Obstacle Bypass in Protein Motion along DNA by Two-dimensional Rather than One-dimensional Sliding*[

Martin Kampmann‡
From the MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, United Kingdom

Site-specific DNA-binding proteins locate their target sites by facilitated diffusion. Several proteins have been shown to slide along DNA in vitro. However, whereas sliding is often envisaged as one-dimensional tracking of the DNA major groove, such a mechanism would not allow linear diffusion over long distances in vivo, where short stretches of free DNA are delimited by bound proteins. I propose a two-dimensional sliding mechanism, in which the protein diffuses freely on the cylindrical DNA surface, and I present experiments that can distinguish between one- and higher-dimensional diffusion along the DNA contour length. At 100 mM NaCl, translocation of EcoRI restriction endonuclease between sites on two DNA helices connected by a Holliday junction is as efficient as between sites on the same helix, indicating a three-dimensional mechanism. At 25 mM NaCl, translocation between sites on the same DNA helix is more efficient, indicating a role for sliding at low ionic strength. Obstacles attached to the major groove of one face of the DNA helix did not interfere with sliding, regardless of their orientation relative to the cleavage sites. This result is compatible with two-dimensional but not one-dimensional sliding. As illustrated by Monte-Carlo simulation, two-dimensional sliding may not only allow proteins to move around nucleosomes in vivo but also reduce the redundancy of their search for the target site.

The binding of proteins to specific DNA sequences is a process central to life. Its thermodynamic basis, the network of noncovalent protein-DNA interactions that confer specificity, is well characterized for a large number of cases. However, the kinetics and mechanism of target location by DNA-binding proteins are not understood in detail.

The simplest kinetic model for complex formation between a protein (P) and a specific target site (T) on the DNA would be as follows.

\[ P + T \rightleftharpoons P\cdot T \]  
(Eq. 1)

However, some of the measured association rates are 1000-fold higher than the theoretical diffusion-controlled rate for this model (1). Results from many kinetic studies suggest that nonspecific DNA sequences flanking a specific target site facilitate target location (2). These findings are compatible with the generally accepted two-step binding model involving nonspecific DNA (N).

\[ P + N + T \rightleftharpoons P\cdot N + T \rightleftharpoons P\cdot T + N \]  
(Eq. 2)

The first step is fast because the concentration of nonspecific sites is much higher than that of specific sites. The second step requires a transfer mechanism between sites on the DNA molecule.

Two main classes of hypothetical transfer mechanisms can be distinguished (3): 1) Mechanisms based on dissociation and reassociation of the protein. This class comprises microscopic dissociation-reassociation events (“hopping”), where the protein usually reassociates with the site of dissociation or with a site in the immediate vicinity thereof, and macroscopic dissociation-reassociation events involving diffusion through three-dimensional (3D) space and binding to an uncorrelated site. 2) Mechanisms during which the protein remains DNA-bound. This class comprises “sliding,” which is defined as diffusion of the protein along the contour length of the DNA, and “direct transfer” of the protein between two non-contiguous DNA segments that are transiently both bound by the same protein molecule. Whereas dissociation-reassociation mechanisms can be expected to be available to all DNA-binding proteins, only proteins that meet certain criteria may use the mechanisms of the second category. In particular, “direct transfer” requires the protein to have two DNA-binding sites.

It is important to note that sliding does not necessarily amount to one-dimensional (1D) diffusion of a protein along the major groove of the DNA, although early definitions of sliding seem to imply 1D random walks (3) and experimentally observed diffusion along the contour length is often interpreted or labeled as 1D diffusion. Instead, a protein may be free to diffuse on the two-dimensional (2D) cylindrical surface of the DNA between sites that are not adjacent in the primary sequence of the DNA. Although other authors have referred to this mechanism as sliding “along one face of the DNA” (4), I prefer the more general term “2D sliding,” especially in light of the experimental results obtained in this study. Theoretical considerations have been used recently to support 1D sliding for BamHI restriction endonuclease (5). No experiments that can distinguish between 1D and 2D sliding have yet been published.

Another unsolved question is whether facilitated diffusion as studied in vitro occurs in vivo: 1D sliding over distances of more than 1000 bp is still a commonly accepted mechanism for facilitated diffusion (2); it is incompatible, however, with eukaryotic chromatin as a substrate. Physiologically relevant models of target location by DNA-binding proteins will have to account for the bypass of other DNA-bound proteins.

Here, I present experiments that address these questions.

* This work was supported by the German Academic Exchange Service (DAAD) and the German National Merit Foundation (Studienstiftung). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed: The Rockefeller University, 1230 York Ave., New York, NY 10021. Tel.: 212-327-8101; Fax: 212-327-7880; E-mail: martin.kampmann@rockefeller.edu.

The on-line version of this article (available at http://www.jbc.org) contains supplemental software source code.

1 The abbreviations used are: 3D, three-dimensional; 1D, one-dimensional; 2D, two-dimensional.
Using EcoRI endonuclease as a model system, I show that the protein can bypass obstacles during translocation on DNA; this result rules out 1D diffusion for distances as short as 21 bp (7 nm). 2D diffusion is seen only at low ionic strength. Using computer simulation, I analyze the physiological implications of 2D diffusion, which have not been discussed in the previous literature.

EXPERIMENTAL PROCEDURES

DNA Substrates—Oligonucleotides corresponding to the two strands of the DNA substrates were chemically synthesized by Charles Hill (Department of Biochemistry, University of Cambridge); all sequences are listed in Table I. Full-length strands were purified by high pressure liquid chromatography and by denaturing gel electrophoresis, and the purity was monitored by capillary electrophoresis. Both strands were radioactively labeled at their 5' ends using [γ-32P]ATP (6). Equal amounts of two oligonucleotides were mixed in annealing buffer (125 mM NaCl, 5 mM MgCl2, and 10 mM Tris-HCl, pH 7.5), heated at 80 °C for 10 min and allowed to cool slowly to room temperature. Folding of substrates containing a Holliday junction structure was analyzed by native gel electrophoresis as described elsewhere (7), except that the running buffer contained 40 mM Tris acetate and 0.1 mM MgCl2.

Cleavage Reactions and Analysis—EcoRI restriction endonuclease and the appropriate diluent buffer were purchased from New England Biolabs. The enzyme was diluted to 2.5 nM immediately before use. 5 nM DNA substrate and 50 pM enzyme were incubated at 37 °C in 50 mM Tris-HCl, pH 7.5, 5 mM MgCl2, and NaCl as indicated. Samples were taken at specified time points and mixed with an equal volume of deionized formamide containing 0.1% bromphenol blue, 0.1% xylene cyanol, and 10 mM EDTA to stop the reaction. These samples were heated at 90 °C for 5 min, and a fraction corresponding to 25 fmol of DNA substrate was subjected to electrophoresis at 50 W on a sequencing-type denaturing 16% polyacrylamide gel. The gels were fixed in 10% methanol, 10% acetic acid for 10 min, dried, and exposed to a PhosphorImager screen (Amersham Biosciences). The radioactivity of the different bands was quantified using ImageQuant software (Amersham Biosciences), corrected for differential labeling of the two strands, and the signals corresponding to the different bands were plotted as a function of time and fitted by linear regression.

Monte Carlo Simulation—Simulations were implemented in Fortran 90 and followed the rules described in the legend to Fig. 4 B and in the text. The source code of the computer programs is available as supplemental online material.

RESULTS

Monitoring Protein Motion on DNA by Measuring Processivity—DNA substrates containing two cleavage sites for a given restriction endonuclease have been used in a number of studies to investigate mechanisms of protein motion on the DNA molecule. I follow the nomenclature introduced by Terry et al. (8) and Stanford et al. (9) for the different cleavage events and the resulting fragments (Fig. 1). Short 5'-end labeled linear DNA molecules containing two EcoRI cleavage sites were incubated in the presence of EcoRI. Upon encounter with an enzyme molecule, different cleavage events will result in the production of different radiolabeled fragments, the relative amounts of which can be quantified after denaturing gel electrophoresis (Fig. 1).

Each event produces defined amounts of the different fragments. The total amounts of the fragments, A, C, AB, and BC are therefore a function of the total occurrence of the different
events that create them, $E_1$, $E_2$, $E_{12}$, and $E_{21}$, as expressed by the following set of equations.

\[ A = E_1 + E_{12} + E_{21} \]  
\[ C = E_2 + E_{12} + E_{21} \]  
\[ AB = E_2 \]  
\[ BC = E_1 \]

From these equations, it is possible to derive the following expression.

\[ A + AB = C + BC \]

This cleavage model describes encounters between one protein and an uncleaved DNA substrate. It therefore reflects the initial phase of the reaction. In all experiments described, the enzyme was incubated with a 100-fold molar excess of substrate to minimize encounters between proteins and partially cleaved DNA fragments. The rates of production of the radiolabeled fragments, $v_A$, $v_C$, $v_{AB}$, and $v_{BC}$, were measured during the initial linear phase, in which less than 10% of the total DNA was digested.

Equation 7 can be formulated in terms of the following initial rates.

\[ v_A + v_{AB} = v_C + v_{BC} \]

Processive cleavage of two sites on one DNA molecule rather than stochastic distributive cleavage of sites on different DNA molecules indicates facilitated diffusion of the restriction endonuclease between the two sites. Processivity is therefore a quantitative measure of the efficiency of protein motion along the DNA. Processivity $p$ is defined as the fraction of proteins that, after cutting one site, also cut the second site of their DNA substrate. It can therefore be expressed as follows.

\[ p = \frac{E_{12} + E_{21}}{(E_1 + E_A + E_{12} + E_{21})} \]

Substitution yields the following.

\[ p = \frac{v_A + v_C - v_{AB} - v_{BC}}{(v_A + v_C + v_{AB} + v_{BC})} \]

It is generally assumed that after the cleavage reaction, the enzyme stays bound to one of the two DNA fragments, whereas the other fragment diffuses away. Therefore, the theoretical upper limit for processivity on linear substrates is 0.5.

The Effect of Obstacles on Processivity—Obstacles that physically impair protein diffusion between the two cleavage sites on the DNA substrate can generally be expected to reduce the processivity. Whether a particular obstacle affects the processivity, however, will depend on the mode of protein diffusion. Therefore, investigation of the effect that different obstacles have on processivity can yield information about the mechanism of protein motion along DNA.

I designed four DNA substrates, as shown in Fig. 2A. Substrate NO is the control substrate without an obstacle. The spacing of the two EcoRI cleavage sites is 21 bp; the sites are therefore located on the same face of the DNA helix. In substrates OA and OS, an obstacle is created by a 6-bp DNA helix capped at both ends with tetraloops, which is connected to the DNA substrate via a four-way junction. In substrate OS, the obstacle faces the same way as the EcoRI recognition sequence in the major groove; in substrate OA, the obstacle faces away from the cleavage sites. Fig. 2, B and C, shows an atomic model of the tertiary structure of the obstacle. The obstacle should not interfere with nonspecific binding of EcoRI endonuclease to the opposite face of the DNA molecule. Substrates OA and OS were obtained by annealing two oligonucleotides of different lengths. The strands were complementary at their ends, but the longer strand contained an additional interior stretch of 20 nucleotides that was designed to fold into the obstacle described above. The short DNA helix was stabilized by 5 CG pairs, as well as GTAA tetraloops, which are known to be particularly stable (10). The sequence at the four-way junction included the trinucleotide ACC, which has been shown to stabilize specific pair-wise stacking of the four DNA helices (11). In substrate FJ, the helices around the conserved four-way junction were rewired such that the two EcoRI sites no longer lie on a continuous DNA helix. The spacing of the sites was chosen such that an enzyme bound to either of the cleavage sites would not experience steric hindrance by the rest of the DNA substrate. The relative mobility of the four DNA substrates during native gel electrophoresis supports their predicted structures with respect to helical stacking (Fig. 2D). It cannot be excluded that the short helix forming the obstacle assumes a slightly different structure, possibly an ensemble of structures, or undergoes thermal fluctuations. In any case, it will present an obstacle to 1D sliding of EcoRI that is localized to one face of the DNA helix.

All substrates were incubated with EcoRI in a buffer containing 25 or 100 mM NaCl. The production of cleavage fragments over time was plotted and $v_A$, $v_C$, $v_{AB}$, and $v_{BC}$ were determined by linear fitting of the initial phase of the reaction (Fig. 3A). Processivity $p$ was calculated as described above. Fig. 3B shows the values for $p$ under different substrate and ion conditions. For substrate NO, the processivity at 25 mM NaCl is close to the theoretical limit of 0.5. At higher ionic strength, the processivity is reduced. There are no significant differences in processivity between the four substrates at 100 mM NaCl. The obstacles in substrates OA and OS do not have an effect on processivity at 25 mM NaCl, either. However, the processivity for substrate FJ at 25 mM NaCl is significantly lower than for the other substrates.

Comparison of 1D and 2D Diffusion of Proteins on DNA by Monte-Carlo Simulation—As I will discuss in more detail below, the experimental results seem to rule out 1D sliding of EcoRI along the major groove of the DNA as a functionally significant mechanism of facilitated diffusion. The results, however, are compatible with 2D sliding at low ionic strengths. In a 2D sliding mode, the protein would move between sites on the DNA but remain nonspecifically bound to it. However, the protein would not track the major groove as in 1D sliding, but it would freely diffuse across the entire cylindrical surface of the DNA. I used Monte-Carlo simulation to compare how efficiently proteins would scan the DNA for their target site using 1D versus 2D diffusion. In a given amount of time determined by the dissociation rate, the protein can sample a given number of sites. Target location will be more efficient and less redundant when the number of different sites sampled is high.

For 1D diffusion on naked DNA, the protein had the same probability of stepping forward or backward along the DNA major groove (Fig. 4A, blue). 2D diffusion was modeled such that each random walk step was either along the major groove or orthogonal to the major groove (Fig. 4A, yellow). Coordinates outside the major groove are likely to be energetically less favorable positions for the protein and represent transition states between the energetically more favorable positions that coincide with major groove sites. The time spent at major groove positions is likely to be much greater than the time spent at all other positions, and dissociation from a major groove site rather than the translocation time between sites will be rate-limiting. The simulation was therefore based on the simplifying assumption that pathways between two major groove sites take the same amount of time irrespective of the length of the pathway.
In vivo, DNA is bound to nucleosomes and other proteins that occlude parts of the major groove. In Fig. 4A, green and red show proteins following the same rules for 1D and 2D diffusion, respectively, on a DNA substrate bound to an obstacle. The simulation assumes that the diffusing protein can still access half of the DNA surface in the nucleosome-bound region; the actual accessible surface in physiological protein-bound DNA is likely to be even smaller because of steric hindrance. Fig. 4B shows the number of different major groove sites sampled as a function of all site-sampling events. On naked DNA, 2D diffusion is more efficient than 1D diffusion. On nucleosomal DNA, proteins using 1D diffusion can sample only the linker DNA between two nucleosomes (55 bp in the simulation). 2D diffusion on nucleosomal DNA is however highly efficient; it is even more efficient than 2D diffusion on naked DNA.

**DISCUSSION**

1D Sliding versus 2D Sliding—1D sliding along the major groove is an appealing model for target location of DNA-binding proteins. Although most authors favor this model, no experimental evidence specifically supports it (2). A recently published computational study (5) predicts 1D sliding for BamHI restriction endonuclease. The most favorable position for BamHI associated with nonspecific DNA is calculated to be at a distance of 34 Å from the DNA axis, facing the major groove. Thus, the protein may be able to slide along the DNA on a water cushion. The authors calculate an energy barrier of 2.6 kcal/mol for the movement of the protein from its position over the major groove to the minor groove and conclude that the protein will therefore follow a helical path along the major groove during diffusion to avoid this energy barrier. In their

**Fig. 2. Overview of DNA substrates.** A, substrate names are given and the constituting oligonucleotides are indicated. Complete sequences are listed in Table I. The tertiary structure is shown schematically. White lines delimit major and minor grooves. Arrows emerging from the major groove represent the dyad axis of EcoRI restriction sites. Secondary structure representations reveal the intended base pairing in the central region of the DNA substrate. A box is drawn around the conserved four-way junction core with known helical stacking of substrates OA, OS, and FJ. The lengths of EcoRI restriction fragments are shown. Where the length differs for the two strands of the DNA substrate, the length of the longer fragment is given in parentheses. B and C, Protein Data Bank coordinates 1DCW, 1F7Y, and BDL084 were used to construct a three-dimensional model of the obstacle (stereo view). The exact nucleotide sequence differs from the sequence of the constructs used in the experiments. B, side view. Arrows indicate the positions in which the 6 bp of the EcoRI cleavage site are exposed in the major groove. Left, substrate OA; right, substrate OS. C, view along the helical axis of the substrate. The turquoise shape indicates how the enzyme may be bound, in analogy to the structure of the nonspecific BamHI-DNA complex (4). One subunit of the homodimeric enzyme is shown as solid shape, the position of the other subunit is indicated by a dotted line. D, PhosphorImager record of a native gel. The relative mobilities of substrates NO, OA, OS, and FJ support their tertiary structure, which is indicated as a diagram to the left of the gel.
argument, they neglect, however, that the protein may also face considerable energy barriers when moving between two major groove sites, because it will transiently have to leave the optimal orientation with respect to the backbone phosphates. Furthermore, helical movement of proteins around DNA has been shown to be unfavorable for hydrodynamic reasons (12). For an accurate prediction of the behavior of proteins bound to nonspecific DNA, molecular dynamics simulations taking into account solvent water molecules will be necessary.

This study presents for the first time experiments that can distinguish between 1D and 2D sliding. Processivity of EcoRI cleavage was investigated for DNA substrates that contain obstacles between two cleavage sites. The obstacle was located in the major groove on one face of the DNA, so that proteins using 1D sliding would not be able to overcome it. The obstacle, however, should not affect proteins using higher-dimensional modes of diffusion. At both low and moderate ionic strength, facilitated diffusion of EcoRI between the two cleavage sites was as efficient as in the absence of an obstacle. This result indicates that 1D sliding does not play a significant role in protein motion over distances as short as 21 bp.

If 2D sliding is involved in facilitated diffusion, it is an interesting question whether the protein prefers pathways along one face of the DNA, or whether a genuine 2D random walk applies. If pathways along one face of the DNA were preferred, processivity on substrate OA could be expected to be higher than on substrate OS. Because processivities on OA and OS are not significantly different from each other, it is concluded that pathways along one face of the DNA are not preferred.
significantly different under both ionic strength conditions, diffusion along one face of the DNA can be ruled out.

**2D Sliding versus 3D Diffusion**—Transfer of EcoRV between distant DNA sites has been shown to involve 3D diffusion (9). RNA polymerase has been shown to diffuse along the contour length of the DNA (13); this finding is compatible with sliding rather than 3D diffusion. It is likely that different proteins have evolved idiosyncratic mechanisms of target location. Furthermore, proteins may use a combination of different diffusion modes: 2D sliding over short distances and 3D diffusion over long distances.

The data presented here support the occurrence of 2D sliding at low ionic strength. The distance between the two cleavage sites in substrate NO is greater than in substrate FJ, yet the processivity is higher on NO at 25 mM NaCl. The most probable explanation for this observation is that the highly efficient transfer between the two sites on NO relies on 2D sliding. Because the two sites on FJ do not lie on the same double helix, 2D sliding cannot be used for the transfer and the protein has to rely on less efficient 3D diffusion.

However, 3D hopping over very short distances along the DNA may also account for the observed effect. Furthermore, the physiological relevance of phenomena observed under low ionic strength conditions is arguable.

**Obstacles and Efficiency of Target Location**—Almost all studies into facilitated diffusion of proteins on DNA were carried out in vitro, using naked DNA (2). In vivo, DNA is bound to numerous proteins: DNA is estimated to account for only one third of the mass of the macromolecular content of eukaryotic nuclei; the other two thirds are equally split between histones and other proteins (14). The enhancement of target location through sliding can be conceptualized as an extension of the effective target site (3). Bound proteins impose a rather low upper limit on the enhancement that can be achieved by 1D sliding because they reduce the maximal effective target size to the length of the intervening naked DNA. 1D sliding is not an efficient mode of scanning protein-bound DNA, as illustrated by the Monte-Carlo simulation. Because a protein sliding along the major groove cannot bypass obstacles bound to the DNA, it will sample the DNA stretch between two obstacles repeatedly in a highly redundant random walk. If the assumptions of the simulation are reasonable, 2D sliding is a more efficient way of target location on both naked and nucleosomal DNA. The most interesting result of the simulation is that site scanning by 2D sliding is far more efficient on nucleosomal DNA than on naked DNA. This result is likely to reflect the real situation, because it does not depend on the exact parameters that describe the 2D random walk. The expediting effect of obstacles on target location should be even more pronounced in reality, in that obstacles such as nucleosomes can be expected to restrict the available surface for 2D sliding more dramatically than assumed in the simulation. Given that DNA sequences of regulatory importance are often found in nucleosome-free regions (15), nucleosomes might function as “highways” that rapidly channel diffusing proteins such as transcription factors between nucleosome-free regions, which are then scanned more carefully.

As illustrated in Fig. 5, a combination of 1D sliding and 3D diffusion is likely to be less efficient than 2D sliding in vivo. Because 1D random walks are limited to short regions of naked DNA, intervening 3D diffusion of the protein is required to explore other DNA segments. The high local concentration of DNA in a chromatin context, however, will increase the probability that the protein reassociates with a DNA segment in the immediate vicinity of the segment it dissociated from. In other words, the probability of escaping from the chromatin domain is low. As a result, the protein will sample a limited 3D chro-

**REFERENCES**

1. von Hippel, P. H., and Berg, O. G. (1989) J. Biol. Chem. 264, 675–678
2. Shimamoto, N. (1999) J. Biol. Chem. 274, 15293–15296
3. Berg, O. G., Winter, R. B., and von Hippel, P. H. (1981) Biochemistry 20, 6929–6948
4. Viadiu, H., and Aggarwal, A. K. (2000) Mol. Cell 5, 889–895
5. Sun, J., Viadiu, H., Aggarwal, A. K., and Weinstein, H. (2003) Biophys. J. 84, 3317–3325
6. Sambrook, J., and Russell, D. W. (2001) Molecular cloning: A Laboratory Manual, 3rd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
7. Duckett, D. R., Murchie, A. I., Diekmann, S., von Kitzing, E., Kemper, B., and Lilley, D. M. (1988) Cell 55, 79–89
8. Terry, B. J., Jack, W. E., and Morrich, P. (1985) J. Biol. Chem. 260, 13130–13137
9. Stanford, N. P., Szczelkun, M. D., Marko, J. F., and Halford, S. E. (2000) EMBO J. 19, 6546–6557
10. Anto, V. P., and Tinoce, I. J. (1992) Nucleic Acids Res. 20, 819–824
11. Eichman, B. F., Vargason, J. M., Moers, B. H., and Ho, P. S. (2000) Proc. Natl. Acad. Sci. U. S. A. 97, 3971–3976
12. Schurr, J. M. (1979) Biophys. Chem. 9, 413–414
13. Guthold, M., Zhu, X., Rivetti, C., Yang, G., Thomas, N. H., Kasas, S., Hansma, H. G., Smith, B., Hansma, P. K., and Bustamante, C. (1999) Biophys. J. 77, 2284–2294
14. Cook, P. R. (2001) Principles of Nuclear Structure and Function, Wiley, New York
15. Wolfe, A. P. (1994) Trends Biochem. Sci. 19, 240–244
16. Polyak, G. (1921) Math. Ann. 84, 149–160
Obstacle Bypass in Protein Motion along DNA by Two-dimensional Rather than
One-dimensional Sliding
Martin Kampmann

J. Biol. Chem. 2004, 279:38715-38720.
doi: 10.1074/jbc.M404504200 originally published online July 2, 2004

Access the most updated version of this article at doi: 10.1074/jbc.M404504200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2004/07/23/M404504200.DC1

This article cites 14 references, 5 of which can be accessed free at
http://www.jbc.org/content/279/37/38715.full.html#ref-list-1