R_c/b)]]$ to a size comparable to the radius of the chain. Thus this mechanism serves primarily to make the steric constraints on the initial binding event less demanding.

Intersegment (*Ring Closure*) *Transfer*. We now consider the facilitating effects of the proposed direct intersegment transfer process on the overall rate of protein transfer to the specific site.

Let us assume that this proposed transfer takes place, on the average, ν times per second. Since it moves the protein to an uncorrelated segment, this transfer has exactly the same physical consequences as transfer via the macroscopic dissociation–reassociation process discussed above. Thus, again neglecting hopping, we can identify the total transfer rate between nonspecific sites, by analogy with eq 5.7, as

$$k_2 D_T = \Lambda + \nu \ (\mathrm{s}^{-1}) \quad (5.9)$$

(A more rigorous derivation in Appendix A gives the same result.) This approach clearly assumes that there is no "memory" in the transfer, i.e., that the protein quickly loses its correlation with the segment from which it is transferred. When the diffusion distances involved are considered, this is a very reasonable assumption, although it does neglect the negative correlation that transfer cannot take place directly *between* two sites which are closer to one another along the chain than about one persistence length. Also this assumption will not be entirely valid for very densely packed or partially ordered DNA (e.g., the native bacterial chromosome) in which higher order packing arrangements may be sufficiently stable to prevent major rearrangements of the entire DNA chain between transfer events [see also discussion in Winter et al. (1981)].

While it appears difficult to derive an expression for ν in terms of molecular parameters, an upper limit can be estimated by considering the segmental diffusion of the DNA chain since the transfer rate must ultimately be limited by the rate at which two segments approach one another. Berg (1979) has calculated the correlation function for the mean-square displacement of a DNA segment in a wormlike chain model. Since all segments are connected, the result is not represented by a simple diffusion process linear in time, as seen in eq 5.12 and 5.14 below; instead this result exhibits the more curious time dependence:

$$\langle \Delta r^2 \rangle_t \simeq (2a)^{1/2}(D_s t)^{3/4} \quad (5.10)$$

where a is the persistence length and D_s is the expected diffusion constant for a *free* DNA segment of length a. Equation 5.10 is actually the first term in a series expansion for short times, but it remains valid for sufficiently large times to be useful for our purpose here.

From this segmental diffusion rate we can estimate a collision time. For a three-dimensional problem with spherical symmetry, the "single passage" time, τ, for the initial collision of free particles is given by

$$6D\tau = R^2 \ln (R/b) \quad (5.11)$$

where D is the diffusion constant, R is an average distance between particles such that the particle concentration is $(4\pi R^3/3)^{-1}$, and b is the reaction radius. D is also related to the mean-square displacement after time t of one particle:

$$\langle \Delta r^2 \rangle_t = 6Dt \quad (5.12)$$

Similarly, for a two-dimensional problem, one has

$$4D\tau = R^2(R/b) \quad (5.13)$$

and

$$\langle \Delta r^2 \rangle_t = 4Dt \quad (5.14)$$

Equation 5.13 is actually identical with eq 4.10 in the diffusion-controlled limit. We note that the collision between one point (where the protein is bound) on a DNA segment and *any* point on an unrelated segment is essentially a problem in cylindrical symmetry; i.e., the problem involves two-dimensional geometry. Consequently, we can estimate the single passage time, τ, for a collision by

$$R^2 \ln (R/B) = \langle \Delta r^2 \rangle_\tau = (2a)^{1/2}(D_s \tau)^{3/4} \quad (5.15)$$

in which the mean-square displacement term from eq 5.10 has been substituted. For the mean distance R we use R_c of eq 4.4 and 4.5, and for the reaction radius here we use $b = 5 \times 10^{-7}$ cm (estimated to be the protein radius). $R_c = 1.2 \times 10^{-5}$ cm for the chain length $2L = 1.7 \times 10^{-3}$ cm (λ DNA), and the persistence length $a = 6 \times 10^{-6}$ cm. The diffusion constant D_s for a DNA rod of length a can be calculated from the rotational diffusion measurements by Hogan et al. (1978), if we assume the Broersma theory to be valid for translational as well as rotational diffusion (Broersma, 1960a,b). This gives $D_s = 1.7 \times 10^{-7}$ cm^2 s^{-1}, and the collision time from eq 5.15 is $\tau = 4 \times 10^{-3}$ s. The maximum transfer rate then is $\nu = 1/(2\tau) \simeq 100$ s^{-1}. This is admittedly crude, but the order of

Downloaded by UNIV MARYLAND COLG PARK on August 31, 2015 | http://pubs.acs.org
Publication Date: November 1, 1981 | doi: 10.1021/bi00527a028

Downloaded by UNIV MARYLAND COLG PARK on August 31, 2015 | http://pubs.acs.org
Publication Date: November 1, 1981 | doi: 10.1021/bi00527a028

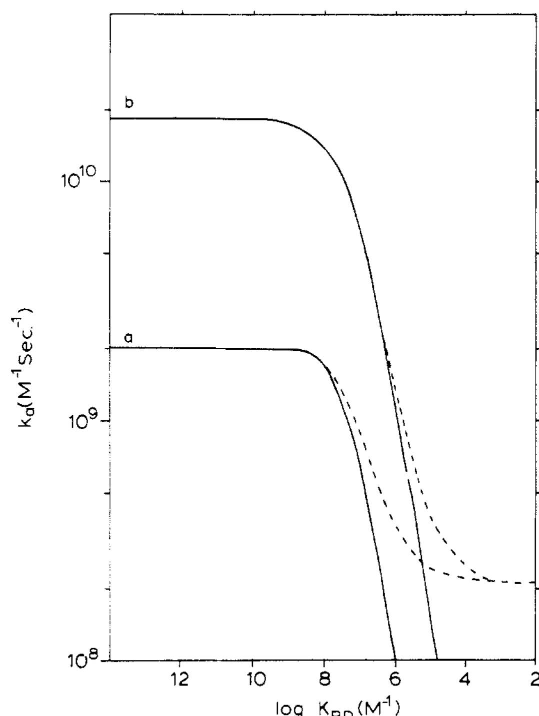


FIGURE 4: Total specific association rate constant as a function of the nonspecific binding constant for a *fixed* intersegment transfer rate, ν. For curve a, $\nu = 10^2$ s^{-1}. For curve b, $\nu = 10^3$ s^{-1}. The dashed portions of the curves are with hopping included, from eq A24. The upper break in the curves occurs at $K_{RD} = D_T^{-1}$, and a value of $D_T = 5 \times 10^{-8}$ M has been used throughout. The diffusion-controlled limit ($k_i \gg 2\pi Dl$) is assumed.

magnitude should be correct as an estimate of the largest possible rate.

The assumptions made tend to maximize the contribution of this transfer mechanism to the rate of the overall process: thus, (i) eq 5.10 overestimates the rate of the segmental motion at longer times; (ii) the average segment density is actually lower than that assumed since for a Gaussian chain only about two-thirds of the chain is really inside the volume defined by the radius of gyration; and (iii) no effects of steric hindrance on the transfer rate have been included. This relatively slow transfer rate also justifies the randomization assumption that underlies eq 5.9 since total chain relaxation (loss of spatial correlations between different segments) will occur on the same time scale as the transfer (Berg, 1979).

As long as the nonspecific association process (eq. 5.3) is not rate limiting, eq 5.2 and 5.9 yield

$$k_a = \frac{\nu K_{RD} + k_{assoc}}{1 + D_T K_{RD}} \simeq \begin{array}{ll} \nu K_{RD} + k_{assoc} & D_T K_{RD} \ll 1 \\ \nu / D_T & D_T K_{RD} \gg 1 \end{array} \quad (5.16)$$

As shown in Figure 4, this contribution to the overall rate is maximal when nonspecific binding is strong, in contrast to the unfacilitated result of Figure 3. In this limit, k_a becomes concentration dependent and consequently pseudo first order. Intersegment transfer is very inefficient for weak nonspecific binding; here transfer by nonspecific dissociation is faster. In Figure 4, we also plot the result (dashed curves) when the hopping process is included as defined by eq A24 of Appendix A. These curves are essentially the same as those one obtains when $k_{assoc}/(1 + D_T K_{RD})$ in eq 5.16 is replaced by k_a of eq 5.8; that is, hopping and intersegment transfer are independent and additive.

The result is chain length dependent primarily through the estimate of ν, which depends on the segment density in the coil. However, for very short chains—of the order of one persistence length or less—intersegment transfer as discussed here becomes impossible, and only the unfacilitated transfer mechanisms of eq 5.4 or 5.8 remain. Of course, for short chains at very high concentrations, one could observe transfer of the intersegment type between *different* chains instead. However, under these conditions, the overall rate of transfer would become *very* DNA concentration dependent.

Transfer by Sliding. The sliding model has been worked out in great detail. Richter & Eigen (1974) assumed that this mechanism could increase the association rate by increasing the target size to the DNA length over which the protein could slide without dissociating. Berg & Blomberg (1976, 1977) extended the sliding model by explicitly incorporating the coupling between free three-dimensional diffusion and one-dimensional diffusion along the chain. In this way they also included the competitive effect of the nonspecific sites as well as the time course of the sliding motion. As it turns out, these effects can totally dominate the *lac* repressor association at low salt concentrations when nonspecific binding is very tight. Schranner & Richter (1978) considered the effects of varying chain length, particularly for short chains. The effects of varying the ionic strength were described by Berg & Blomberg (1978),[3] and from the results of this paper we can identify

$$k_2 O_T = \frac{\Lambda}{(\Lambda L^2/D_1)^{1/2} \coth (\Lambda L^2/D_1)^{1/2} - 1} \; (\mathrm{s}^{-1}) \quad (5.17)$$

as the effective transfer rate to the specific site relevant for the general two-step scheme of section 3. D_1 is the proposed one-dimensional diffusion constant for the protein sliding along the chain. Equation 5.17 has also been calculated in Appendix A as the limiting case of a detailed model incorporating all of the suggested mechanisms.

This result was derived without considering microscopic dissociation processes. However, it can be shown that hopping has only a very marginal effect on the sliding result (cf. Appendix A or Berg & Blomberg, 1978). This is because sliding—when present—provides a much more efficient way of reaching nearby sites that may contain the specific target than do dissociation–reassociation processes.

The effective transfer rate defined in eq 5.17 incorporates both the effective target extension of the specific site and the time delay inherent in the sliding motion; it is simply the inverse of the total time required for nonspecific binding before the specific site is located (see eq A10). It is also strongly dependent on the nonspecific binding constant—through the dissociation rate constant Λ—as well as on the length of the DNA chains. There are two simple limits to this transfer rate:

$$k_2 O_T \simeq \begin{array}{ll} (\Lambda D_1/L^2)^{1/2} & (\Lambda L^2/D_1)^{1/2} \gg 1 \\ 3D_1/L^2 & (\Lambda L^2/D_1)^{1/2} \ll 1 \end{array} \quad (5.18)$$

The lower limit is valid at low salt (strong nonspecific binding) and represents the case where the protein cannot dissociate but has to traverse the entire DNA chain by sliding in order to find the target site. Using eq 5.17, one can obtain the total association rate from eq 5.2 as long as the upper estimate provided by eq 5.3 is not exceeded.

[3] Some misprints in this paper have been corrected in Berg & Blomberg (1979).

$$k_a = \left(\frac{k_{assoc}}{1 + D_T K_{RD}}\right)\left[\frac{M}{(\Lambda L^2/D_1)^{1/2} \coth (\Lambda L^2/D_1)^{1/2} - 1}\right] =$$

$2k_{assoc}[D_1/(\Lambda l^2)]^{1/2}$ when $D_T K_{RD} \ll 1$ and $(\Lambda L^2/D_1)^{1/2} \gg 1$ (5.19a)

$[k_{assoc} D_1/(l^2 D_T)]^{1/2} = k_a^{max}$ when $D_T K_{RD} = 1$ and $(\Lambda L^2/D_1)^{1/2} \gg 1$ (5.19b)

$2(\Lambda D_1/l^2)^{1/2}/D_T$ when $D_T K_{RD} \gg 1$ and $(\Lambda L^2/D_1)^{1/2} \gg 1$ (5.19c)

$3D_1/(L^2 O_T)$ when $D_T K_{RD} \gg 1$ and $(\Lambda L^2/D_1)^{1/2} \ll 1$ (5.19d)

$k_{assoc}/(1 + D_T K_{RD})$ is the unfacilitated limit from eq 5.4. Consequently, the factor $M/[(\Lambda L^2/D_1)^{1/2} \coth (\Lambda L^2/D_1)^{1/2} - 1]$ represents the *enhancement* due to sliding. The upper limit (eq 5.19a) is valid at high salt (weak nonspecific binding), which is also the limit at which our result agrees with that of Richter & Eigen (1974). Here the association rate constant is given by the unfacilitated rate, k_{assoc}, as in eq 5.4 times the *effective target extension* $[D_1/(\Lambda l^2)]^{1/2}$, which is simply the number of base pairs over which the protein can slide without dissociating. k_a of eq 5.19 then increases with increasing nonspecific binding affinity ($k_a \propto K_{RD}^{1/2}$) and eventually goes through a maximum for $D_T K_{RD} \simeq 1$, where $k_a \simeq [k_{assoc} D_1/(l^2 D_T)]^{1/2}$ (eq 5.19b). In the limit of eq 5.19c, k_a decreases with increasing nonspecific binding ($k_a \propto K_{RD}^{-1/2}$) until it reaches the limit (eq 5.19d), where it becomes independent of K_{RD}. Both of these low salt limits (eq 5.19c,d) are DNA concentration dependent; i.e., k_a corresponds to a pseudo-first-order reaction in these cases. At sufficiently low DNA concentrations k_a becomes large, and the first nonspecific association event (see eq 5.3) becomes the rate-limiting step. For sufficiently high DNA concentrations, the low salt limit (eq 5.19d) is strongly length dependent, $k_a \propto L^{-2}$. Also at lower DNA concentrations, where the limit defined by eq 5.3 is valid, k_a becomes length dependent; thus k_a increases with increasing chain length, and $k_a \propto r_g \propto L^{1/2}$. In contrast, the high salt limit (eq 5.19a) is hardly length dependent at all, with only a logarithmic length dependence through the segment density which enters the nonspecific association rate constant, k_{assoc}, via eq 4.6. Some representative examples of the total association rate as a function of nonspecific binding have been plotted in Figure 5.

For very short rodlike DNA chains—about one persistence length or shorter—the hierarchy of nonspecific rates described in section 3 will collapse. Since there are no domains other than those defined by the axial extension of the rod, there is no distinction between intradomain and interdomain dissociation events. These rodlike chains have been described by Schranner & Richter (1978) using spheroidal geometry. Our results agree with theirs if the dissociation distance R_c is chosen to be the chain length, $R_c = 2L$. In these terms, our equations would read

$$k_{-1} = \Lambda = \frac{\lambda(2\pi Dl/k_i)}{\ln (2L/b) + 2\pi Dl/k_i}$$
$$k_1 = k_{assoc} = \frac{2\pi Dl}{\ln (2L/b) + 2\pi Dl/k_i} \quad (5.20)$$

Furthermore, for the short chains, $D_T K_{RD}$ should always be negligible, i.e., $D_T K_{RD} \ll 1$. Consequently from eq 5.1 and 5.20 we obtain

$$k_a = 2k_{assoc}[D_1/(\Lambda l^2)]^{1/2} \tanh (\Lambda L^2/D_1)^{1/2} \quad (5.21)$$

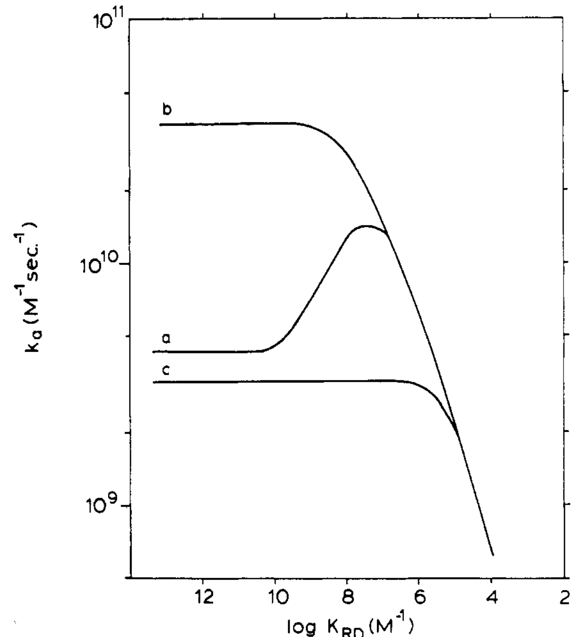


FIGURE 5: Specific association rate constant as a function of the nonspecific binding constant for various DNA chain lengths for the sliding model. (Curve a) M = 50 000 base pairs; long-chain behavior (eq 5.19) is exhibited. (Curve b) M = 6700 base pairs; an intermediate size chain that can still form a domain. The plateau level depends on the nonspecific association rate constant; see eq 5.3. (Curve c) M = 200 base pairs; this corresponds to a rodlike chain described by eq 5.21. The extreme length dependence of the final plateau levels at large values of K_{RD} is evident. Values of $O_T = 10^{-12}$ M and $D_1 = 9 \times 10^{-10}$ cm^2 s^{-1} have been used throughout. The diffusion-controlled limit ($k_i \gg 2\pi Dl$) is assumed.

In the diffusion-controlled limit ($k_i \gg 2\pi Dl$), this gives the explicit length dependence[4]

$$k_a = 4\pi D\left[\frac{D_1 K_{RD}}{2\pi Dl \ln (2L/b)}\right]^{1/2} \tanh \left[\frac{L^2 2\pi Dl}{D_1 K_{RD} \ln (2L/b)}\right]^{1/2} \quad (5.22)$$

At low salt concentrations (strong nonspecific binding, $(\Lambda L^2/D_1)^{1/2} \ll 1$), k_a is equal to the nonspecific association rate of repressor to the whole chain. That is, $k_a = Mk_{assoc}$ and k_a is roughly proportional to the chain length M. However, for very short DNA fragments, the effective diffusion rate (D) will be dominated by the diffusion constant of the DNA fragment itself, which is roughly inversely proportional to the DNA chain length. In this way, k_{assoc} can compensate for the factor M, and k_a becomes length independent for very short fragments at low salt concentrations. At higher salt concentrations, the whole DNA fragment no longer serves as an effective target, and the specific association rate decreases as the argument of the hyperbolic function in eq 5.21 becomes small.

In the same way, the expected specific dissociation rate constant can be calculated from eq 3.8 and 5.17 for the sliding model; thus for rodlike chains, one finds, using also eq 5.20, that

[4] As it turns out (O. G. Berg and M. Ehrenberg, unpublished results), eq 5.22 is a reasonable approximation to the correct expression (including hopping), even for very long rods. As expected, deviations appear for very short sliding lengths when hopping cannot be neglected. The decrease in k_a for very large L [proportional to $[\ln (2L/b)]^{-1/2}$] is an artifact in this approximation, although the effect is not as pronounced as in the result given by Schranner & Richter (1978) where k_a goes to zero with increasing L in proportion to $[\ln (2L/b)]^{-1}$.

Downloaded by UNIV MARYLAND COLG PARK on August 31, 2015 | http://pubs.acs.org
Publication Date: November 1, 1981 | doi: 10.1021/bi00527a028

Downloaded by UNIV MARYLAND COLG PARK on August 31, 2015 | http://pubs.acs.org
Publication Date: November 1, 1981 | doi: 10.1021/bi00527a028

$$k_d = \frac{k_{assoc}}{K_{RO}[\Lambda l^2/(4D_1)]^{1/2} \coth (\Lambda L^2/D_1)^{1/2} + K_{RD}} \quad (5.23)$$

Brief Comparison with Previous Theories. Our treatment of the sliding, model follows the coupled-diffusion approach of Berg & Blomberg (1976, 1977, 1978). The main difference from the original formulation by Richter & Eigen (1974) is the explicit diffusion–flux balancing at the chain surface through the use of a proper boundary condition. This serves as a basis for the definition of the hierarchy of nonspecific dissociation rates (Figure 2) and also makes it possible to include correlated transfer, e.g., hopping, in the description of the diffusion process. If our results are to agree in the limit of weak nonspecific binding, the unspecified dissociation rate used by Richter & Eigen should be identified with our intradomain dissociation rate constant, Λ. Schranner & Richter (1978) have also used a coupled-diffusion approach with particular emphasis on short chains. However, the flux-balancing approach employed by these authors still neglects correlated events (hopping). Also, the single dissociation rate constant used must, in fact, be length dependent. This is the reason why our result (eq 5.22) carries a weaker length dependence than that given by Schranner & Richter.

Lohman et al. (1978) have formulated a theory to describe the effects of salt concentration on a general two-step (transfer) scheme like that of eq 1.3, defining the binding parameters of the nonspecific transfer complex in terms of the approach of Record et al. (1976, 1978). However, they use only the weak-binding limit (eq 5.2a) of the preequilibrium result, in which k_a increases with increasing nonspecific binding affinity until the total upper limit (eq 5.3) is reached. Thus they neglect the concentration-dependent limit (eq 5.2b) of the process. Also, their assumption that the general transfer rate constant (k_2), is salt concentration independent does not hold. As is obvious from eq 5.17, the effective transfer rate in the sliding model *is* strongly salt concentration dependent as a consequence of the salt dependence of the nonspecific dissociation rate constant, Λ. The intersegment transfer rate can also be expected to be salt dependent, though in a less obvious way.

6. More Than One Specific Site per DNA Chain

Independent Target Sites. The results above need to be modified when several specific sites are present on each DNA molecule. If the experimental measurements can be interpreted to determine which specific site is occupied, the presence of extra sites will appear, in essence, as an extra source of binding competition. Here we consider the (real) case in which a complex is "counted" (retained on the filter) regardless of which of the specific sites is occupied.

There is, of course, no change in the nonspecific association processes described in section 4. The simplest case is that in which the specific sites are *independent* of one another; i.e., the probability of hitting any specific site is simply N/M, where N is the number of specific sites and $M \gg N$ is the total number of sites—specific and nonspecific—per chain. This will be the case for the intersegment transfer mechanism where all transfers are assumed uncorrelated. Then the total specific association rate constant per chain will be (cf. eq 3.5 and 5.9)

$$k_a^{obsd} = \left[(Mk_1)^{-1} + \left(N\frac{k_2D_TK_{RD}}{1 + D_TK_{RD}} \right)^{-1} \right]^{-1} \quad (6.1)$$

where $k_2D_T = \nu + \Lambda$. This is valid in the unfacilitated case as well when $\nu = 0$ (above).

When hopping is included, the specific sites will still be effectively independent unless they are very close together, i.e., offset by less than approximately 20 base pairs and thus effectively overlapping, since at this or a greater separation they cannot be bridged by a hopping process. As long as nonspecific association is not rate limiting, the total association rate to a chain with N *independent* specific sites is N-fold larger than that to a single site. We note that eq 6.1 actually gives the initial slope of the customary bimolecular plot (see case a of Figure 6).

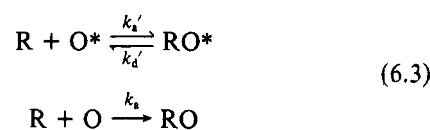
Similarly, the effective specific dissociation rate constant (per DNA chain) for N independent sites would be (cf. eq 3.8)

$$k_d^{obsd} = \gamma \left[(k_2D_T)^{-1} + \left(\frac{Mk_{-1}}{N + \gamma M} \right)^{-1} \right]^{-1} \quad (6.2)$$

As long as the first term is rate limiting, there will be no N dependence. Thus, for independent sites, a dependence on N appears primarily as a multiplicative factor in the effective association rate.

Two Specific (Independent) Sites. The association to a chain with several specific sites cannot be viewed as a proper bimolecular reaction. Let us consider in more detail the case for a system carrying one extra (secondary) specific site which has a weaker binding affinity than the primary one [see Winter & von Hippel (1981) and Winter et al. (1981)]. The extension to several sites is obvious. For simplicity, the discussions will be confined to the case for which nonspecific association is not rate limiting. The kinetics for several independent and equally strong sites have also been discussed in this limit, both for association (Giacomoni, 1979) and dissociation (Giacomoni, 1976).

In the following treatment, we use R to denote the protein, O the primary specific site, and O* the secondary (specific) site. If conditions are such that the protein–primary site complex is stable on the time scale of the experiment, the total association scheme involves two parallel reactions:



$$\mathrm{R} + \mathrm{O}^* \underset{k_d'}{\overset{k_a'}{\rightleftharpoons}} \mathrm{RO}^* \qquad \mathrm{R} + \mathrm{O} \xrightarrow{k_a} \mathrm{RO} \quad (6.3)$$

We assume further that the association rate constant to the secondary site is the same as that to the primary one; i.e., $k_a' = k_a$. *Total* concentrations of primary and secondary sites are the same and are set equal to O_T. We let the fractional saturations of the two operators be $\theta = [\mathrm{RO}]/O_T$ and $\theta^* = [\mathrm{RO}^*]/O_T$ and the ratio of protein molecules to DNA chains be $\mu = R_T/O_T$. Then kinetic scheme 6.3 gives

$$\frac{d\theta}{dt} = k_aO_T(\mu - \theta - \theta^*)(1 - \theta)$$
$$\frac{d\theta^*}{dt} = k_aO_T(\mu - \theta - \theta^*)(1 - \theta^*) - k_d'\theta^* \quad (6.4)$$

The sites were assumed independent; thus the probability P that a chain has at least one site occupied is

$$P = 1 - (1 - \theta)(1 - \theta^*) \quad (6.5)$$

The filter is then counted, and the results are calculated as if PO_T = concentration of bound complexes. The data are then plotted in the form of a bimolecular association:

$$k_a^{obsd}O_Tt = \frac{1}{\mu - 1} \ln \left[\frac{1 - P(t)/\mu}{1 - P(t)} \right] \quad (6.6)$$

Downloaded by UNIV MARYLAND COLG PARK on August 31, 2015 | http://pubs.acs.org
Publication Date: November 1, 1981 | doi: 10.1021/bi00527a028

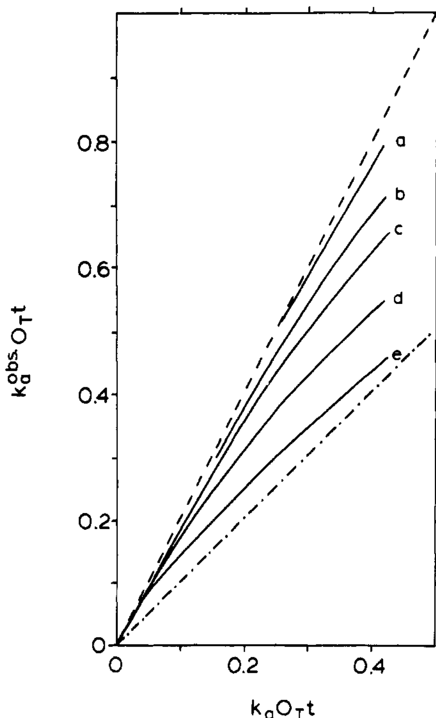


FIGURE 6: Effective association rate, $k_a^{obsd}O_Tt$ (calculated from eq 6.6 as a function of the rate of association to the primary site, k_aO_Tt), for various values of the rate constant for dissociation from the secondary site (k_d'). $R_T/O_T = 2$ is assumed throughout. (Curve a) For $k_d' = 0$ from eq 6.8; (curve b) for $k_d' = k_aO_T$; (curve c) for $k_d' = 2k_aO_T$; (curve d) for $k_d' = 5k_aO_T$; (curve e) for $k_d' = 20k_aO_T$. The dashed curve represents the "ideal" situation for $k_a^{obsd} = 2k_a$; the dotted curve represents this situation for $k_a^{obsd} = k_a$.

which defines the *observed* association rate constant, k_a^{obsd}.

If a repressor bound to the secondary site is stable on the time scale of the experiment, i.e., $k_d' < k_aO_T$, we can set $k_d' = 0$. The kinetic equations are then easily solved to give

$$1 - \theta = 1 - \theta^* = \frac{1 - 2/\mu}{\exp[(\mu - 2)k_aO_Tt] - 2/\mu} \qquad \mu \neq 2 \tag{6.7}$$

and the usual limit when binding sites and protein are equimolar is

$$1 - \theta = 1 - \theta^* = \frac{1}{2k_aO_Tt + 1} \qquad \mu = 2 \tag{6.8}$$

Inserting eq 6.5 into eq 6.6 and plotting, we obtain an approximately straight line with the slope $k_a^{obsd} = 2k_a$ (see Figure 6, curve a). Thus, in this case, the observed association rate is twice the rate for each site, as would be expected. However, as we shall see below, the requirement that the secondary complex be stable on the time scale of the experiment is not very restrictive.

When $k_d' \neq 0$, the kinetic equations (eq 6.4) can only be solved numerically. Such solutions have been carried out for $\mu = 2$ and various values of $k_d'/(k_aO_T)$. When this latter ratio becomes larger, the influence from the secondary site becomes smaller. However, as seen in Figure 6, even a rather weak site can be important, and this "secondary site" effect will be concentration dependent. Such an effect for the *lac* repressor has recently been observed by Pfahl et al. (1979). Thus the slower association observed by these workers on deletion of the secondary site could be a consequence of the measurement process, rather than indicating any real change in primary operator affinity as suggested by these authors.

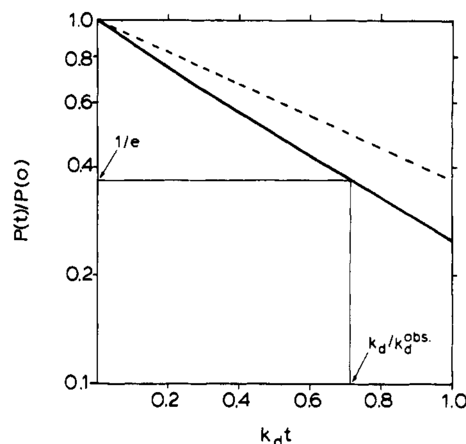


FIGURE 7: Fraction of operator-containing DNA molecules bound to the filter as a function of k_dt calculated by using eq 6.14 with $\alpha = 2$. The solid line is the result for $\beta = 2$, which gives close to the maximum deviation from a single exponential. The dashed line is for $\beta = 1$, which gives a single exponential. The definition of k_d^{obsd} from eq 6.15 is also indicated.

We can calculate the overall dissociation rate constant similarly. Let k_d be the rate constant for dissociation from the primary site. Then in a dissociation experiment

$$\theta(t) = \theta_0 \exp(-k_dt)$$
$$\theta^*(t) = \theta_0^* \exp(-k_d't) \tag{6.9}$$

where θ_0 and θ_0^* are the fractions of the primary and secondary sites, respectively, which are complexed at time $t = 0$. Thus, from eq 6.5, the probability, $P(t)$, that a chain has at least one site occupied at time t is given by

$$\frac{P(t)}{P(0)} = \frac{\exp(-k_dt) + (\theta_0^*/\theta_0)\exp(-k_d't) - \theta_0^* \exp(-k_dt - k_d't)}{1 + (\theta_0^*/\theta_0) - \theta_0^*} \tag{6.10}$$

If the binding sites are far from saturated initially, then θ_0, $\theta_0^* \ll 1$ and eq 6.10 can be approximated by

$$\frac{P(t)}{P(0)} = \frac{\exp(-k_dt) + (\theta_0^*/\theta_0)\exp(-k_d't)}{1 + (\theta_0^*/\theta_0)} \tag{6.11}$$

With the assumption that the dissociation process starts from equilibrium, the relative degree of saturation of the two sites is

$$\frac{\theta_0}{\theta_0^*} = \left(\frac{k_a}{k_d}\right)\left(\frac{k_d'}{k_a'}\right)\left(\frac{1-\theta_0}{1-\theta_0^*}\right) \simeq \left(\frac{k_a}{k_a'}\right)\left(\frac{k_d'}{k_d}\right) \tag{6.12}$$

This approximation again is based on the assumption that the sites are not initially saturated. Let the ratios of the association and dissociation rate constants be

$$\alpha \equiv k_a'/k_a \qquad \beta \equiv k_d'/k_d \tag{6.13}$$

If both associations are diffusion limited, α will be of the order of one.

Equations 6.11–6.13 then give

$$\frac{P(t)}{P(0)} = \frac{\beta}{\alpha + \beta}\left[\exp(-k_dt) + \frac{\alpha}{\beta}\exp(-\beta k_dt)\right] \tag{6.14}$$

This represents a single exponential decay process only for $\beta \ll 1$ or $\beta = 1$. However, it can also appear as an approximately straight line in a semilog plot for other situations (cf.

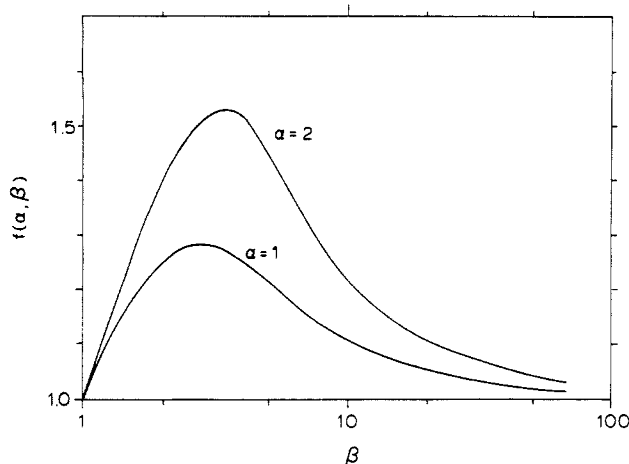


FIGURE 8: Correction factor $f(\alpha,\beta)$ from eq 6.15 and 6.16 as a function of β, for $\alpha = 1$ and $\alpha = 2$.

Downloaded by UNIV MARYLAND COLG PARK on August 31, 2015 | http://pubs.acs.org
Publication Date: November 1, 1981 | doi: 10.1021/bi00527a028

Figure 7). The total dissociation rate, k_d^{obsd}, can be defined as the reciprocal of the time for which $P(t)/P(0) = 1/e$ (e is the base of the natural logarithms); i.e., from eq 6.14

$$\frac{\beta}{\alpha+\beta}\left[\exp(-k_d/k_d^{obsd}) + \frac{\alpha}{\beta}\exp(-\beta k_d/k_d^{obsd})\right] = \frac{1}{e} \quad (6.15)$$

A numerical investigation reveals that

$$k_d^{obsd} = k_d f(\alpha,\beta) \quad (6.16)$$

where $f(\alpha,\beta)$ is a correction factor given in Figure 8 as a function of β for $\alpha = 1$ and $\alpha = 2$. The value of $f(\alpha,\beta)$ is always larger than one, and for β/α greater than approximately 5, it can be represented by

$$f(\alpha,\beta) \simeq [1 - \ln(1 + \alpha/\beta)]^{-1} \quad (6.17)$$

Thus the influence of a secondary site on the observed dissociation rate constant will be rather small unless the rate of association to the secondary site is faster than that to the primary one. In contrast, the *observed* association rate constant can be strongly perturbed even by a relatively weak secondary site.

Several "Dependent" Target Sites. When sliding is present, sites which are far apart can also become "nonindependent" if they are close enough together to be "connected" by one sliding event. Again, let us consider the case for which nonspecific association is *not* rate limiting. Also, for simplicity, we consider only the case for which the sliding distance is much smaller than the total chain length; i.e., end effects can be neglected. (The general result is given in Appendix A, eq A21.) Then the association rate to a *single* site is

$$k_a = 2k_{assoc}[D_1/(\Lambda l^2)]^{1/2}/(1 + D_T K_{RD}) \quad (6.18)$$

from eq 5.19a–c. $[D_1/(\Lambda l^2)]^{1/2}$ is the *sliding length* or the *effective target extension*, i.e. the distance the protein can slide without dissociation. When there are two specific sites located n_0 base pairs apart, the effective target length to the *left* of the left site and to the *right* of the right site will extend $[D_1/(\Lambda l^2)]^{1/2}$ base pairs in each direction. For a protein binding nonspecifically somewhere *between* the sites, the probability P_0 of association to either specific site without an intervening dissociation will be

$$P_0 = 2[D_1/(\Lambda L_0^2)]^{1/2}\tanh[\Lambda L_0^2/(4D_1)]^{1/2} \quad (6.19)$$

where $L_0 = n_0 l$ is the distance between the sites. The effective target extension between the sites is n_0P_0. Thus, the total specific association rate to the chain is

$$k_a^{obsd} = k_a[1 + \tanh[\Lambda L_0^2/(4D_1)]^{1/2}] \quad (6.20)$$

where k_a is the association rate constant for a *single* site as defined by eq 6.18.

This two-site result is easily extended to any number of specific sites; e.g., for N equally spaced (distance L_0 apart) sites, we obtain

$$k_a^{obsd} = k_a[1 + (N-1)\tanh[\Lambda L_0^2/(4D_1)]^{1/2}] \quad (6.21)$$

When nonspecific binding is weak, $[\Lambda L_0^2/(4D_1)]^{1/2} \gg 1$, and the sites become independent, i.e., $k_a^{obsd} = Nk_a$. For strong nonspecific binding, i.e., $N[\Lambda L_0^2/(4D_1)]^{1/2} \ll 1$, the target sites effectively merge and $k_a^{obsd} = k_a$. [Belintsev et al. (1980) have also discussed these limiting cases for the binding of RNA polymerase to T7 promoters.]

The effective dissociation rate can also be calculated for such situations. For a group of *N identical* sites with equal intersite spacing (distance L_0), we find (Appendix B) that

$$k_d^{obsd} = k_d\tanh[\Lambda L_0^2/(4D_1)]^{1/2}\left\{1 - \frac{1}{N}\left[\frac{1-\exp[-N(\Lambda L_0^2/D_1)^{1/2}]}{\sinh(\Lambda L_0^2/D_1)^{1/2}}\right]\right\}^{-1} \quad (6.22)$$

where k_d is the dissociation rate for a single specific site. This result is valid in an unsaturated case where initially not more than one protein is bound to each chain.

Equations 6.21 and 6.22 conform only approximately to

$$k_a^{obsd}/k_d^{obsd} = Nk_a/k_d \quad (6.23)$$

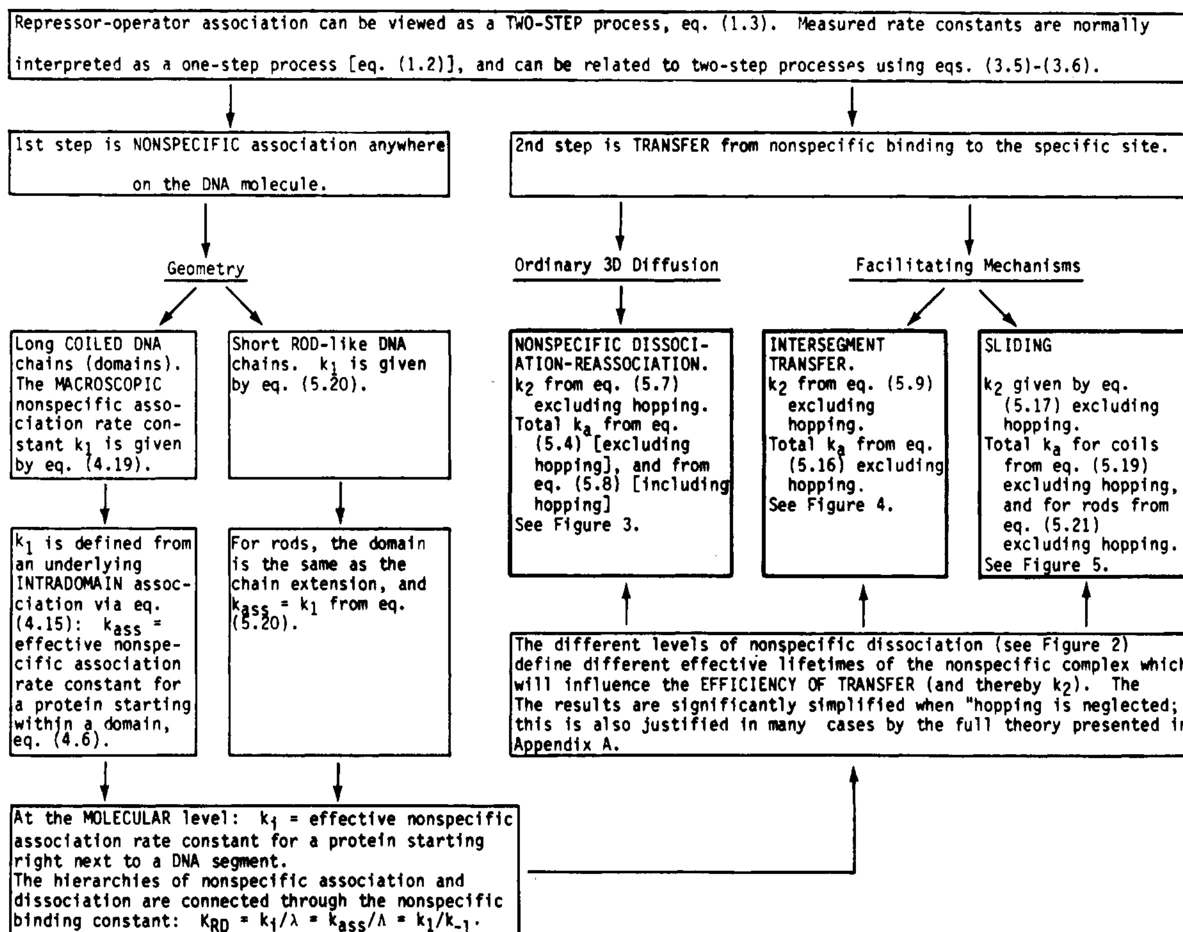
For independent sites eq 6.23 is valid as long as nonspecific binding is not rate limiting, and then the dependence on N comes in only through k_a^{obsd}; see eq 6.1 and 6.2. The effective specific equilibrium binding constant to the chain will not be a simple product as in eq 6.23 (C. P. Woodbury and P. H. von Hippel, unpublished results).

When sliding occurs, the effective association and dissociation rates as a function of the number of sites per chain will be strongly dependent on the distance between the sites and on the strength of the nonspecific binding (specifically, on salt concentration). This provides another important test for the sliding model. Sadler et al. (1980) have interpreted their data for *lac* repressor dissociating from tandem operators as suggestive of a sliding mechanism. The expressions provided above make possible a quantitative interpretation of these data [see Winter et al. (1981)].

7. Conclusions

In this paper, we have developed a general theory for the kinetics of interaction of genome regulatory proteins with their physiologically relevant target sites on the DNA chromosome. To facilitate its use, the logical connections and key equations which tie together the basic theory are outlined in the form of a "flow chart" in Chart I. The theory has been developed in terms of the *E. coli lac* repressor–operator system, in particular, to permit easy comparison with experimental results in the following paper (Winter et al., 1981). However, the results are quite general for any biological system where the search for a target or receptor site can be speeded up by utilizing nonspecific binding to reduce the dimensionality (or volume) of the search process. [A direct two-dimensional analogue to the effectively one-dimensional genome-regulatory protein–DNA system can be found in ligand–membrane interactions, as previously pointed out by Richter & Eigen (1974).] The purely electrostatic nonspecific binding of *lac*

Chart I: "Flow Chart" for the Two-Step Association Theory



Downloaded by UNIV MARYLAND COLG PARK on August 31, 2015 | http://pubs.acs.org
Publication Date: November 1, 1981 | doi: 10.1021/bi00527a028

repressor to DNA, coupled with the Record et al. (1976, 1978) approach to the interpretation of such binding, can lead to particularly effective protein translocation mechanisms [see Winter et al. (1981)]. Such binding interactions are almost certainly involved in other genome regulatory systems and may well operate in membrane–ligand systems as well, given that most of the lipid constituents of the membrane bilayer feature highly charged head groups on the membrane surface. The electrostatic basis of the binding of protein to nonspecific sites within the overall target molecule or structure provides an easily manipulated (via salt concentration) experimental variable; similar manipulation should be possible in other systems.

Detailed tests of some of the theoretical (and physical) models of protein translocation mechanisms developed here, as well as further comments on the biological relevance of such notions, are included in Winter et al. (1981).

Added in Proof

More detailed calculations (O. G. Berg and P. H. von Hippel, unpublished results) employing the intrachain reaction theory of Wilemshi and Fixman (Wilemshi & Fixman, 1974; Doi, 1975) indicate that the diffusion-controlled intrachain transfer rate constant (ν) could possibly be as large as 10^4 s^{-1} under some conditions. This result is based on the same correlation function (eq 5.10) used above (Berg, 1979), but supercedes the crude analogy used to derive eq 5.15. The conclusions reached in the companion paper (Winter et al., 1981) for mechanisms involved in the translocation of *lac* repressor are not affected by this revised estimate. However, this result does strengthen our view that intersegment transfer could play an important role in other systems [see General Discussion, Winter et al. (1981)].

Appendix

(*A*) *Derivation of a Complete Description for the Kinetics of Repressor–Operator Binding. General Formulation.* For completeness, we describe here the mathematical model which includes all the mechanisms discussed above, i.e., intersegment transfer, sliding, and local "hopping". These calculations also provide a justification for the intuitive identification of the effective transfer rate in eq 5.7 and 5.9.

Sliding has previously been described as a continuous diffusion process. When the other effects are included in the formalism, it is more convenient to consider the chain as built up by discrete binding sites; it can be assumed that each base pair constitutes the beginning of a new binding site. For simplicity, it is also assumed that the specific site is in the middle of the chain. Then the base pairs can be numbered with the discrete coordinate j, $-m \leq j \leq m$, with $j = 0$ denoting the specific site. Thus, the total number of sites is $M = 2m + 1$ per chain, and we shall be interested in the case $M \gg 1$ such that end effects are unimportant.

When a protein is nonspecifically bound at site j, it can slide to a neighboring site ($j - 1$ or $j + 1$) with a rate constant Γ_1 (s^{-1}); it can also be translocated directly by intersegment transfer with a rate constant ν (s^{-1}). (As pointed out above, this intersegment transfer mechanism can be treated as totally uncorrelated; that is, the probability of going from site j to j' is independent of the locations of j and j'.) A third possibility, of course, is that the protein dissociates from site j with the rate constant λ (s^{-1}). As sites j and $-j$ are symmetrical with respect to the operator site, they can be considered together. Let $u_j(t)$ be the probability that the protein is bound

at site j (or $-j$) at time t. Then the "time evolution" expression is determined by

$$\frac{du_j}{dt} = \Gamma_1(u_{j+1} - 2u_j + u_{j-1}) + \frac{\nu}{m}\sum_{\substack{j'=1,\\ j'\neq j}}^{m} u_{j'} - \nu u_j[1 - 1/(2m)] - \lambda u_j + \sum_{j'=1}^{m}\lambda\int_0^t F_{jj'}(t-t')u_{j'}(t')\,dt' + G(t) \quad (A1)$$

This expression is analogous to that for continuous sliding (Berg & Blomberg, 1976).

The terms in eq A1 can be identified as follows: the first corresponds to sliding, the second represents intersegment transfer to site j from any other site j', the third corresponds to intersegment transfer from site j to any other site (except $-j$), and the fourth corresponds to the return to site j at time t of a protein that had dissociated from site j' at time t', with $F_{jj'}(t-t')$ expressing the probability for such a return. Finally, the fifth term, $G(t)$, represents the rate for the *first* attachment to the chain, which should be independent of j. From section 4 this is simply

$$G(t) = 2k_1O_T\exp(-k_1D_Tt) \quad (A2)$$

The factors $F_{jj'}$ carry all the correlations between sites and describe the microscopic dissociation events ("hopping") as well as reassociations to segments on the same or a different chain. Consequently, as functions of t, these factors are represented by complicated expressions; however, fortunately, only the time-integrated forms are needed to derive the overall operator-association rate.

At the operator site the protein will be absorbed. This can be described by the absorbing boundary condition $u_0(t) = 0$ which gives

$$\frac{du_0}{dt} = \Gamma_1u_1 + \frac{\nu}{2m}\sum_{j'=1}^{m}u_{j'} + \lambda\sum_{j'=1}^{m}\int_0^t F_{0j'}(t-t')u_{j'}(t')\,dt' + k_1O_T\exp(-k_1D_Tt) \quad (A3)$$

Summing equations (eq A1) over j from 1 to m, one finds the flux into the operator from (A3):

$$\frac{du_0}{dt} = -\sum_1^m\frac{du_j}{dt} - \sum_1^m\lambda u_j + \lambda\sum_{j=0}^{m}\sum_{j'=1}^{m}\int_0^t F_{jj'}(t-t')u_{j'}(t')\,dt' + k_1D_T\exp(-k_1D_Tt) \quad (A4)$$

This expression describes the probability flux into any operator site for *one* protein starting somewhere in a solution of DNA chains with total concentration O_T of operator sites and D_T of nonspecific sites.

Mean Times. The mean time, τ, for operator association is found to be

$$\tau \equiv \int_0^\infty t\frac{du_0}{dt}\,dt = \sum_1^m\int_0^\infty u_j(t)\,dt + \lambda\sum_{j=0}^{m}\sum_{j'=0}^{m}\int_0^\infty tF_{jj'}(t)\,dt\int_0^\infty u_{j'}(t)\,dt + \frac{1}{k_1D_T} \quad (A5)$$

Here the condition $\sum_{j=0}^m\int_0^\infty F_{jj'}(t)\,dt = 1$ has been used, which simply expresses the fact that a dissociated protein will eventually reassociate *somewhere*. Equation A5 demonstrates that the total association time is a sum over the residence times in the intermediate states. Thus, the first term is the total mean residence time for nonspecific binding, the second term is the total mean time in solution between dissociation events, and the last term is the mean time before the *first* nonspecific association. The factor

$$\sum_{j=0}^{m}\int_0^\infty tF_{jj'}(t)\,dt \equiv \tau_2 \quad (A6)$$

is the mean time of reassociation anywhere for a protein dissociating from j'. This must be related to the nonspecific binding constant such that

$$\tau_2 = 1/(\lambda K_{RD}D_T) = 1/(k_iD_T) \quad (A7)$$

where k_i is the microscopic association rate constant from eq 4.1. Intuitively, this relation is obvious since a binding constant must express the ratio of the mean time the protein is bound (λ^{-1}) to the mean time (τ_2) that it is free after a dissociation. This has been demonstrated previously in this context (Berg & Blomberg, 1976, 1977) and for the more general situation (Berg, 1978). Thus

$$\tau = 1/(k_1D_T) + \tau_2(1 + D_TK_{RD})\sum_1^m\lambda\tilde{u}_j \quad (A8)$$

where

$$\tilde{u}_j \equiv \int_0^\infty u_j(t)\,dt$$

is the total mean time the protein is bound at site j before the final association at the operator. $\sum\lambda\tilde{u}_j$ can be interpreted as the mean number of nonspecific dissociations needed before the operator is found. Thus, for determination of the mean operator association time, only the time-integrated expressions are needed.

From the mean time of association, we can define the association rate constant:

$$k_a \equiv \frac{1}{\tau O_T} = \frac{D_TK_{RD}/(O_T\sum\tilde{u}_j)}{1 + D_TK_{RD} + 1/(k_{-1}\sum\tilde{u}_j)} \quad (A9)$$

The structure of this expression is identical with that of eq 3.6, which shows that this detailed model is compatible with the more common formulation of the two-step scheme. Thus, we can identify the effective transfer rate to the operator:

$$k_2O_T = 1/\sum\tilde{u}_j \quad (A10)$$

This is again a reasonable result, with the transfer rate to the specific site being simply the inverse of the total mean time for nonspecific binding.

From (A1) one finds

$$0 = \Gamma_1(\tilde{u}_{j+1} - 2\tilde{u}_j + \tilde{u}_{j-1}) + \frac{\nu}{m}\sum_{j'=1}^{m}\tilde{u}_{j'} - \nu\tilde{u}_j[1 + 1/(2m)] - \lambda\tilde{u}_j + \lambda\sum_{j'=1}^{m}\tilde{F}_{jj'}\tilde{u}_{j'} + 2/M \qquad j = 1, ..., m \quad (A11)$$

Here

$$\tilde{F}_{jj'} \equiv \int_0^\infty F_{jj'}(t)\,dt = 2/M + \sum_{n=1}^{\infty}\psi_n[4/(n\pi)]\sin(n\pi/M)\cos(2jn\pi/M)\cos(2j'n\pi/M) \quad (A12)$$

is the probability for a return to site j (or $-j$) if the protein had dissociated from site j'. This expression has been calculated by considering the free diffusion outside the chain, and the Fourier coefficients (ψ_n) are the same complicated functions of chain geometry as those calculated previously (Berg & Blomberg, 1977, 1978):

$$\psi_n = 1 - \left[1 + \left(\frac{k_i}{2\pi Dl}\right)\times\left(\frac{1}{n\beta}\right)\left[\frac{I_0(n\alpha)K_0(n\beta) - K_0(n\alpha)I_0(n\beta)}{I_0(n\alpha)K_1(n\beta) - K_0(n\alpha)I_1(n\beta)}\right]\right]^{-1} \quad (A13)$$

Downloaded by UNIV MARYLAND COLG PARK on August 31, 2015 | http://pubs.acs.org Publication Date: November 1, 1981 | doi: 10.1021/bi00527a028

where $\alpha \equiv \pi R_c/L$, $\beta \equiv \pi b/L$, and I and K are modified Bessel functions.

Since $\tilde{F}_{jj'}$ couples the continuous diffusion in free space with the discrete sites on the chain, there is some ambiguity in its construction. The expression (eq A12) was calculated for a protein that dissociates from the middle of site j' and reassociates somewhere within the length l of site j.

Solution without Hopping. A simple solution of (A11) can be achieved with the approximation

$$\tilde{F}_{jj'} = (2/M)(1 - \psi) + \psi\delta_{jj'} \qquad \text{(A14)}$$

where $\delta_{jj'}$ is the *Kronecker* δ. This means that a fraction ψ of the total number of proteins dissociated return to the same site and that the rest are distributed with equal probability to any site. Thus, the fraction ψ is just the reassociation probability defined in eq 4.2, or equivalently, it is the limit $n = 0$ for the Fourier coefficients in eq A13. As discussed in section 5, this approximation disregards the possibility of a site change during short-lived microscopic dissociations, but this is not important if a sliding or intersegment-transfer mechanism dominates the process. Using eq A14, we find from eq A11 that

$$\tilde{u}_j = \frac{4}{M}\sum_{n=1}^{m} \frac{1 - \cos(2jn\pi/M)}{4\Gamma_1 \sin^2(n\pi/M) + \nu M/(2m) + \lambda(1 - \psi)} \qquad \text{(A15)}$$

which satisfies a reflecting boundary condition at the chain end, $j = m$, and the absorbing condition at $j = 0$ (the specific site). Then

$$\sum_{j=1}^{m} \tilde{u}_j = \sum_{n=1}^{m} \frac{2}{4\Gamma_1 \sin^2(n\pi/M) + \Lambda + \nu} \qquad \text{(A16)}$$

where $\Lambda = \lambda(1 - \psi)$ has been introduced from eq 4.3, and $M = 2m + 1 \simeq 2m \gg 1$ has been assumed. The summation over n can be carried (cf. Jolley, 1961, eq 485), giving for the effective transfer rate from eq A10

$$k_2 O_T = (\Lambda + \nu)[M \tanh(\Omega) \coth(M\Omega) - 1]^{-1} \qquad \text{(A17)}$$

where

$$\Omega \equiv \ln\left[\left(\frac{\Lambda + \nu}{4\Gamma_1}\right)^{1/2} + \left(1 + \frac{\Lambda + \nu}{4\Gamma_1}\right)^{1/2}\right]$$

In the limit of no sliding, ($\Gamma_1 = 0$), this gives

$$k_2 O_T = (\Lambda + \nu)/M \qquad \text{(A18)}$$

as in eq 5.9. If sliding dominates, $\Gamma_1 \gg \Lambda + \nu$, and then

$$\Omega \simeq \left(\frac{\Lambda + \nu}{4\Gamma_1}\right)^{1/2} \ll 1$$

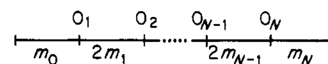
$$k_2 O_T = \frac{\Lambda + \nu}{M\left(\frac{\Lambda + \nu}{4\Gamma_1}\right)^{1/2} \coth\left[M\left(\frac{\Lambda + \nu}{4\Gamma_1}\right)^{1/2}\right] - 1} \qquad \text{(A19)}$$

With the identification $D_1 = \Gamma_1 l^2 = \Gamma_1 4L^2/M^2$, this coincides with eq 5.17 when intersegment transfer does not contribute ($\nu = 0$):

$$k_2 O_T = \frac{\Lambda}{(\Lambda L^2/D_1)^{1/2} \coth(\Lambda L^2/D_1)^{1/2} - 1} \qquad \text{(A20)}$$

Thus, the discrete sliding model gives the same result as the continuous one, except when the nonspecific dissociation rate Λ is as large as, or larger than, the step rate, Γ_1. The advantage of (A17) is that $k_2 O_T$ levels off at the correct unfacilitated value, Λ/M, when $\nu = \Gamma_1 = 0$, whereas in eq A19, $k_2 O_T$ is zero in this limit.

Also the case for which there are several specific sites on each DNA chain can be treated in the same fashion. Consider the following arrangement of N specific sites $O_1, O_2, ..., O_N$:



where m_0 and m_N are distances (in base pairs) to the ends of the chain, and $2m_1, 2m_2, ..., 2m_{N-1}$ are distances between the specific sites. The total number of sites—specific and nonspecific—is

$$M = m_0 + m_N + 2\sum_{i=1}^{N} m_i + N$$

Then the total mean time of nonspecific binding before *one* of the specific sites is found will be

$$\sum_{1}^{M} \tilde{u}_j = (\Lambda + \nu)^{-1}\{[(M \tanh \Omega)/[\tfrac{1}{2} \tanh(2m_0 + 1)\Omega + \sum_{i=1}^{N-1} \tanh(2m_i + 1)\Omega + \tfrac{1}{2} \tanh(2m_N + 1)\Omega]] - 1\} \qquad \text{(A21)}$$

where Ω is the same as in eq A17. The total specific association rate to the chain is still given by (A9), with O_T identified as the concentration of DNA chains. The various limits for independent and dependent sites discussed in section 6 then follow directly. Also the case of one specific site not placed in the middle of the chain is defined by eq A21.

General Solution. The solution for the general expression for $\tilde{F}_{jj}$ can be derived from eq A12 in the same way. One finds

$$\sum_{1}^{m} \tilde{u}_j = 2\sum_{n=1}^{m}\left[4\Gamma_1 \sin^2\left(\frac{n\pi}{M}\right) + \frac{\nu M}{2m} + \lambda - \lambda\frac{M}{\pi}\sin\left(\frac{n\pi}{M}\right)\left\{\frac{1}{n}\psi_n - \sum_{i=1}^{\infty}(-1)^i\left(\frac{\psi_{iM-n}}{iM - n} - \frac{\psi_{iM+n}}{iM + n}\right)\right\}\right]^{-1} \qquad \text{(A22)}$$

This reduces to eq A16 when all $\psi_m = \psi$. To get useful results from this expression, we have to make certain approximations. First, it is noted that the infinite sum over i involves ψ factors with large subscripts, the smallest one being m. From (A13), one finds that in the limit of large subscripts

$$\psi_n \xrightarrow[n > m]{} \frac{1}{1 + n\pi[4\pi Db/(Mk_i)]}$$

Then the summation over i can be carried out. Neglecting terms of order $(2\pi Dl/k_i)^2$ and higher (i.e., in the diffusion-controlled limit when $k_i \gg 2\pi Dl$), one finds

$$\sum_{i=1}^{\infty}(-1)^i\left(\frac{\psi_{iM-n}}{iM - n} - \frac{\psi_{iM+n}}{iM + n}\right) \simeq \frac{1}{n} - \frac{\pi/M}{\sin(n\pi/M)} \qquad \text{(A23)}$$

Then eq A22 reduces to

$$\sum_{1}^{m} \tilde{u}_j = 2\sum_{n=1}^{m}\left[4\Gamma_1 \sin^2\left(\frac{n\pi}{M}\right) + \nu + \lambda(1 - \psi_n)\frac{M}{n\pi}\sin\left(\frac{n\pi}{M}\right)\right]^{-1} \qquad \text{(A24)}$$

and from (A10) the effective transfer rate to the specific site,

Downloaded by UNIV MARYLAND COLG PARK on August 31, 2015 | http://pubs.acs.org
Publication Date: November 1, 1981 | doi: 10.1021/bi00527a028

k_2O_T, is the inverse of this expression.

In the limit of no facilitating mechanism, $\Gamma_1 = \nu = 0$, one obtains

$$k_2O_T = \lambda\left[2\sum_{n=1}^{m}\frac{n\pi/M}{(1-\psi_n)\sin(n\pi/M)}\right]^{-1} \quad (A25)$$

Then, using the Fourier coefficients from eq A13 in the diffusion-controlled limit ($k_i \gg 2\pi Dl$), one finds the total association rate from eq A9, A10, and A25 to be

$$k_a \simeq \frac{k_2D_TK_{RD}}{1 + D_TK_{RD}} \simeq \frac{4\pi bD}{1 + D_TK_{RD}}\left[\frac{2}{M}\sum_{n=1}^{m}\left[\frac{1}{\sin(n\pi/M)}\right]\times\left[\frac{I_0(n\alpha)K_0(n\beta) - K_0(n\alpha)I_0(n\beta)}{I_0(n\alpha)K_1(n\beta) - K_0(n\alpha)I_1(n\beta)}\right]\right]^{-1} \quad (A26)$$

A numerical investigation shows that for long chains ($L \gg l$ and $R_c \gg b$), the result is approximately

$$k_a \simeq 1.5\pi bD/(1 + D_TK_{RD}) \quad (A27)$$

The numerical factor ($\sim$1.5) depends primarily on l/b and is essentially independent of R_c if $R_c \gg b$. A similar result has been derived in a standard steady-state analysis of the rate of association of a protein to a specific site on an infinitely long cylinder (O. G. Berg and M. Ehrenberg, unpublished results), thus demonstrating more clearly that the hopping process is an integral part of the diffusion equation.

As discussed in section 5, this is a highly idealized result; however, it does show how the microscopic dissociation process can contribute by relieving steric constraints.

(*B*) *Complete Derivation for Tandem Operators "Connected" by Sliding.* Consider the case where N operators are positioned in tandem sequence. In accord with the treatment above, the repressor will recognize only its exact binding positions. Thus, these specific sites will be separated by a stretch of nonspecific sites, the number of which is at least the number of base pairs in one operator. Sliding provides a means for the repressor to transfer between adjacent operators. However, during such a transfer, it will be nonspecifically bound and consequently more easily lost in solution. To calculate the total lifetime for a repressor bound to such a group of operators, one must first know the relevant transfer and dissociation rates in the gaps *between* the operators.

Consider a gap of length L_0 flanked by absorbing barriers (the operators). The repressor, while nonspecifically bound in the gap, is characterized by a one-dimensional diffusion coefficient D_1 along the chain and a dissociation rate constant, Λ. Thus, the probability distribution $u(x,t)$ for the repressor in the interoperator gap satisfies the relation

$$\frac{\partial u}{\partial t} = D_1\frac{\partial^2 u}{\partial x^2} - \Lambda u \qquad 0 < x < L_0 \quad (B1)$$

with the absorbing boundary conditions

$$u(0,t) = u(L_0,t) = 0 \quad (B2)$$

If the repressor starts just outside the operator at the left, what is the probability, P_c, that it crosses the gap and finds the operator at the right? What is the probability, P_d, that it dissociates from the chain and is lost in solution? The initial condition for the diffusion equation is

$$u(x,0) = \delta(x - l) \quad (B3)$$

which assumes that at $t = 0$ the repressor has moved from the specific site at $x = 0$ to the nearest nonspecific site, $x = l$. The probabilities, P_c and P_d, are given by

$$P_c = -D_1\frac{\partial\tilde{u}}{\partial x}\bigg|_{x=L_0} \quad (B4)$$

$$P_d = \Lambda\int_0^{L_0}\tilde{u}(x)\,dt \quad (B5)$$

where

$$\tilde{u}(x) \equiv \int_0^\infty u(x,t)\,dt$$

satisfies

$$\frac{\partial^2\tilde{u}}{\partial x^2} - \frac{\Lambda}{D_1}\tilde{u} = \frac{1}{D_1}\delta(x - l) \quad (B6)$$

with the same boundary conditions as in (B2). The Green's function solution is

$$\tilde{u}(x) = (\Lambda D_1)^{-1/2}\frac{\sinh[q(L_0 - l)]\sinh(qx)}{\sinh(qL_0)} \qquad 0 < x < l$$

$$\tilde{u}(x) = (\Lambda D_1)^{-1/2}\frac{\sinh(ql)\sinh[q(L_0 - x)]}{\sinh(qL_0)} \qquad l < x < L_0$$

where $q = (\Lambda/D_1)^{1/2}$ is the inverse of the diffusion distance. This gives

$$P_d = 1 - \cosh(ql) + \tanh(qL_0/2)\sinh(ql) \quad (B7)$$

$$P_c = \sinh(ql)/\sinh(qL_0) \quad (B8)$$

The interesting case is that for which $ql = (\Lambda l^2/D_1)^{1/2} \ll 1$; i.e., the repressor can slide a distance much longer than 1 base pair along the chain without falling off. Then

$$P_d = (\Lambda l^2/D_1)^{1/2}\tanh[\Lambda L_0^2/(4D_1)]^{1/2} \quad (B9)$$

$$P_c = (\Lambda l^2/D_1)^{1/2}/\sinh(\Lambda L_0^2/D_1)^{1/2} \quad (B10)$$

When $L_0 \rightarrow \infty$, $P_d \rightarrow (\Lambda l^2/D_1)^{1/2}$. Thus, the dissociation rate constant from a single operator is

$$k_d = 2(\gamma D_1/l^2)(\Lambda l^2/D_1)^{1/2} = 2\gamma(\Lambda D_1/l^2)^{1/2}$$

where $\gamma D_1/l^2$ is the rate for the "elementary" step from the operator onto the closest nonspecific site, and the factor 2 takes care of the possibility of dissociating both to the left and the right. Consequently, we can express the transfer rate from one operator to an adjacent one as

$$k_c = (\gamma D_1/l^2)P_c = \tfrac{1}{2}k_d/\sinh(\Lambda L_0^2/D_1)^{1/2} \quad (B11)$$

and the effective dissociation rate via nonspecific binding in the gap as

$$k_g = (\gamma D_1/l^2)P_d = \tfrac{1}{2}k_d\tanh(\Lambda L_0^2/(4D_1))]^{1/2} \quad (B12)$$

We note that it would be more consistent to consider the diffusion along the chain as a discrete random walk over the nonspecific sites. Although this problem can be solved in a similar manner, the result is represented by the following unwieldy sums:

$$P_c = \frac{1}{m}\sum_{k=1}^{m-1}\frac{\sin^2(k\pi/m)(-1)^{k+1}}{1 + \Lambda/(2\Gamma_1) - \cos(k\pi/m)}$$

$$P_d = 1 - \frac{1}{m}\sum_{k=1}^{m-1}\frac{\sin^2(k\pi/m)[1 - (-1)^k]}{1 + \Lambda/(2\Gamma_1) - \cos(k\pi/m)}$$

where $m - 1$ is the number of nonspecific sites in the gap, and Γ_1 is the rate of the elementary step in going from one of these sites to the next. Substituting $m = L_0/l$ and $\Gamma_1 = D_1/l^2$, it can be shown that these sums are well approximated by eq B7 and B8. This is true in particular for large m, and in this case, m must be equal to or larger than the number of base

Downloaded by UNIV MARYLAND COLG PARK on August 31, 2015 | http://pubs.acs.org
Publication Date: November 1, 1981 | doi: 10.1021/bi00527a028

pairs in one operator ($m > \sim 25$), which is sufficient. The approximation is not as valid for values of $\Lambda > \sim D_1/l^2$, but this case is not interesting in terms of the sliding model as used with the *lac* system.

We are now ready to consider the whole group of N operators. We label them $n = 1, 2, ..., N$, and let $P_n(t)$ be the probability that the repressor is bound to site n at time t. Then the probability distribution is governed by the following master equations:

$$\frac{dP_1}{dt} = k_c P_2 - (\tfrac{1}{2}k_d + k_g + k_c)P_1$$

$$\frac{dP_n}{dt} = k_c(P_{n-1} + P_{n+1}) - 2(k_g + k_c)P_n \qquad n = 2, 3, ..., N-1$$

$$\frac{dP_N}{dt} = k_c P_{N-1} - (\tfrac{1}{2}k_d + k_g + k_c)P_N \qquad \text{(B13)}$$

The repressor bound at site n can be transferred to either site $n + 1$ or site $n - 1$ with a rate constant k_c, or it can be lost via dissociation from the interoperator gaps with a rate constant k_g. The end operators, $n = 0$ and $n = N$, have a different dissociation probability at their free sides, i.e., $^1/_2k_d$, which is just one-half the total dissociation rate for a single operator.

Under the assumption that the repressor is bound somewhere (homogeneously) within this group of sites at time $t = 0$, what is the mean time, τ_N, before it has dissociated from this group altogether? This can be calculated as

$$\tau_N = \sum_{n=1}^{N} \tilde{P}_n \qquad \text{(B14)}$$

where $\tilde{P}_n \equiv \int_0^\infty P_n(t)\, dt$ is the total mean time spent at site n before dissociating from the group. Integrating eq B13 with $P_n(0) = 1/N$ for all n as an initial condition gives

$$-1/N = k_c\tilde{P}_2 - (\tfrac{1}{2}k_d + k_g + k_c)\tilde{P}_1$$

$$-1/N = k_c(\tilde{P}_{n-1} + \tilde{P}_{n+1}) - 2(k_g + k_c)\tilde{P}_n \qquad n = 2, 3, ..., N-1$$

$$-1/N = k_c\tilde{P}_{N-1} - (\tfrac{1}{2}k_d + k_g + k_c)\tilde{P}_N \qquad \text{(B15)}$$

Summing these equations and using the fact that $\tilde{P}_N = \tilde{P}_1$ for symmetry reasons, we obtain, using eq B14

$$\tau_N = \frac{1}{2k_g}[1 - (k_d - 2k_g)\tilde{P}_1] \qquad \text{(B16)}$$

with $\tilde{P}_1$ still remaining to be determined. This could be accomplished by calculating all $\tilde{P}_n$ values recursively using eq B15. As only $\tilde{P}_1$ is needed, a *much* simpler way is to use the generating function G(z) defined as

$$G(z) = \sum_{n=1}^{N} z^{n-1}\tilde{P}_n \qquad \text{(B17)}$$

We multiply each equation n in the set B15 by z^{n-1} and sum all equations. This gives

$$G(z)[(z + z^{-1})k_c - 2(k_g + k_c)] = -\frac{1}{N}\left(\frac{1 - z^N}{1 - z}\right) + k_c\tilde{P}_1(z^N + z^{-1}) - (k_c + k_g - \tfrac{1}{2}k_d)\tilde{P}_1(z^{N-1} + 1) \qquad \text{(B18)}$$

This relation is valid for all values of the dummy variable, z. In particular, we can *choose*

$$z = z_1 = (1 + k_g/k_c) + [(1 + k_g/k_c)^2 - 1]^{1/2} \qquad \text{(B19)}$$

which makes the left-hand side of eq B18 equal to zero. $\tilde{P}_1$ can then be solved to give (for τ_N from eq B16)

$$\tau_N = \frac{1}{2k_g}\left[1 - \frac{(2/N)(1 - Z_1^N)/(1 - z_1)}{1 + z_1^{N-1} + (z_1^{-1} - 1 + z_1^N - z_1^{N-1})k_c/(\tfrac{1}{2}k_d - k_g)}\right] \qquad \text{(B20)}$$

With the expression B19 for z_1, this gives the mean dissociation time for all possible choices of k_c, k_g, and k_d. This result is considerably simplified by using eq B11 and B12, which apply to the sliding model. Then z_1 is simply $z_1 = \exp(\Lambda L_0^2/D_1)^{1/2}$, and τ_N reduces to

$$\tau_N = \tau_1 \coth [\Lambda L_0^2/(4D_1)]^{1/2} \left\{1 - \frac{1}{N}\left[\frac{1 - \exp[-N(\Lambda L_0^2/D_1)^{1/2}]}{\sinh (\Lambda L_0^2/D_1)^{1/2}}\right]\right\} \qquad \text{(B21)}$$

where $\tau_1 = k_d^{-1}$ is the dissociation time for a single operator.

References

Adam, G., and Delbrück, M. (1968) in *Structural Chemistry and Molecular Biology* (Rich, A., & Davidson, N., Eds.) p 198, W. H. Freeman, San Francisco, CA.

Barkley, M. D. (1981) *Biochemistry 20*, 3833.

Barkley, M. D., Lewis, P. A., & Sullivan, G. E. (1981) *Biochemistry 20*, 3842.

Belintsev, B. N., Zavriev, S. K., & Shemyakin, M. F. (1980) *Nucleic Acids Res. 8*, 1391.

Berg, O. G. (1978) *Chem. Phys. 31*, 47.

Berg, O. G. (1979) *Biopolymers 18*, 2161.

Berg, O. G., & Blomberg, C. (1976) *Biophys. Chem. 4*, 367.

Berg, O. G., & Blomberg, C. (1977) *Biophys. Chem. 7*, 33.

Berg, O. G., & Blomberg, C. (1978) *Biophys. Chem. 8*, 271.

Berg, O. G., & Blomberg, C. (1979) *Biophys. Chem. 9*, 415.

Broersma, S. (1960a) *J. Chem. Phys. 32*, 1626.

Broersma, S. (1960b) *J. Chem. Phys. 32*, 1632.

deHaseth, P. L., Lohman, T. M., & Record, M. T., Jr. (1977) *Biochemistry 16*, 4783.

Doi, M. (1975) *Chem. Phys. 9*, 455.

Epstein, I. R. (1979) *Biopolymers 18*, 2037.

Giacomoni, P. U. (1976) *FEBS Lett. 72*, 83.

Giacomoni, P. U. (1979) *Eur. J. Biochem. 98*, 557.

Hogan, M., Dattagupta, N., & Crothers, D. M. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 195.

Jolley, L. B. W. (1961) *Summation of Series*, Dover Publications, New York.

Kornberg, A. (1980) *DNA Replication*, W. H. Freeman, San Francisco, CA.

Kowalczykowski, S. C., Lonberg, N., Newport, J. W., Paul, L. S., & von Hippel, P. H. (1980) *Biophys. J. 32*, 403.

Lohman, T. M. (1980) *Biophys. J. 32*, 458.

Lohman, T. M., & Kowalczykowski, S. C. (1981) *J. Mol. Biol.* (in press).

Lohman, T. M., deHaseth, P. L., & Record, M. T., Jr. (1978) *Biophys. Chem. 8*, 281.

Manning, G. S. (1978) *Q. Rev. Biophys. 11*, 179.

O'Gorman, R. B., Dunaway, M., & Matthews, K. S. (1980) *J. Biol. Chem. 255*, 10100.

Pfahl, M., Gulde, V., & Bourgeois, S. (1979) *J. Mol. Biol. 127*, 339.

Downloaded by UNIV MARYLAND COLG PARK on August 31, 2015 | http://pubs.acs.org
Publication Date: November 1, 1981 | doi: 10.1021/bi00527a028

Record, M. T., Jr., Lohman, T. M., & deHaseth, P. L. (1976) *J. Mol. Biol. 107*, 145.
Record, M. T., Jr., Anderson, C. F., & Lohman, T. M. (1978) *Q. Rev. Biophys. 72*, 103.
Revzin, A., & von Hippel, P. H. (1977) *Biochemistry 16*, 4769.
Richter, P. H., & Eigen, M. (1974) *Biophys. Chem. 2*, 255.
Riggs, A. D., Bourgeois, S., & Cohn, M. (1970) *J. Mol. Biol. 53*, 401.
Sadler, J. R., Tecklenburg, M., & Betz, J. L. (1980) *Gene 8*, 279.
Schranner, R., & Richter, P. H. (1978) *Biophys. Chem. 8*, 135.
Schurr, J. M. (1979) *Biophys. Chem. 9*, 413.
von Hippel, P. H. (1979) in *Biological Regulation and Development* (Goldberger, R. F., Ed.) pp 279–347, Plenum Press, New York.
von Hippel, P. H., Revzin, A., Gross, C. A., & Wang, A. C. (1975) in *Protein-Ligand Interactions* (Sund, H., & Blauer, G., Eds.) pp 270–288, Walter de Gruyter, Berlin.
Wilemshi, G., & Fixman, M. (1974) *J. Chem. Phys. 60*, 866.
Winter, R. B., & von Hippel, P. H. (1981) *Biochemistry* (second paper of three in this issue).
Winter, R. B., Berg, O. G., & von Hippel, P. H. (1981) *Biochemistry* (third paper of three in this issue).

Diffusion-Driven Mechanisms of Protein Translocation on Nucleic Acids. 2. The *Escherichia coli* Repressor–Operator Interaction: Equilibrium Measurements[†]

Robert B. Winter[‡] and Peter H. von Hippel*

Downloaded by UNIV MARYLAND COLG PARK on August 31, 2015 | http://pubs.acs.org
Publication Date: November 1, 1981 | doi: 10.1021/bi00527a028

ABSTRACT: In this paper the equilibrium binding of *lac* repressor to operator sites has been studied as a function of monovalent salt concentration, of length of the DNA molecule containing the operator, and (by using various natural *lac* "pseudo"-operators) of operator base pair sequence. The nitrocellulose filter assay has been used to obtain values of repressor–operator association constants (K_{RO}), both directly and as ratios of association to dissociation rate constants (k_a/k_d). Measurements of K_{RO} have been made in the absence of Mg^{2+} or other divalent ions, allowing a direct estimate [Record, M. T., Jr., Lohman, T. M., & deHaseth, P. L. (1976) *J. Mol. Biol. 107*, 145] of the contribution of electrostatic (charge–charge) interactions to the stability of the RO complexes. Using *lac* operator containing DNA restriction fragments of known size, we have shown the following: (i) The magnitide of the RO interaction is salt concentration dependent. A plot of log K_{RO} vs. log [KCl] is linear over the 0.1–0.2 M KCl range, and from the slope of this plot, we can determine that RO complex formation involves six to seven charge–charge interactions. This value is independent of operator type and of DNA fragment size for fragments greater than ~170 base pairs in length. (ii) This number of charge–charge interactions is appreciably less than the 11 such interactions involved in RD complex formation [deHaseth, P. L., Lohman, T. M., & Record, M. T., Jr. (1977) *Biochemistry 16*, 4783; Revzin, A., & von Hippel, P. H. (1977) *Biochemistry 16*, 4769], suggesting that repressor binds to operator and to nonoperator DNA in different conformations. (iii) The RO interaction involves a substantial (>50%) nonelectrostatic component of the binding free energy, in contrast to the RD interaction for which all the binding free energy appears to be electrostatic in nature. (iv) The binding constant (K_{RO_2}) for the secondary (*lacZ* gene) pseudooperator is 5-fold weaker than K_{RO_1} for the primary (physiological) operator when both are measured on separate pieces of DNA. When both operators are on the *same* piece of DNA, the measured value of K_{RO_2} is ~25-fold smaller than that of K_{RO_1}. (v) K_{RO_3}, the binding constant for the tertiary (I gene) pseudooperator, has been estimated at $<10^{10}$ M^{-1} at salt concentrations where $K_{RO_1} \simeq 10^{13}$ M^{-1}. (vi) K_{RO_1} for repressor binding to short DNA fragments is smaller than that for binding to long DNA fragments under the same environmental conditions. Several of these findings, together with others in the literature, are suggestive of "long-range" effects on RO binding constants; possible molecular bases for such effects are discussed. These measurements provide the equilibrium "underpinnings" of our analysis of RO kinetic binding mechanisms [Winter, R. B., Berg, O. G., & von Hippel, P. H. (1981) *Biochemistry* (following paper in this issue)] and also allow comparisons of repressor binding affinities for operator, pseudooperator, and nonoperator DNA. In addition, these results further demonstrate the importance of the surrounding (nonspecific) DNA in controlling the equilibrium stability as well as the rates of formation and dissociation of RO complexes.

In the preceding paper (Berg et al., 1981), we described and quantitatively formulated theoretical models for mechanisms of diffusional (thermal fluctuation driven) translocation of proteins on nucleic acids. As pointed out in that paper, in order to test the applicability of these theories to a real system, one must determine the equilibrium and kinetic parameters for the binding of the protein to specific target sites and to nonspecific sites as a function of salt concentration and of the length of the DNA molecules that contain the target sites. In this paper we report some relevant equilibrium measurements for the *Escherichia coli lac* repressor–operator system; additional equilibrium measurments on this system have also recently been presented by Barkley et al. (1981). Equilibrium pa-

[†] From the Institute of Molecular Biology and Department of Chemistry, University of Oregon, Eugene, Oregon 97403. *Received April 20, 1981.* A portion of this work was presented by R.B.W. to the Graduate School of the University of Oregon in partial fulfillment of the requirements for the Ph.D. degree in Chemistry. This research was supported in part by U.S. Public Health Service Grant GM-15792 (to P.H.v.H.). R.B.W. was a predoctoral trainee on U.S. Public Health Service Training Grants GM-00715 and GM-07759.
[‡] Present address: Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, CO 80302.

 © 1981 American Chemical Society