Eur Biophys J (2002) 31: 257–267
DOI 10.1007/s00249-002-0224-4

ARTICLE

Stephen E. Halford · Mark D. Szczelkun

How to get from A to B: strategies for analysing protein motion on DNA

Received: 12 October 2001 / Accepted: 21 February 2002 / Published online: 30 May 2002

Abstract Essentially all genetic events require proteins to move from one location in a DNA polymer to another location in the same chain. A protein will seldom bind to a specific site in the DNA by colliding directly with that site. Instead, the protein will almost always collide first with a random site anywhere in the DNA and then migrate to the specific site by a facilitated-diffusion process that is constrained to the zone of that DNA molecule. Thereafter, many proteins bound to their target sites translocate in a specified direction along the DNA by a energy-dependent vectorial mechanism. This review will discuss some of the strategies that have been developed to analyse the motion of proteins on DNA, with respect to both the random diffusion processes involved in target-site location by DNA-binding proteins and the vectorial processes involved in unidirectional translocation along DNA.

Keywords Diffusion · DNA · Restriction enzyme · Translocation

Abbreviations *AFM*: atomic force microscope · *bp*: base pair(s) · *kb*: 1000 bp · *TFO*: triplex-forming oligonucleotide

Introduction

A protein that acts at a specific sequence of bases in DNA first has to locate its target sequence from amongst a very large number of alternative sequences that are present in the same DNA molecule. As a consequence of random diffusion through solution (Gutfreund 1995), the protein will occasionally collide with the DNA molecule. In the "diffusion limit", when all of the collisions possess sufficient activation energy to permit complex formation, such collisions will always result in the binding of the protein to the DNA. However, if the target sequence is 6 bp long, as is the case with many restriction endonucleases (Roberts and Halford 1993), and the DNA a plasmid of 6 kb, it is most unlikely that the initial encounter will be at the target sequence. Instead, it will almost certainly be at a non-specific site elsewhere in the DNA. Starting from free solution, the protein would therefore need to collide with DNA many times over before encountering the target site. Yet proteins can bind to specific sites along DNA at extraordinarily rapid rates, at or even seemingly above the diffusion limit (Riggs et al. 1970). DNA-binding proteins are thus thought to locate their target sites in DNA by facilitated diffusion, in which the protein first binds to a random site anywhere along the DNA and then translocates to its specific site by an intramolecular process, without departing from the domain of solution occupied by that DNA molecule (von Hippel and Berg 1989).

Once bound to their target sites, many enzymes catalyse reactions on DNA that involve the progressive motion of the protein along the DNA in a specified direction (or vice versa, the motion of the DNA past the protein), in steps of one or more base pairs at a time. Enzymes that undergo this type of vectorial motion include the DNA and RNA polymerases, that move from a specified start-point in the DNA (an origin of replication or a promoter, respectively) to a specified end-point (Kornberg and Baker 1992; Uptain et al. 1997), and the DNA helicases that progressively unwind the DNA duplex prior to replication, repair or many other DNA transactions (Lohman and Bjornson 1996).

Amongst the most amenable systems for analysing both sorts of motion are the type I and II restriction

Dedicated to Professor H. Gutfreund on the occasion of his 80th birthday

S.E. Halford · M.D. Szczelkun (✉)
Department of Biochemistry,
School of Medical Sciences,
University of Bristol, University Walk,
Bristol BS8 1TD, UK
E-mail: mark.szczelkun@bristol.ac.uk
Tel.: +44-117-9287439
Fax: +44-117-9288274

endonucleases. The type II enzymes recognize specific DNA sequences (4–8 bp long) and cleave both strands at fixed loci in or adjacent to the site, using Mg^{2+} or a similar divalent cation as a cofactor (Roberts and Halford 1993). The type I enzymes also recognize specific DNA sequences (13–17 bp long), but cleave the DNA at random loci that can be many kilobase pairs away from the recognition site, in a reaction that requires both Mg^{2+} ions and ATP hydrolysis (Murray 2000). This review will focus on some of the strategies that have been developed to analyse protein motion on DNA: firstly, with respect to the random diffusion processes that lead to the target-site location by type II endonucleases (Stanford et al. 2000); and secondly, with regard to the vectorial motion of type I endonucleases tracking along the DNA in a fixed direction. (Szczelkun 2000). Nevertheless, the underlying principles for these long-range communications on DNA apply to virtually all DNA-binding proteins.

Schemes for facilitated diffusion

A number of different schemes have been proposed to account for the intermolecular transfer of a protein from non-specific to specific sites within the same molecule of DNA (von Hippel and Berg 1989; Shimamoto 1999). In a widely accepted model known as "sliding", the protein diffuses linearly between consecutive non-specific sites in the DNA without dissociating from the DNA (Fig. 1a). Each time the protein executes a step along DNA, the probabilities of moving "leftward" or "rightward" are the same. Hence, the mean location of the protein during linear diffusion always remains its initial position but the length of the excursions increase as the total number of steps increases (Berg 1993). Thus, given enough time, the protein can reach its target-site.

The transfer could alternatively occur through 3-D space, via multiple rounds of dissociation and re-association (Fig. 1b). In dilute solution, each DNA molecule occupies a discrete volume – a domain – that contains a high concentration of DNA segments separated by larger volumes that lack DNA (Winter et al. 1981). The mean distances between DNA molecules are much larger those between different segments of the same molecule (Bellomy and Record 1989), particularly in the case of supercoiled DNA (Vologodskii and Cozzarelli 1994). Consequently, following its dissociation from the DNA, a protein is much more likely to re-associate with the same molecule of DNA than to a different one. However, a search through 3-D space is likely to be slower, at least under certain circumstances (Berg 1993), than one restrained to 1-D events such as sliding (Richter and Eigen 1974).

In a further scheme called "intersegmental transfer" (Fig. 1c), a protein bound to one site in the DNA binds concurrently to another site to give a complex in which it is transiently bound to both sites; by then releasing one of the two sites, the protein may either remain at its original site or transfer to the new site (Milsom et al. 2001). This scheme only applies to those proteins with two DNA-binding sites, e.g. the Lac repressor or the *Sfi*I and *Ngo*MIV endonucleases (Halford et al. 2000). In addition, the formation of a complex in which a protein is bound concurrently to two DNA sites first requires the juxtaposition of the sites in 3-D space to within the requisite reaction radius. The probability of juxtaposition of two *cis* DNA sites varies with the separation of the sites, increasing steeply as the intervening length is increased to ~300 bp and then declining gradually for lengths > 300 bp (Bellomy and Record 1989; Ringrose et al. 1999; Podtelezhnikov and Vologodskii 2000). Intersegmental transfers are therefore likely to translocate a protein by ≥300 bp at a time, and are thus only suitable for large steps on long DNA molecules by a particular subset of DNA-binding proteins.

The above schemes can be classified by whether they result in positionally correlated or uncorrelated transfers, i.e. whether the protein at time $t+\Delta t$ is near its position at time t (Berg et al. 1981). Sliding is a correlated process while intersegmental transfer is uncorrelated, but transfers by dissociation/re-association can be

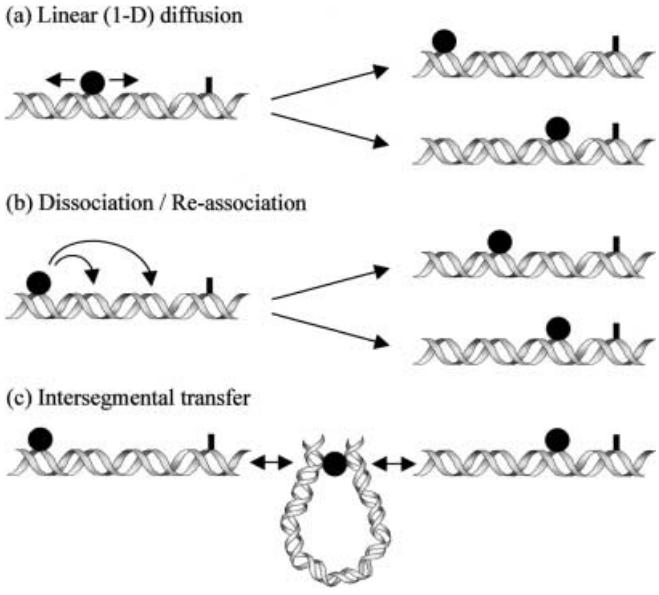

Fig. 1a–c. Schemes for facilitated diffusion. In all three schemes the specific target-site is indicated by a *hatch mark* on the DNA, and the protein by a *black circle*. (**a**) Linear diffusion. The protein diffuses linearly along the DNA without dissociating. The 1-D motion has no fixed directionality so each movement of the protein has an equal probability of taking it towards or away from the target site (*lower* and *upper schemes*, respectively). (**b**) Dissociation/re-association. Dissociation of the protein from the DNA is followed by its re-association to the same molecule, to a site either close to or distant from the initial site along the DNA contour. These are called hopping (*upper scheme*) or jumping (*lower scheme*), respectively. (**c**) Intersegmental transfer. The protein bound to one site binds concurrently to a second site in the same DNA, sequestering the intervening DNA in a loop. The protein can then release the second site, leaving itself at the same position as before, or it can release the initial site and so transfer itself to the second site

either correlated or uncorrelated. The dissociation of the protein from the DNA will usually be followed by its re-association at or near the original site, since two macromolecules in close vicinity of each other tend to remain so for a relatively long period of time before they diffuse apart (Berg 1978). Consequently, most transfers by dissociation/re-association will be correlated events and these are known as "hopping" (Fig. 1b, upper scheme). However, on occasions, the protein may diffuse some distance away from its initial site yet still remain within the domain of the DNA molecule. The re-association can then be a "jumping" event to an uncorrelated site elsewhere in the chain (Fig. 1b, lower scheme). Only an arbitrary distinction can be made between "hopping" and "jumping": say, movements of <20 bp for hopping and of >20 bp for jumping (Stanford et al. 2000). On linear or relaxed DNA, hopping will be more frequent than jumping but this will not necessarily be true on supercoiled DNA. In the latter case, segments that are far apart along the 1-D contour may be interwound around each other and thus close together in 3-D space (Vologodskii and Cozzarelli 1994), and a protein that dissociates from one site could easily re-associate with a distant site directly across the superhelical axis.

Strategies for facilitated diffusion

Perhaps the most widely used strategy for studying the mechanism of target-site location uses a series of DNA molecules that each have one copy of the target-site for the protein but which differ from each other in their overall lengths (Fig. 2a). The rates (k_{bind}) and/or the equilibria (K_B) for protein association with the site are then analysed as a function of the DNA length (Berg et al. 1981). Many DNA-binding proteins have been examined by this procedure (Shimamoto 1999), including several type II restriction enzymes (Jack et al. 1982; Ehbrecht et al. 1985; Jeltsch et al. 1996; Jeltsch and Pingoud 1998). This strategy can reveal directly whether or not the protein finds its target by a facilitated diffusion mechanism, i.e. whether the non-specific sequences in the DNA molecule are on the pathway to the specific site. Unfortunately, essentially all schemes for facilitated diffusion allow for decreases in k_{bind} as the DNA length is decreased (Fig. 2a), simply because a shorter DNA provides a smaller target for the initial encounter with the protein. Consequently, it can be extremely difficult to use this approach to distinguish between the various schemes for facilitated diffusion noted above. Indeed, in most applications of this procedure (Shimamoto 1999) the data are reconciled to sliding without demonstrating that the data are inconsistent with other pathways.

For example, in one study (Jeltsch et al. 1996), the reaction rates of a number of mutants of the *Eco*RV restriction endonuclease were measured in vitro on two substrates: a long (958 bp) and a short (26 bp) DNA molecule, each with a single *Eco*RV site. The ratio of the cleavage rates on the two substrates was found to vary in proportion to the efficiency of these mutants in restricting foreign DNA in vivo, from which it was concluded that *Eco*RV locates its target site in vivo by sliding. However, whilst these data suggest that *Eco*RV locates its target site by *facilitated diffusion*, the various 1-D and 3-D mechanisms cannot be distinguished from each other on the basis of length dependence alone. To make such a distinction, it is necessary to vary not only the length of the DNA molecule but also the equilibrium constant for the binding of the protein to non-specific DNA (Berg et al 1981). This can be modulated by altering the concentration(s) of the salts in the reaction

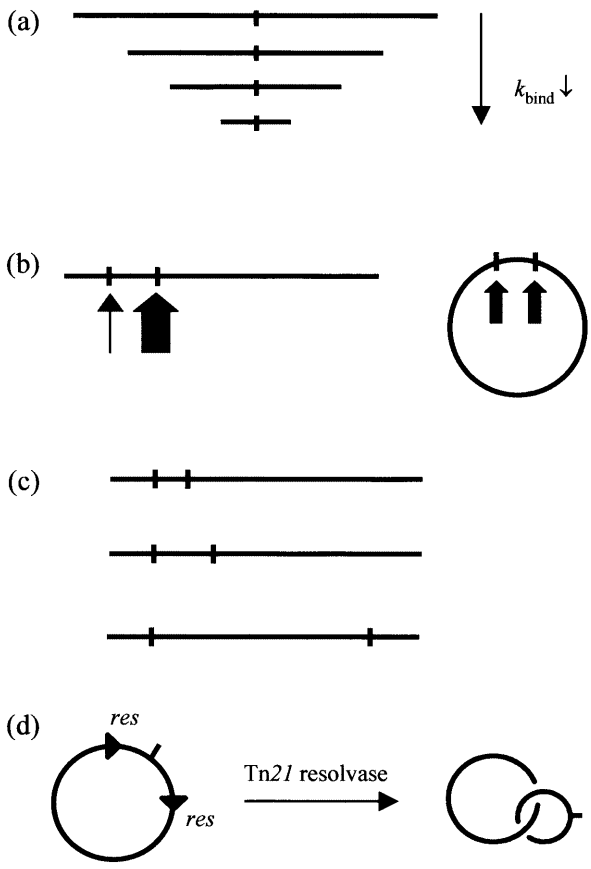

Fig. 2a–d. Strategies for facilitated diffusion. (**a**) Variations in the overall length of the DNA. The kinetics (k_{bind}) and/or equilibria (not shown) for the binding of the protein to its target site are studied with a series of DNA molecules of different lengths, that each has one copy of the target sequence. (**b**) Variations in the positions of two target sites. Each DNA molecule has two target sites separated by a fixed length of DNA: one has both sites close to one end of the DNA and the other is the circular form of the same DNA. The likely preference of the enzyme for one site over the other is indicated by the *thickness of the arrows*. (**c**) Variations in the length of DNA between two target sites. The degree of processivity of an enzyme on a DNA with two target sites is measured with a series of molecules that each carries two target sites separated by varied lengths of DNA. (**d**) Variations in DNA topology. Site-specific recombination by resolvase will convert a plasmid with two *res* sites (*filled triangles*) into a catenane containing two interlinked circles of covalently closed DNA; if the plasmid has a single target site for the test enzyme in the short arc between the *res* sites, the catenane will carry this site in the small ring

buffer (Lohman 1986). The various schemes for facilitated diffusion can then be distinguished by analysing k_{bind} and K_B with each of the DNA molecules, in each case at a series of different salt concentrations (Winter et al. 1981).

An alternative approach uses substrates with two or more target sites separated by a fixed length of DNA (Fig. 2b); the preference of the enzyme for one site over another, depending on its position in the chain, is then examined at a range of different salt concentrations (Terry et al. 1985; Jeltsch et al. 1994). Additional information can also be obtained by determining whether or not the enzyme acts processively or distributively on a substrate with two sites, i.e. in the case of a type II restriction enzyme, the extents to which the endonuclease cleaves both sites during a single DNA-binding event or each site in a distinct binding event.

An elegant example of this strategy was provided by studies of the *Eco*RI restriction enzyme on DNA molecules of 388 bp with two target-sites 51 bp apart (Terry et al. 1985). In one case, the two sites were located closer to the "left-hand" end of a linear DNA than the "right-hand" end (Fig. 2b). The *Eco*RI enzyme acted on this DNA in a partly processive manner but it made its initial cleavage more frequently at the site nearer the middle of the DNA than at the site near the "left" end. In contrast, on the circular form of this DNA (Fig. 2b), or on an alternative linear form in which the two sites were more or less equidistant from the ends (not shown), the endonuclease still acted processively but showed no preference for one site over the other. The *Eco*RI enzyme thus binds first to DNA at random and then locates the nearest target site by a correlated pathway before proceeding to the second site in the same molecule of DNA. On the DNA with two sites close to the left-hand end (Fig. 2b), the initial binding is most likely to occur somewhere in the 319 bp segment to the "right" of the innermost site, so that the enzyme finds the innermost site first, in preference to the outermost site. Strikingly, the degree of processivity on the circular DNA (Fig. 2b) was almost double that on the linear forms. This indicates that the translocation of the protein from one site to the other occurs with almost equal facility over either the short (51 bp) or the long (337 bp) arc of the circular DNA. If the correlated pathway was sliding, translocation across 51 bp should have been more efficient that that across 337 bp (Stanford et al. 2000), and the degree of processivity on the circular DNA should thus have been only marginally above that on the linear DNA.

A further strategy (Stanford et al. 2000) integrates the two approaches noted above by using DNA substrates with two sites (viz. Fig. 2b), but where instead of varying the overall length of the DNA molecule (viz. Fig. 2a), the length of DNA between the two sites is varied (Fig. 2c). A processivity factor (number of reactions in which both sites are cleaved during a single binding event/total number of reactions) is then measured for each substrate. This strategy (Fig. 2c) has a major advantage over that involving variations in the overall length of a DNA with one site (Fig. 2a). On a DNA with one site, the translocation from the initial random site to the final specific site occurs over an indeterminate distance; on some occasions, the protein will land near the target site, but on others, far away. In contrast, during processive action at two DNA sites, the enzyme translocates over a known distance, namely the separation of the sites (n bp). Consequently, the schemes for facilitated diffusion give rise to much simpler relationships between the extent of processivity and the intersite spacing (Stanford et al. 2000) than those between the association rate constant and the overall length of the DNA (Berg et al. 1981). As a first approximation, the processivity factor should decline with increasing site separation as a function of n^2 for 1-D transfer along the DNA contour, and as a function of $n^{1/2}$ for a 3-D pathway. The n^2 term for the former comes from the statistics of a random walk along a linear lattice (Berg 1993), while the $n^{1/2}$ term for the latter reflects the mean distance in 3-D space between two loci in a random coil polymer (Doi and Edwards 1984).

On a series of DNA molecules with two *Eco*RV sites separated by 54–764 bp, the degree of processivity displayed by the *Eco*RV endonuclease declined with increasing inter-site spacing in a manner that deviated by several orders of magnitude from that required for a 1-D sliding process (Stanford et al. 2000). Instead, the relationship matched, to within a factor of two, that expected for transfers through 3-D space by hopping and jumping.

Another unambiguous method for distinguishing between mechanisms of facilitated diffusion uses a DNA catenane substrate containing two interlinked rings of covalently closed DNA (Fig. 2d). Such catenanes are readily constructed in vitro by the action of a site-specific recombination enzyme on a plasmid with the appropriate targets for the recombinase: *att*P and *att*B for the integrase from phage λ or two directly repeated *res* sites for a resolvase from a Tn*3*-like transposon (Hallet and Sherratt 1997). For communications through 3-D space, a catenane carrying one site in each ring will be almost as effective a substrate as the parental plasmid carrying both sites in a single ring (Craigie and Mizuuchi 1986), but such a catenane will be ineffective as a substrate for communications that occur along the 1-D DNA contour (Szczelkun et al. 1996). To date, this strategy has been used primarily to analyse communications between two specific sites (Craigie and Mizuuchi 1986; Szczelkun and Halford 1996). However, if the catenane carries one specific site in a small ring of DNA and no sites in the larger ring, it is then possible to examine whether the non-specific sequences in the large ring help the protein find its specific site in the small ring, even though there is no 1-D pathway from one ring to the other. In preliminary experiments with the *Eco*RV restriction enzyme, a small circle of DNA carrying an *Eco*RV site was indeed cleaved more readily when it was tethered by catenation to a large ring of non-specific

DNA than when it was separated from the non-specific sequences (D.M. Gowers and S.E. Halford, in preparation). Thus, the *Eco*RV endonuclease must transfer primarily from non-specific to specific DNA through 3-D space.

It has always been hard to reconcile DNA sliding with protein motion in a cell where, in the absence of an input of chemical energy, the thermal energy that drives diffusion will be insufficient to displace the numerous protein "roadblocks" that would otherwise impede 1-D motion. Instead, it has been demonstrated that target-site location in the mammalian nucleus is via 3-D jumping (Misteli 2001). These conclusions have come from observations of protein mobility monitored in vivo by fluorescence recovery after photobleaching (FRAP). The measured 3-D protein flux is remarkably fast, given the spatial constraints of the nucleus, but is nonetheless slower than expected on the basis of molecular weight alone (Misteli 2001). Instead, effective mobility is determined by interactions with DNA and other nuclear components; for example, the linker histone H1 exchanges rapidly on the order of seconds whilst the core histone takes many hours (Lever et al. 2000; Misteli et al. 2000). If DNA-binding activity is removed, then diffusion becomes 10–200 times faster. This suggests that 3-D diffusion through the nucleus is potentiated by transient interactions with non-specific DNA – this is equivalent to the hopping and jumping events observed in vitro for *Eco*RV (Stanford et al. 2000; D.M. Gowers and S.E. Halford, in preparation).

Schemes for vectorial motion

Once bound to a specific site, many proteins then undertake a series of directed 1-D steps along DNA. At first glance, these translocation events appear similar to DNA sliding except that motion is directional and thus requires fewer steps to reach a target site. However, although sliding is driven randomly by thermal energy from collisions with solvent molecules, translocation must overcome an entropy barrier associated with its directionality and is powered by free energy liberated from a high-energy phosphodiester. This can come from nucleotide polymerization (e.g., transcription and replication), breakage of the nucleic acid polymer backbone (e.g., exonuclease III) or from removal of the γ-phosphate from a nucleoside triphosphate (e.g., helicases and type I restriction enzymes). The coupling of this chemical energy to the mechanical energy of motion can vary: in some cases, such as the polymerases, it is necessarily tight – one nucleotide is added to a nascent chain at every step; in other cases the coupling is more loose and several ATP molecules may be utilized to undertake a single step.

In many examples, the initiation of translocation will require the "melting" of the DNA duplex to reveal single strands of DNA. The protein then contacts and moves actively along one strand with a defined polarity (3′-5′ or 5′-3′); passive transport of the opposite strand past the protein may still be important to stability of the complex and thus its processivity (Zhou et al. 1995). In other examples, such as the RuvB or RecG helicases (George et al. 2000; Singleton et al. 2001) and possibly the type I restriction enzymes (see below), the duplex remains annealed, though probably only one of the two strands is actively transported. In either case, after initiation the protein will continue translocating until a termination signal is reached, or it collides with another translocase or a stationary roadblock, or it releases the DNA track. This last event is dependent upon the relative values of two rate constants: k_{step}, controlling the rate for moving one "step" forward, and k_{off}, the rate of dissociating from the lattice.

The probability of making a step rather than dissociating can be measured in terms of the processivity of translocation. This can be defined as $P_T = k_{step}/(k_{step} + k_{off})$. If $P_T = 1$, then the protein will never dissociate and will continue translocating until it exhausts its energy supply or reaches a termination signal. However, even if $k_{step} >> k_{off}$, then the chance of moving long distance can be seriously reduced. For instance, if k_{step} were 1000-fold faster than k_{off}, the probability of making 1000 steps would be $(P_T)^{1000} = 0.37$, i.e. just over one third of the molecules make 1000 steps despite a relatively large difference in the rate constants. For processive enzymes such as the RecBCD helicase or the type I enzymes that can translocate over tens of thousands of base pairs (Roman et al. 1992; Bianco et al. 2001; Szczelkun 2000), k_{step} must be at least 30,000-fold faster than k_{off}. Both rate constants can be affected by the local sequence being translocated. For the polymerases which have a complex kinetic relationship between dwell times and template sequence, pausing, stalling and even back steps are common (Uptain et al. 1997). However, the kinetics of translocation by helicases (Ali and Lohman 1997; Jankowsky et al. 2000) and exonuclease III (Linxweiler and Horz 1982) clearly suggest translocation rates that are independent of sequence. It is not known how local sequence may affect facilitated diffusion: sliding would be more efficient if the off rate is as small as possible (Berg 1993; Stanford et al. 2000), whereas hopping/jumping rely on an appreciable off rate for motion.

The kinetic constant k_{step} can relate to one of several rate-limiting physical events: for example, in strand-separation mechanisms, a distance physically *moved* per step or an amount of DNA *unwound* per step. Whereas sliding most likely occurs in single base pair increments and "step sizes" of 1 bp are fundamental to the activity of translocases such as polymerases, the value is not unity in all cases. Indeed, although step sizes of 1 bp have been estimated for some helicases (Dillingham et al. 2000), larger steps sizes of 3–5 bp (Ali and Lohman 1997; Jankowsky et al. 2000) and even 23 bp (Bianco and Kowalczykowski 2000) have also been recorded. It is not yet clear exactly what these values reflect in terms of mechanism.

During motion, the majority of translocases will follow the helical path of DNA. Two topological problems can ensue: torsional stress (Liu and Wang 1987) and/or the entwining of a nascent nucleic acid polymer around the track (Cook 1999). If one imagines the DNA as a stationary rod, then a translocating protein will rotate in a right-handed direction around the helical axis (Fig. 3a, upper scheme). Alternatively, if the protein was somehow anchored to an immobile matrix or was unable to rotate due to its "bulk", then the DNA must be pulled past the stationary protein and will rotate about its axis (Fig. 3a, lower scheme). This can then lead to changes in the DNA twist ahead and behind the complex (Liu and Wang 1987). On linear DNA these effects will diffuse away due to rotation at the free ends, whilst on circular DNA the equal and opposite changes will cancel each other out by diffusion around the ring. However, if the DNA is organized into separate and discrete topological domains, then the changes in twist will be trapped and result in changes in supercoiling (Liu and Wang 1987). This has been clearly demonstrated using a fusion protein of the GAL4 repressor and T7 RNA polymerase (Ostrander et al. 1990). On a topologically relaxed plasmid carrying a GAL4 operator and a T7 promoter, the chimeric protein binds to form a figure-of-eight structure as shown in Fig. 3b. Upon addition of ribonucleotides, the polymerase starts to transcribe and, in doing so, pulls DNA from one loop into the other. Because of anchoring via the GAL4 protein, the changes in twist are separated into the loops such that the contracting loop ahead of the polymerase becomes positively supercoiled whilst the expanding loop behind becomes negatively supercoiled (Fig. 3b). Thus, the topology of a substrate can affect both diffusion, favouring one facilitated scheme over another, and 1-D translocation, modulating the rate of motion and the outcome of DNA processing (Szczelkun et al. 1996).

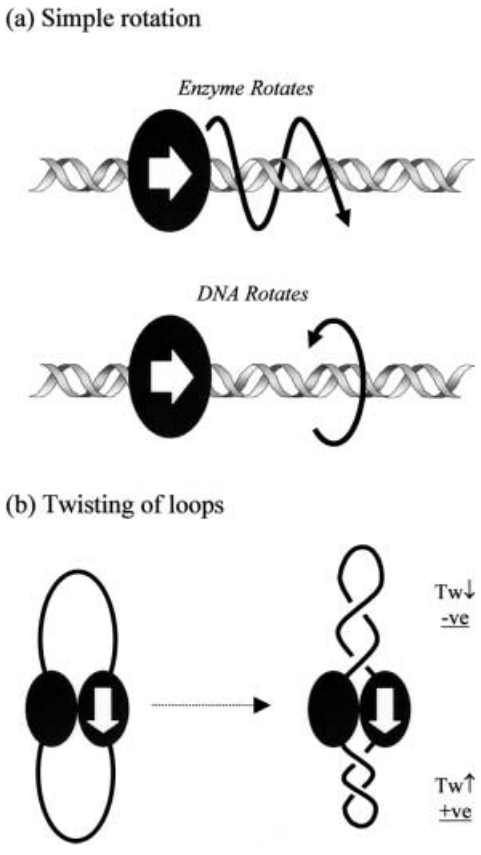

Fig. 3a, b. The affect of protein motion upon DNA topology. In each case, the protein is represented as a *filled oval*. The *white arrow* indicates motion; *no arrow* indicates the protein is fixed. (**a**) Simple rotation. During 1-D motion on DNA, a protein will follow the helical path of the track. If the protein is freely mobile, it will rotate around the DNA axis (*upper scheme*). If the protein is fixed, the DNA will rotate (*lower scheme*). (**b**) Twisting of distinct topological domains. A bi-functional protein binds to two sites on a relaxed DNA (*black line*), forming a figure-of-eight structure. 1-D motion of one of the subunits (*dotted line*) results in DNA from the lower contracting loop passing into the upper expanding loop. Twist (*Tw*) is increased in the contracting loop and reduced in the expanding loop, producing positive (+ve) and negative (–ve) supercoiling, respectively. These changes are not permanent as no strands have been broken and rejoined (if the protein were removed, the DNA would still be relaxed)

Strategies for vectorial motion

There are three basic enzyme activities associated with DNA cleavage by the type I restriction enzymes: an endonuclease, an ATPase and a translocase. The first two activities can be easily measured and have been well characterized (Murray 2000; Szczelkun 2000), but the third activity is less easily assessed. Although it was appreciated more than 27 years ago that these enzymes must move along DNA (Shulman 1974), much of the early work was unable to define the mechanism reliably. One problem is the difficulty in determining directly translocation velocity from ATPase rates without knowledge of the degree of coupling [there are certain exceptions: for example, the distinct ATPase rates of the PcrA helicase can be related to different translocation states (Dillingham et al. 2000)]. Moreover, the ATPase activity of the type I enzymes continues unabated even after DNA cleavage has occurred (Murray 2000), so distinguishing where translocation-dependent ATP hydrolysis stops and some other activity starts is impossible with current methods. Similarly, it is not trivial to apportion the rate and location of the random DNA cleavage sites to the preceding translocation events. For instance, during cleavage of linear DNA – which generally requires the convergence and collision of a pair of translocating enzymes (Studier and Bandyopadhyay 1988; Szczelkun et al. 1996) – the enzymes can dissociate from the DNA both during motion and upon collision before cleavage occurs (Szczelkun 2002). In this case, an independent measure of DNA translocation is needed to explain the observed patterns.

An unequivocal proof of a 1-D communication event between site-specific recognition and distant non-specific cleavage came from assays using DNA catenanes

(Szczelkun et al. 1996). Cleavage of a one-site plasmid and its equivalent singly interlinked catenane was compared (Fig. 2d). On the plasmid, cleavage occurred throughout the DNA whilst cleavage of the catenane occurred only in the ring carrying the type I recognition site. The connected ring remained intact despite its physical proximity. This discounted a looping event, which should have been capable of communicating with the connected ring through 3-D space (see above).

The type I restriction enzymes carry motifs found in superfamily 2 helicases (Gorbalenya and Koonin 1991). The basic assay of true helicase activity involves measuring the time taken to unwind a series of duplex DNA substrates of different length into their corresponding free single strands (Lohman and Bjornson 1996). These assays can reveal a great deal of kinetic information, including translocation rates, steps sizes and processivity (Ali and Lohman 1997; Jankowsky et al. 2000). However, evidence for helicase-like strand separation by type I restriction enzymes has never been found (Szczelkun 2000), and it is possible that they move on intact duplex DNA as suggested for some other superfamily 2 members such as RecG (Singleton et al. 2001). Thus, the standard assays cannot be used with these enzymes.

An alternative approach to monitoring translocation is to monitor the kinetics of collision between a translocating enzyme and a static roadblock. For the type I enzymes this was achieved by following protein-directed displacement of the third strand of a DNA triplex (Fig. 4a). Linear DNA substrates were generated that carried target sites for the type I enzyme *Eco*R124I and for a triplex-forming oligonucleotide (TFO), with varied lengths of DNA between the two sites (Firman and Szczelkun 2000). The reaction conditions were set such that the triplex was stable (off rate <1 day), but if it were displaced, either thermally or by a mobile protein, the TFO never re-bound. A direct test of DNA translocation could then be achieved by measuring the lag time between the initiation of the reaction (by adding ATP to an enzyme-DNA complex) and the exponential phase in the increase in free TFO on each DNA substrate (Fig. 4b). A linear relationship between lag time and distance is indicative of a 1-D process and, for *Eco*R124I, gave a translocation velocity of ~400 bp/s (Firman and Szczelkun 2000).

As had been seen with DNA and RNA helicases (Ali and Lohman 1997; Jankowsky et al. 2000), a rigorous kinetic analysis of this sort can produce a great deal of information beyond basic rate measurements. For instance, the same kinetics were observed when the asymmetric *Eco*R124I site was orientated in reverse with respect to the triplex (Firman and Szczelkun 2000), showing that motion was bi-directional [occurring simultaneously in both "leftward" and "rightward" conditions as predicted by Studier and Bandyopadhyay (1988)]. Furthermore, the kinetic profiles are informative about the efficiency of motion. For *Eco*R124I, the amplitude and rate of the exponential phases that represent triplex displacement were invariant with distance (Firman and Szczelkun 2000). This indicates a highly

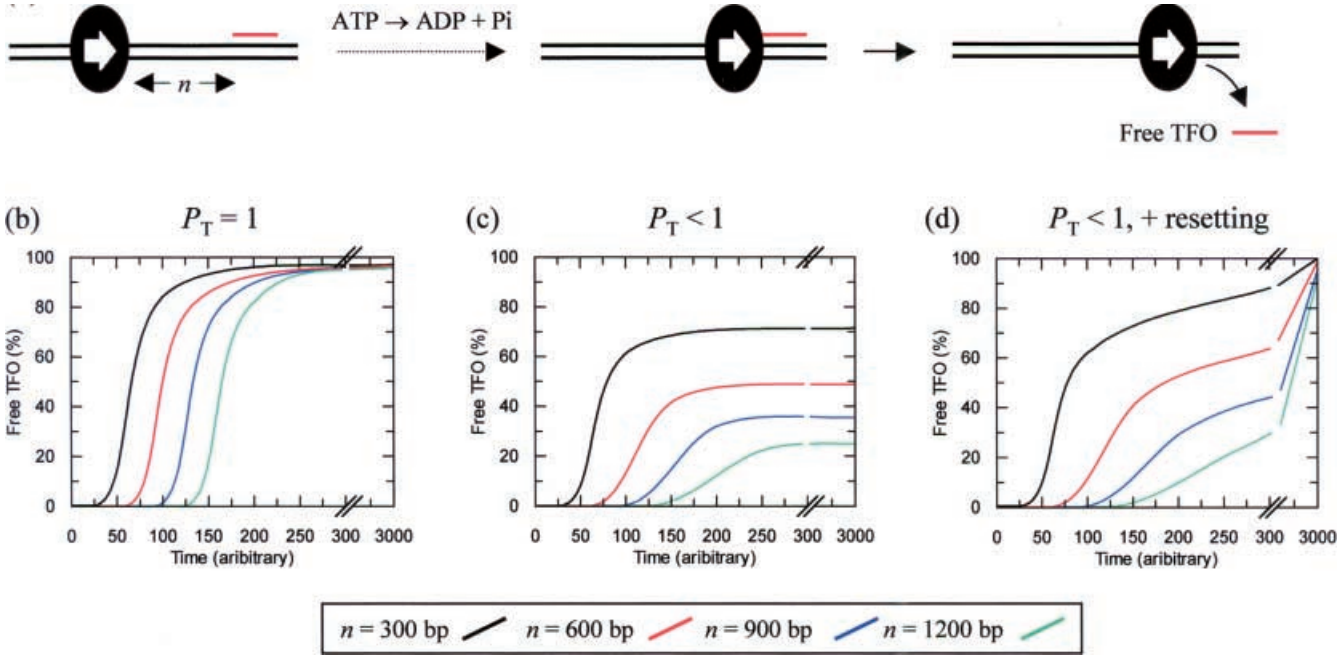

Fig. 4a–d. The kinetics of protein-directed "roadblock" displacement. (**a**) A triplex forming oligonucleotide (TFO, *red line*) is bound to duplex DNA (*two black lines*). The protein (*filled oval*) is bound to a specific site *n* bp from the triplex. The reaction is initiated with ATP and the protein moves towards the triplex, as indicated by the *white arrow*. Upon collision, the triplex is displaced to give free TFO, at a rate slower that the individual translocation steps. (**b–d**) Kinetics of triplex displacement. The percentage of free TFO displaced with time is indicated for $n = 300$, 600, 900 and 1200 bp, and for $P_T = 1$ or 0.9998 as indicated. For (**c**), dissociated enzymes cannot restart the reaction. For (**d**), dissociated enzymes can restart the reaction at the recognition site (resetting)

processive reaction ($P_T \rightarrow 1$): even after travelling >4300 bp, practically none of the proteins had dissociated, stalled or paused. If dissociation had occurred and re-initiation at the recognition site had been prevented ($P_T < 1$), then the resulting profiles would have been similar to those in Fig. 4c. In this situation, P_T can be estimated from the change in the amplitude of displacement with increasing distance (Fig. 4c). Translocation of *Eco*R124I could be made less processive by eliminating one of the two subunits involved in ATPase activity and protein motion (Firman and Szczelkun 2000). This complex has no DNA cleavage activity (Janscak et al. 1998), but was still capable of displacing a triplex. However, as the inter-site distance was increased, the resulting profiles became markedly non-exponential, as in Fig. 4d. These profiles are indicative of non-processive translocation in which a dissociated species can rebind at the recognition site and restart translocation in either direction. This "resetting" event produces a continuum of populated states where every enzyme eventually reaches the target site but does so increasingly slowly as the distance increases (Szczelkun 2002).

As noted above (Fig. 3b), protein motion on DNA can have radical affects on DNA topology. This is particularly pertinent to type I enzymes because of the architecture of the translocation complex (Murray 2000). Studies using both electron microscopy (reviewed in Murray 2000) and atomic force microscopy (AFM) (see below) have shown that the complex remains bound at its recognition site throughout the reaction, pulling non-specific DNA past the complex and thus forming two expanding DNA loops (Szczelkun 2000). On circular substrates this results in the DNA being partitioned into three distinct topological domains. If translocation follows the helical path of DNA, the twist (Tw) in the expanding loops would decrease whilst Tw in the contracting loop would increase (Szczelkun et al. 1996). Starting from relaxed DNA, this would lead to the production of negative and positive supercoils, respectively, as seen for the GAL4:T7 fusion described above (Fig. 3b). To test this proposal, Janscak and Bickle (2000) utilized a mutant of a type I endonuclease that could no longer cleave DNA but still had wild-type ATPase activity. After addition of ATP to initiate tracking, *Escherichia coli* topoisomerase I was added at set times and the topology of the DNA monitored. As translocation proceeded, the degree of positive supercoiling increased – the topoisomerase having presumably removed an equal number of negative supercoils from the DNA. Although these results confirmed the changes in topology expected from models of motion on circular DNA by type I enzymes, they also through up some challenging questions. For instance, if type I enzymes start translocating close to their recognition sites, how can large amounts of untwisting be accommodated in such small expanding loops? Precisely how the changes in twist are generated during motion needs further examination.

One of the greatest challenges in monitoring protein motion on DNA is to make measurements in vivo. An elegant strategy that addresses this question for type I enzymes has been described by Garcia and Molineux (1999). The translocation activity of *Eco*KI was assayed by following the rate of transfer of the T7 genome from phage particles into bacterial cells. A mutant of T7 was isolated whose DNA carried a single *Eco*KI site in the 850 bp leader sequence that is ejected into bacterial cells in the early stages of infection. In a strain that lacked an active *Eco*KI endonuclease, very little of the remaining 39 kb of DNA was able to enter the cells. However, in a strain with an active *Eco*KI enzyme, complete internalization was observed in ~9 min. Thus, translocation by the restriction enzyme was necessary to pull the DNA out of the phage head and into the bacterial cell. By following the progressive methylation of *Sau*3AI restriction sites, a translocation velocity of ~100 bp/s was estimated (Garcia and Molineux 1999). Mutations in the helicase motifs that impaired ATPase activity (Davies et al. 1998) and DNA restriction (Webb et al. 1996) also prevented entry of the phage DNA (Davies et al. 1999). Conversely, mutations in the endonuclease motif that abolished DNA cleavage did not affect ATPase activity and translocation. These results show that motion is independent of cleavage whilst cleavage absolutely requires a preceding translocation event. The helicase motifs are undoubtedly vital to the type I enzymes, but it remains to be seen exactly how these motifs convert the chemical energy of ATP into directed 1-D steps along duplex DNA.

Single molecule strategies

The classical methods for studying protein-DNA interactions involve bulk analysis of solutions where statistical averaging over the molecular population is required. Whilst reliable for simple schemes (Gutfreund 1995), the complexity of many protein-DNA systems alongside the structural anisotropy of DNA leads to problems in resolving mechanistic information from empirical data. However, recent technical developments have allowed manipulation of single molecules and complex processes have become more amenable to study (Bianco et al. 2001; Ishijima and Yanagida 2001). For example, transcription has been monitored directly by measuring the displacement and stall force of an RNA polymerase using an optical trap and polystyrene beads attached to the DNA (Yin et al. 1995; Wang et al. 1998). Not only can the forces involved in such interactions be measured, but external forces can also be exerted to investigate the fate of a reaction (Ishijima and Yanagida 2001). Furthermore, topographic images of protein motion on DNA can be recorded. Recently, an AFM has been used to observe DNA looping, translocation and cleavage by the type I restriction endonuclease *Eco*KI (Ellis et al. 1999; Berge et al. 2000). Upon addition of ATP to protein-DNA complexes weakly attached

to a mica surface, DNA loops were observed that changed in size, indicative of translocation. Eventually, cleavage of the loops resulted in progressive loss of DNA fragments from the field of view. Unfortunately, many features of the translocation process, such as bi-directional motion, remained indistinct. Moreover, the rate of this process would be affected by the attachment, albeit weakly, of the DNA to the mica.

In principle, single-molecule methods also ought to be able to provide direct demonstrations of random diffusion motion along DNA. In one attempt to observe diffusional motion by fluorescence (Kabata et al. 1993), a large number of λ phage DNA were attached to a solid surface and aligned parallel to each other by hydrodynamic flow; the motion of the protein was then monitored with an antibody tagged with multiple fluorescent labels. However, the resultant motion was across a 2-D plane of DNA molecules, rather than around a single molecule, and was no longer a random diffusion process but instead a unidirectional event, on account of the hydrodynamic flow. AFM has also been used to report on 1-D diffusion of proteins along DNA (Bustamante et al. 1999), but the observation that the protein is located at different position on the DNA after successive scans at 40 s time intervals leaves open the question of the route taken by the protein in the time between each scan.

Summary

In conclusion, it now appears that 1-D free diffusion of proteins along DNA lattices (i.e., sliding) plays a minor role, if any, in target-site location by DNA-binding proteins. Instead, and in contrast to previous views (viz. Shimamomto 1999), proteins find their target sites in DNA, both in vitro and in vivo, primarily by 3-D pathways involving multiple dissociation/re-association events (Stanford et al. 2000; Mistelli 2001). Nevertheless, a role for sliding may exist after the protein is positioned close to its recognition site, in the final docking of the protein with that sequence. The speed with which proteins bind to their specific sites is therefore not due to a reduction in the dimensionality of the search (Richter and Eigen 1974) but rather to a reduction in volume, in that the search is confined to the domain occupied by a single chain of DNA. On the other hand, it is now well established that the vectorial motion of proteins along DNA in a specified direction is often confined strictly to the 1-D contour of the DNA, even for processes that cover thousands of base pairs. Our understanding of the processive nature of these 1-D motions have been advanced considerably by biochemical, structural and single-molecule studies (e.g., Bianco et al. 2001; Soultanas and Wigley 2001).

The principal difference between diffusional and vectorial motions along DNA is that the former relies solely on thermal fluctuations while the latter requires an input of energy, derived in most cases from the hydrolysis of nucleoside triphosphates. Consequently, it is thermodynamically impossible for diffusional motion to change the mean position of a protein along a DNA, while the energy consumed during vectorial motion can be utilized to overcome the entropy barrier in moving the protein from one position to another in a specified direction. For any given ratio of rate constants for the dissociation of the protein from the DNA relative to that for its movement along the DNA, a protein that moves in a specified direction will cover a much greater distance before leaving the chain than one that has equal probabilities of moving in either direction.

Acknowledgements We thank members of our groups both past and present for their contributions to the work discussed above. We also thank Katie Wood for illustrating Fig. 1. Our work is currently funded by the Wellcome Trust, the BBSRC and the EPSRC. M.D.S. is a Wellcome Trust Research Career Development Fellow.

References

Ali JA, Lohman TM (1997) Kinetic measurement of the step size of DNA unwinding by *Escherichia coli* UvrD helicase. Science 275:377–380

Bellomy GR, Record MT Jr (1989) Stable DNA loops in vivo and in vitro: roles in gene regulation at a distance and in biophysical characterization of DNA. Prog Nucleic Acid Res Mol Biol 39:81–128

Berg HC (1993) Random walks in biology. Princeton University Press, Princeton

Berg OG (1978) On diffusion-controlled dissociation. Chem Phys 31:47–57

Berg OG, Winter RB, von Hippel PH (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. Biochemistry 20:6929–6948

Berge T, Ellis DJ, Dryden DTF, Edwardson JM, Henderson RM (2000) Translocation-independent dimerization of the *Eco*KI endonuclease visualized by atomic force microscopy. Biophys J 79:479–484

Bianco PR, Kowalczykowski SC (2000) Translocation step size and mechanism of the RecBC DNA helicase. Nature 405:368–372

Bianco PR, Brewer LR, Corzett M, Balhorn R, Yeh Y, Kowalczykowski SC, Baskin RJ (2001) Processive translocation and DNA unwinding by individual RecBCD enzyme molecules. Nature 409:374–378

Bustamante C, Guthold M, Zhu X, Yang X (1999) Facilitated target location on DNA by individual *Escherichia coli* RNA polymerase molecules observed with the scanning force microscope operating in liquid. J Biol Chem 274:16665–16668

Cook PR (1999) The organization of replication and transcription. Science 284:1790–1795

Craigie R, Mizuuchi K (1986) Role of DNA topology in Mu transposition: mechanism of sensing the relative orientation of two DNA segments. Cell 45:793–800

Davies GP, Powell LM, Webb JL, Cooper LP, Murray NE (1998) *Eco*KI with an amino acid substitution in any one of seven DEAD-box motifs has impaired ATPase and endonuclease activities Nucleic Acids Res 26:1775–1782

Davies GP, Kemp P, Molineux IJ, Murray NE (1999) The DNA translcoation and ATPase activities of restriction-deficient mutants of *Eco*KI. J Mol Biol 292:787–796

Dillingham MS, Wigley DB, Webb MR (2000) Demonstration of unidirectional single-stranded DNA translocation by PcrA helicase: measurement of step size and translocation speed. Biochemistry 39:205–212

Doi M, Edwards SF (1984) The theory of polymer dynamics. Oxford University Press, Oxford
Ehbrecht H-J, Pingoud A, Urbanke C, Maass G, Gualerzi C (1985) Linear diffusion of restriction endonucleases on DNA. J Biol Chem 260:6160–6166
Ellis DJ, Dryden DTF, Berge T, Edwardson JM, Henderson RM (1999) Direct observation of DNA translocation and cleavage by the *Eco*KI endonuclease using atomic force microscopy. Nat Struct Biol 6:15–17
Firman K, Szczelkun MD (2000) Measuring motion on DNA by the type I restriction endonuclease *Eco*R124I using triplex displacement. EMBO J 19:2094–2102
Garcia LR, Molineux IJ (1999) Translocation and specific cleavage of bacteriophage T7 in vivo by *Eco*KI. Proc Natl Acad Sci USA 96:12430–12435
George H, Kuraoka I, Nauman DA, Kobertz WR, Wood RD, West SC (2000) RuvAB-mediated branch migration does not involve extensive DNA opening within the RuvB hexamer. Curr Biol 10:103–106
Gorbalenya AE, Koonin EV (1991) Endonuclease (R) subunits of type-I and type-III restriction-modification enzymes contain a helicase-like domain. FEBS Lett 291:277–281
Gutfreund H (1995) Kinetics for the life sciences. Cambridge University Press, Cambridge
Halford SE, Gowers DM, Sessions RB (2000) Two are better than one. Nat Struct Biol 7:705–707
Hallet B, Sherratt DJ (1997) Transposition and site-specific recombination: adapting DNA cut-and-paste mechanisms to a variety of DNA rearrangements. FEMS Microbiol Rev 21:157–178
Ishijima A, Yanagida T (2001) Single molecule nanobioscience. Trends Biochem Sci 26:438–444
Jack WE, Terry BJ, Modrich P (1982) Involvement of outside sequences in the major kinetic path by which *Eco*RI endonuclease locates and leaves its recognition sequence. Proc Natl Acad Sci USA 79:4010–4014
Jankowsky E, Gross CH, Shuman S, Pyle AM (2000) The DExH protein NPH-II is a processive and directional motor for unwinding RNA. Nature 403:447–451
Janscak P, Bickle TA (2000) DNA supercoiling during ATP-dependent DNA translocation by the type I restriction enzyme *Eco*AI. J Mol Biol 295:1089–1099
Janscak P, Dryden DT, Firman K (1998) Analysis of the subunit assembly of the type IC restriction-modification enzyme *Eco*R124I. Nucleic Acids Res 26:4439–4445
Jeltsch A, Pingoud A (1998) Kinetic characterisation of linear diffusion of the restriction endonuclease *Eco*RV on DNA. Biochemistry 37:2160–2169
Jeltsch A, Alves J, Wolfes H, Maass G, Pingoud A (1994) Pausing of the restriction endonuclease *Eco*RI during linear diffusion on DNA. Biochemistry 34:10215–10219
Jeltsch A, Wenz C, Stahl F, Pingoud A (1996) Linear diffusion of the restriction endonuclease *Eco*RV on DNA is essential for the in vivo function of the enzyme. EMBO J 15:5104–5111
Kabata H, Kurosava O, Arai I, Washizu M, Marganson SA, Glass RE, Shimamoto N (1993) Visualization of single molecules of RNA polymerase sliding along DNA. Science 262:1561–1563
Kornberg A, Baker T (1992) DNA replication. Freeman, New York
Lever MA, Th'ng JP, Sun X, Hendzel MJ (2000) Rapid exchange of histone H1.1 on chromatin in living human cells. Nature 408:873–876
Linxweiler W, Horz W (1982) Sequence specificity of exonuclease III from *E. coli*. Nucleic Acids Res 10:4845–4859
Liu LF, Wang JC (1987) Supercoiling of the DNA template during transcription. Proc Natl Acad Sci USA 84:7024–7027
Lohman TM (1986) Kinetics of protein-nucleic acid interactions: use of salt effects to probe mechanisms of interactions. CRC Crit Rev Biochem 19:191–245
Lohman TM, Bjornson KP (1996) Mechanisms of helicase-catalyzed DNA unwinding. Annu Rev Biochem 65:169–214
Milsom SE, Halford SE, Embleton ML, Szczelkun MD (2001) Analysis of DNA looping interactions by type II restriction enzymes that require two copies of their recognition sites. J Mol Biol 31:517–528
Misteli T (2001) Protein dynamics: implications for nuclear architecture and gene expression. Science 291:843–847
Misteli T, Gunjan A, Hock R, Bustin M, Brown DT (2000) Dynamic binding of histone H1 to chromatin in living cells. Nature 408:877–881
Murray NE (2000) Type I restriction systems: sophisticated molecular machines. Microbiol Mol Biol Rev 64:412–434
Ostrander EA, Benedetti P, Wang JC (1990) Template supercoiling by a chimera of yeast GAL4 protein and phage T7 RNA polymerase. Science 14:1261–1265
Podtelezhnikov AA, Vologodskii AV (2000) Dynamics of small loops in DNA molecules. Macromolecules 33:2767–2771
Richter PH, Eigen M (1974) Diffusion controlled reaction rates in spheroidal geometry. Application to repressor-operator association and membrane bound enzymes. Biophys Chem 2:255–263
Riggs AD, Bourgeois S, Cohn M (1970) The *lac* repressor-operator interaction. III. Kinetic studies. J Mol Biol 53: 401–417
Ringrose L, Chabanis S, Angrand PO, Woodroofe C, Stewart AF (1999) Quantitative comparison of DNA looping in vitro and in vivo: chromatin increases effective DNA flexibility at short distances. EMBO J 18:6630–6641
Roberts RJ, Halford SE (1993) Type II restriction endonucleases. In: Linn SM, Lloyd RS, Roberts RJ (eds) Nucleases. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 35–88
Roman LJ, Eggleston AK, Kowalczykowski SC (1992) Processivity of the DNA helicase activity of *Escherichia coli* recBCD enzyme. J Biol Chem 267:4207–4214
Shimamoto N (1999) One-dimensional diffusion of proteins along DNA. J Biol Chem 274:15293–15296
Shulman MJ (1974) Model for wandering restriction enzymes. Nature 252:76–78
Singleton MR, Scaife S, Wigley DB (2001) Structural analysis of DNA replication fork reversal by RecG. Cell (in press)
Soultanas P, Wigley DB (2001) Unwinding the "Gordian knot" of helicase action. Trends Biochem Sci. 26:47–54
Stanford NP, Szczelkun MD, Marko JF, Halford SE (2000) One- and three-dimensional pathways for proteins to reach specific DNA sites. EMBO J 19:6546–6557
Studier FW, Bandyopadhyay PK (1988) Model for how type I restriction enzymes select cleavage sites in DNA. Proc Natl Acad Sci USA 85:4677–4681
Szczelkun MD (2000) How do proteins move along DNA? Lessons from type-I and type-III restriction endonucleases. Essays Biochem 35:131–143
Szczelkun MD (2002) Kinetic models of translocation, head-on collision and DNA cleavage by type I restriction endonucleases. Biochemistry 41:2067–2074
Szczelkun MD, Halford SE (1996) Recombination by resolvase to analyse DNA communications by the *Sfi*I restriction endonuclease. EMBO J 15:1460–1469
Szczelkun MD, Dillingham MS, Janscak P, Firman K, Halford SE (1996) Repercussions of DNA tracking by the type IC restriction endonuclease *Eco*R124I on linear, circular and catenated substrates. EMBO J 15:6335–6347
Terry BJ, Jack WE, Modrich P (1985) Facilitated diffusion during catalysis by *Eco*RI endonuclease. J Biol Chem 260:13130–13137
Uptain SM, Kane CM, Chamberlin MJ (1997) Basic mechanisms of transcript elongation and its regulation. Annu Rev Biochem 66:117–172
Vologodskii AV, Cozzarelli NR (1994) Conformational and thermodynamic properties of supercoiled DNA. Annu Rev Biophys Biomol Struct 23:609–643
von Hippel PH, Berg OG (1989) Facilitated target location in biological systems. J Biol Chem 264:675–678

Wang MD, Schnitzer MJ, Yin H, Landick R, Gelles J, Block SM (1998) Force and velocity measured for single molecules of RNA polymerase. Science 282:902–907

Webb JL, King G, Ternet D, Titheradge AJB, Murray NE (1996) Restriction by *Eco*KI is enhanced by co-operative interactions between target sequences and is dependent on DEAD box motifs. EMBO J 15:2003–2009

Winter RB, Berg OG, von Hippel PH (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The *Escherichia coli lac* repressor-operator interaction: kinetic measurements and conclusions. Biochemistry 20:6961–6977

Yin H, Wang MD, Svoboda K, Landick R, Block SM, Gelles J (1995) Transcription against an applied force. Science 270:1653–1657

Zhou W, Reines D, Doetsch PW (1995) T7 RNA polymerase bypass of large gaps on the template strand reveals a critical role of the nontemplate strand in elongation. Cell 82:577–585