Action at a distance: DNA-looping and initiation of transcription

Karsten Rippe, Peter H. von Hippel and Jörg Langowski

Effective initiation of transcription, especially in eukaryotes, requires the specific assembly of large protein complexes at promoters. We ask here how activator proteins that are bound hundreds or thousands of base pairs away from the promoter might facilitate this process if protein–protein interactions occur via looping of the intervening DNA. We show that the local concentration at the promoter of activator proteins bound at vicinal DNA sites can be substantially regulated by intrinsic or protein-induced distortion of the regular DNA conformation.

A MAJOR PART of the control of gene expression in both prokaryotes and eukaryotes occurs at the level of transcription initiation. Transcript 'throughput' for a given promoter is driven by a peculiar mixture of thermodynamics and kinetics; thermodynamics because, to maximize transcript yield, a promoter must initially bind a polymerase (in competition with other promoters and nonspecific sites), and kinetics because it must then move this polymerase through the initiation phase of transcription ('melting' of the DNA at the promoter, transcript initiation, etc.) into elongation as rapidly as possible to 'clear' the promoter and make it available for re-use[1].

In prokaryotes, a specificity factor (sigma factor) binds to the core polymerase in solution and directs the resultant holopolymerase to the relevant class of promoters. The accessibility of a given promoter to the polymerase is modulated by protein activators and repressors that bind at or near the promoter and facilitate or impede the binding of polymerase and the subsequent steps of initiation. The ability of these regulating proteins to interact with a promoter is a direct result of their binding affinity for specific DNA sequences in the promoter region[2].

DNA-binding sites for regulatory proteins at prokaryote promoters are of two types. Some are located close enough to the promoter sequence to permit proteins bound at these sites to make direct regulatory contact with the polymerase. Others may be located 50–200 base pairs (bp) upstream and thus make contact with the initiation complex at the promoter by interactions requiring looping of the DNA[3,4]. These latter protein-binding sites appear to be equivalent to the upstream elements located near eukaryotic promoters, and access to the initiation complex of proteins bound to these 'looping' binding sites should follow the same rules as those that apply in eukaryotes (see below). A notable example involves promoters that require the σ^{54} sigma factor[3,5]. At these promoters the nitrogen regulatory protein C (NTRC) activator protein binds near position −110 and cannot contact the polymerase without looping of the intervening DNA.

In eukaryotes, promoter activation is rendered more complex by the greatly increased numbers and types of promoters that are present in the DNA, and the need to engage in massive switching in promoter use as a consequence of developmental regulation and tissue-specific metabolism. The temporal and spatial expression pattern of the RNA transcripts produced is controlled by a large number of gene-specific transcription factors; thus, active transcription complexes in eukaryotes typically contain dozens of proteins.

As a consequence, the protein–DNA and protein–protein specificity problems involved in assembling these complexes at the proper promoters is formidable. This process cannot rely solely on specific binding of proteins directly to the promoter DNA, since specificity determinants at the promoter are limited and cannot offer the required regulatory flexibility. The problem appears to be solved, at least in part, by interactions between protein factors and specific binding sites on the surrounding (largely upstream) DNA. Such *cis* interactions permit sequences significantly removed along the DNA to participate in controlling the transcription initiation process at specific promoters, since proteins bound to such sequence-specific sites can be brought into contact with, and assemble into, specific activation complexes at the promoter as a consequence of DNA looping.

These *cis* interactions are of two types. One involves the binding of proteins to DNA sequences called 'upstream elements', which are located 100–200 bp from the promoter as defined by the TATA box. Related interactions occur also in prokaryotes. The other involves binding of protein factors to 'enhancer' sequences, which are found at positions up to several thousand base pairs from the promoter, and tend to be confined to eukaryotes. A schematic model illustrating these two types of interactions at eukaryotic RNA polymerase II (Pol II) promoters is shown in Fig. 1.

As indicated above, regulatory proteins bound at enhancers and upstream elements, either singly or as partially preassembled complexes, are thought to participate in the formation of the active transcription initiation complex by looping of the DNA to permit interaction with the promoter (for recent reviews, see Refs 6, 7). Although experimental evidence supports such a looping mechanism[8,9], it is notable that there are also other possible mechanisms by which enhancers could work at a distance. These mechanisms include nucleosomal rearrangements induced by proteins binding to enhancers[10], transcriptional activation by tracking proteins[11] and transmission of an altered DNA structure as, for example, by changes in the degree of DNA supercoiling[12].

K. Rippe and J. Langowski are at the German Cancer Research Center, Department Biophysics of Macromolecules, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany; and P. H. von Hippel is at the Institute of Molecular Biology, University of Oregon, Eugene, OR 97403, USA.

Clearly, the effectiveness of DNA looping in bringing protein factors to the promoter at high enough concentrations and with appropriate orientations, is a quantitative problem and will depend on the flexibility and conformation of the intervening DNA. Both DNA flexibility and conformation are determined by structural features inherent to the DNA sequence as well as the effect of bound complexes, such as nucleosomes. Here, we describe some recent progress in assessing the thermodynamic contributions of DNA-loop formation. We hope that this work might serve as a guide for interpreting experimental results, and for estimating the potential contributions of enhancers located at different distances from the promoter. The results described may also help to focus consideration of the effects of different conformations (induced by the binding of other regulatory proteins or DNA supercoiling) on the looping properties of the intervening DNA. Interactions between proteins that are bound on the same DNA molecule at separated sites can play important roles in many biological processes in addition to their roles in transcription initiation as discussed here, for example, DNA recombination, DNA replication and transcription termination. All of these interactions are regulated by the same principles and, thus, should be amenable to treatment by appropriately modified approaches of the sort presented here.

In general, the tethering of two proteins (or protein complexes) on the same piece of DNA can increase the probability of protein–protein contacts, with the intervening DNA being looped out to potentiate interactions that would not occur if the proteins were free in solution (see Fig. 2). The probability of finding two proteins that are bound on the same DNA molecule within a volume that allows them to contact each other can be expressed as the local concentration in moles per liter of one binding site in the proximity of the other. This concentration will be called j_M in the following treatment, and corresponds to a probability density.

As described above, transcription by eukaryotic Pol II is controlled by regulatory proteins that bind at sites from which DNA looping is required to allow physical contact between these proteins and the transcription factors bound at the promoter. We will show that a simple DNA linkage between promoter and upstream element/enhancer, with no constraints on the conformation of the intervening DNA, leads to an effective local concentration of activator protein of only approximately 10^{-8}–10^{-9} M in the proximity of the promoter for typical separation distances of 100–200 bp (upstream elements) or several thousand bp (enhancers).

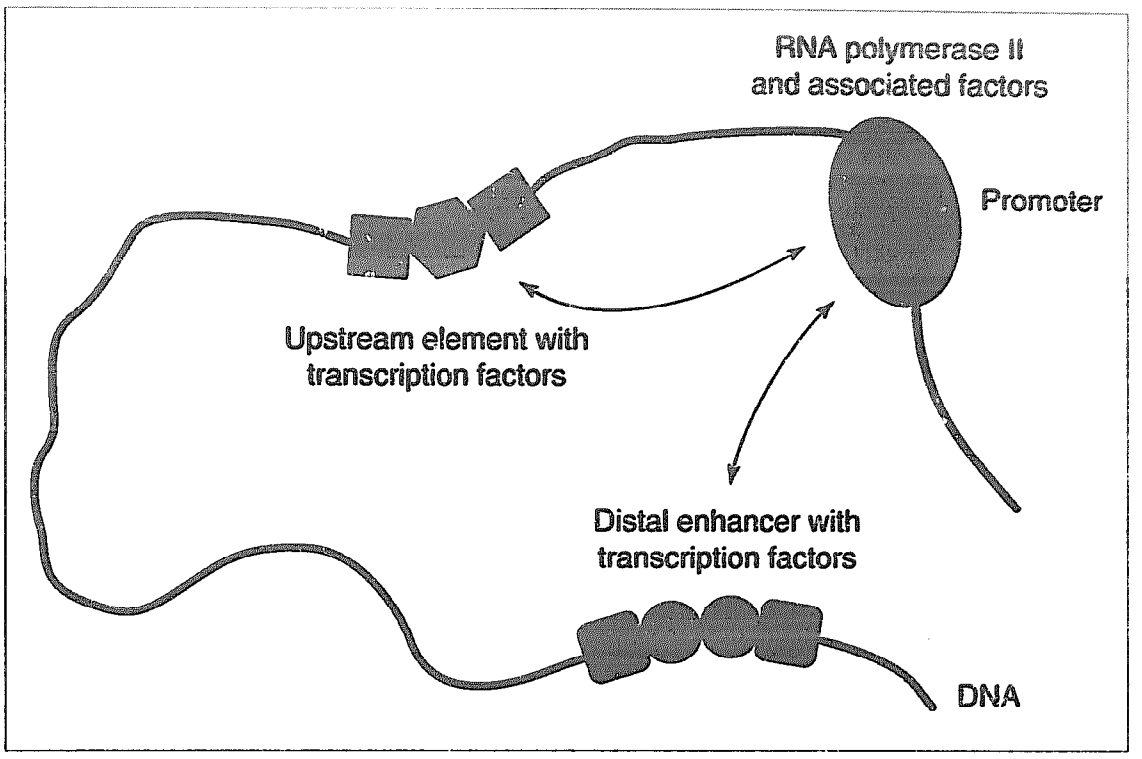

Figure 1

DNA looping brings proteins bound at upstream elements and enhancers to the promoter. In this simplified view, the many protein subunits comprising the RNA polymerase II (Pol II) transcription complex are not shown separately. Upstream elements are typically found at a distance of 100–200 bp away from the transcription start, whereas enhancers can be located several thousand base pairs upstream or downstream of their target promoters.

It is generally accepted that activator proteins bound at upstream elements or enhancers serve to recruit a protein or a protein complex that is a part of the general transcription machinery to a given promoter. This recruitment is thought to facilitate the assembly of the active transcription complex[6]. Clearly, such a thermodynamic recruitment mechanism cannot operate usefully if the end result is to generate effective protein concentrations at the promoter that are as low as 10^{-8} to 10^{-9} M, since the proteins of the general Pol II transcription machinery are likely to be present in free solution at concentrations of at least this order of magnitude within the eukaryotic nucleus (100–1000 molecules of a given protein factor per nucleus corresponds to a concentration between 10^{-7} and 10^{-9} M). These estimates suggest that either the activation process proceeds via a different mechanism or that j_M, the effective local concentration of a given protein at the promoter that is induced by looping, is actually much higher. The aim of this review is to provide estimates for j_M for a number of loop sizes, and to show how the effective local protein concentrations can be changed by manipulating the conformation of the intervening DNA. We will demonstrate that the three-dimensional organization of the DNA by other proteins, by DNA curvature or by topological constraints, can raise j_M to concentrations of 10^{-4}–10^{-5} M, thus permitting assembly of the entire complex to be driven by much more modest protein–protein association constants.

DNA cyclization as a model for protein–protein interaction by DNA looping

To a first approximation, protein–protein interactions mediated by DNA looping can be compared with DNA cyclization reactions in which a protein ligase drives the transformation of a linear DNA fragment to a covalently closed circle as a consequence of DNA ring closure. This process has been much studied, both experimentally and theoretically (see Refs 13–21). The apparent similarity of cyclization reactions and DNA looping has also been used to describe interactions between *lac* repressor proteins bound at separate DNA sites *in vivo* (reviewed in Ref. 22). In the following section we will review results that are derived from DNA cyclization experiments and theoretical work on DNA flexibility to obtain estimates of the effective local protein concentration increase that can be induced by DNA looping. In this context it is

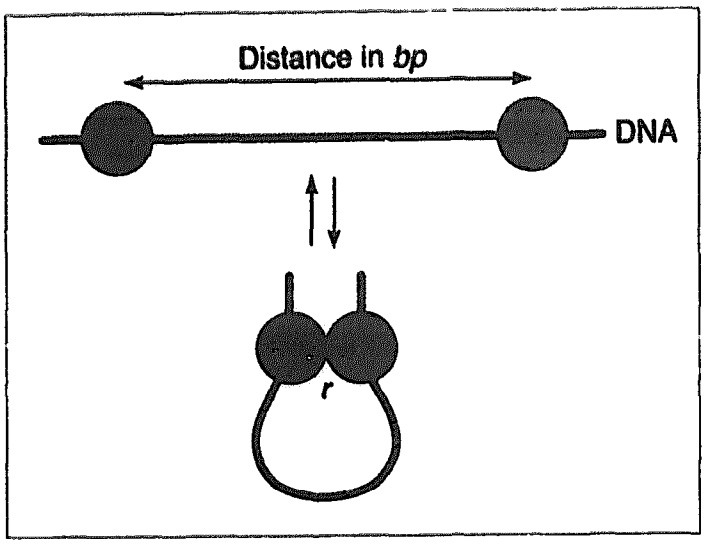

Figure 2
Schematic view of protein–protein contacts mediated by DNA looping. The two proteins are bound on the same DNA strand. The centers of the two binding sites are separated on the DNA by a distance of *bp* base pairs and must approach one another to a distance *r* to permit the two proteins to make contact. For the model calculations presented here we use a value of 10 nm for both *r* and the protein diameter.

important to appreciate the following points.

(1) To permit cyclization, the ends of the DNA fragment must come very close together, i.e. the end-to-end distance *r* must approach zero. By contrast, protein–protein contacts can occur if the interacting DNA sites are brought to a distance (*r*) of only 5–15 nm, depending on the size of the proteins. This difference becomes noticeable if the intervening DNA length is <350 bp (see below). In our calculations we will use a value of $r = 10$ nm as an average separation distance for typical protein–protein contacts induced by looping (see Fig. 2).

(2) The occupancies, θ_1 and θ_2, of the two protein binding sites that contact one another, as well as interactions with the same proteins not bound to the DNA, also need to be considered. In the balance of this presentation we assume that the interacting sites are always fully occupied and that the concentration of protein not bound to the DNA, c_{free}, is much smaller than j_M. Under these conditions, the effective concentration of one protein with respect to the other, c_{eff} is equal to j_M. If this is not the case c_{eff} can be calculated according to Eqn 1:

$$c_{eff} = \theta_1 \times \theta_2 \times j_M + c_{free} \qquad (1)$$

(3) Both DNA cyclization and biologically functional protein–protein contacts require correct torsional alignment of the two sites. The concentrations given by j_M are equivalent to the same concentration of a species free in solution. If the torsional orientation of the sites on the DNA is favorable, the interaction is facilitated as compared to interactions free in solution; for an unfavorable orientation, the interaction will be inhibited. For DNA cyclization, this effect can change j_M as much as tenfold for fragments <800 bp in length[16,17]. For protein–protein interactions, we can use two experimental studies to estimate that unfavorable torsional alignment may reduce j_M as much as tenfold if the length of the intervening DNA is 60–130 bp[23], and as much as fivefold for a DNA length of 130–200 bp[24]. In general, the dependence on the helical periodicity of the DNA will be less pronounced for longer separation distances, and should be hardly noticeable above 800 bp. For simplicity, we assume in what follows that no twisting of the DNA is required to align the interacting proteins. Thus, the j_M values presented should be considered as averages, though probably accurate within less than an order of magnitude.

(4) The effective protein concentrations (j_M values) produced by DNA looping that are presented here are based on theoretical and experimental work that describes the distribution of end-to-end or equivalent site-to-site distances between pairs of proteins bound to various types and conformations of intervening DNA. The effects on this parameter that are due to the excluded volumes of the proteins themselves are not included in these calculations. This excluded volume effect can be safely neglected if the total protein volume is smaller than 5% of the volume occupied by the DNA fragment. For two proteins with 10 nm diameters this corresponds to an end-to-end DNA separation distance of approximately 100 bp for B-form double-stranded DNA.

(5) At physiological salt concentrations DNA has an effective diameter of about 4 nm (see Ref. 25 and references therein). This can lead to a <10% error in the calculated values of j_M for linear DNA fragments of 2000–10 000 bp in length on the basis of the freely jointed chain model (see below). For all other values of j_M presented here, excluded volume effects are not relevant because they are included explicitly in the computer simulations or can be neglected for short DNA fragments.

Long linear DNA fragments

Linear double-stranded DNA molecules with lengths greater than 2000 bp can be described as idealized chains of *n* segments of statistical segment length *l*, in which the chain segments are not restricted in their movement with respect to one another. Such a chain is termed a freely jointed chain, and the end-to-end distribution function $W(r)$ can be shown to be a Gaussian distribution according to Eqn 2 (reviewed in Ref. 14).

$$W(r) = \left(\frac{3}{2\pi n l^2}\right)^{\frac{3}{2}} \times \exp\left(\frac{-3r^2}{2nl^2}\right) \qquad (2)$$

Here $W(r)$ is the probability density that the two ends are separated by a certain end-to-end distance *r*. For double-stranded B-form DNA *l* is, to a good approximation, 100 nm (300 bp) at monovalent salt concentrations >0.01 M (Ref. 26). Since the function $W(r)$ gives the probability density as a probability per unit volume, this is equivalent to the concentration at which one end of a polymer chain is present at distance *r* from the other end. By analysing Eqn 2, it turns out that $W(r)$ changes very little if *r* is between 0 and 20 nm. For our purposes we can therefore use the simpler expression (Eqn 3), that permits us to calculate the parameter $j_M(bp)$ (Eqn 3) in $mol\,l^{-1}$ for a DNA fragment that is a certain number of bp in length:

$$j_M(bp) = 0.0028 \times bp^{-\frac{3}{2}} \left[\frac{\text{mol}}{\text{l}}\right] \qquad (3)$$

for $r = 0$–20 nm and $2000 < bp < 10\,000$.

A statistical segment length *l* of 100 nm indicates that DNA is a relatively stiff polymer. For single-stranded RNA [such as poly(U)] a much smaller statistical segment length of 4 nm has been determined[27] (mixed RNA sequences are likely to have a somewhat higher statistical segment length because of the formation of secondary structures). Thus protein–protein contacts mediated by RNA looping that, for example, have been observed for the antitermination protein N from phage λ bound to particular hairpin structures on the nascent RNA in transcription

antitermination processes[28], can increase j_M up to concentrations in the range of 10^{-3} M, because the single-stranded RNA is much more flexible than double-stranded B-DNA. The appropriate expression for treating such RNA systems can be derived by using Eqn 2 of this review and Eqn 50 of Ref. 17 with $l = 4$ nm instead of $l = 100$ nm, and 6.5 Å (instead of 3.4 Å) as the average internucleotide distance[27] to derive the equivalent of Eqns 3 and 4 for RNA. This approach has been used elsewhere to consider RNA looping interactions in N-dependent transcriptional antitermination (W. A. Rees, S. E. Weitzel, A. Das and P. H. von Hippel, unpublished).

Short linear DNA fragments

The freely jointed chain model does not describe DNA looping adequately for linear DNA fragments shorter than 2000 bp (or for RNA looping for single-strand chain lengths shorter than 50 nucleotides). In this range another model, the Kratky–Porod wormlike chain[29], provides a good approximation and has been widely used. On the basis of this model for DNA cyclization, Eqn 50 of Ref. 17 gives a numerical approximation for the probability of finding the two ends of the polymer at a separation distance $r = 0$.

The Shimada–Yamakawa equation can be converted to molar concentrations (J_{M_0}; the index zero is used to indicate that $r = 0$) for DNA lengths in base pairs, using 3.4 Å per base pair (B-form DNA) and 100 nm for the statistical segment length. Equation 4 is the approximation obtained and is valid for DNA lengths of about 100–1500 bp.

$$j_{M_0}(bp) = \frac{1.024 \times 10^8}{bp^5} \times \exp\left(\frac{-2067}{bp} + 0.001673 \times bp\right)\left[\frac{\text{mol}}{1}\right] \quad (4)$$

for $r = 0$ nm and $100 < bp < 1500$.

Eqns 3 and 4 have been used to calculate j_M for ring-closure for DNA fragments ranging from 100 to several thousand bp in length. The results are shown in Fig. 3 (see Ref. 30).

Eqn 4 was derived for the DNA cyclization situation, i.e. for $r = 0$. In contrast to the situation for long DNA fragments (>2000 bp; see above), where j_M was shown to be about the same for values of r ranging from 0 to 20 nm, we expect that for short DNA fragments the dependence of j_M on r will become significant. This is very relevant for protein–protein interactions in eukaryotic transcription, where many regulatory proteins bind at upstream elements located at distances of 100 to 200 bp from the RNA polymerase at the transcription start site.

Unfortunately, it is very difficult to obtain an analytical solution for the probability of finding the ends of a Kratky–Porod wormlike chain at a certain value of r. For this reason, this calculation has generally been made by using computer simulations to determine the entire distribution of end-to-end or site-to-site distances for a given DNA fragment[18,19]. This procedure has the additional advantage that parameters such as DNA curvature or superhelicity can be incorporated into the model[21,31,32].

In Fig. 4 the local concentration (j_M) at $r = 0$ nm and at $r = 10$ nm is plotted versus the intervening DNA length. The data were derived from brownian dynamics simulations (H. Merlitz, K. Klenin, K. Rippe and J. Langowski, unpublished) using a DNA model similar to that in Ref. 32, but with an explicit treatment of electrostatic interactions. We note that the DNA length scales in Figs 3 and 4 are very different, and that the data presented in Fig. 4 correspond to the very DNA-length-dependent region of the plot located near the origin in Fig. 3.

Figure 4 shows that at DNA lengths of >350 bp there is no significant difference between the curves for $r = 0$ nm and for $r = 10$ nm (the distance relevant for protein–protein interactions). At 150 bp we have a tenfold higher value at $r = 10$ nm as compared to $r = 0$ nm. However, the concentration at this point is still lower by a factor of ten than at the maximum around 500 bp (compare also with Fig. 3). In connection with the upstream activation elements of eukaryotic promoters, it might be asked why these are generally located 100–200 bp away from the promoter if the local concentration at the promoter would be higher at a separation distance of 500 bp. One possible explanation is that DNA curvature increases the local concentration.

To illustrate this effect we have calculated, and present in Fig. 4, values for j_M at $r = 10$ nm that apply for DNA fragments that have been 'kinked' by introducing a central bend of 120° into the DNA. This bend loses its effect on j_M

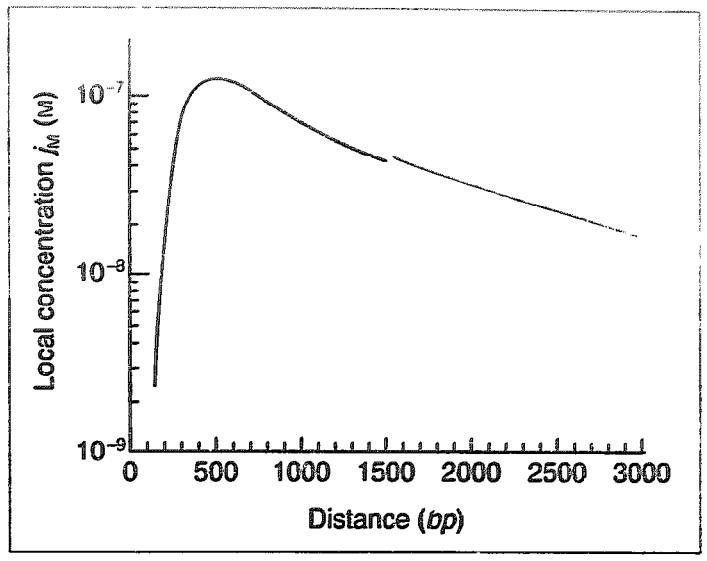

Figure 3
Dependence of j_M on the length of the intervening DNA. The parameter j_M is the local concentration of one end of a linear DNA fragment in the proximity of the other end. The values were calculated according to Eqn 3 for the freely jointed chain (red), and according to Eqn 4 for the wormlike chain (blue). Note that in the region from 1500 to 2000 bp the freely jointed chain model gives j_M values that are a few percent too high, as this model cannot accurately describe DNA sequences shorter than 2000 bp. It can be seen that j_M decays only slowly above 1000 bp, which fits with the observation that the activation efficiency of enhancers is relatively insensitive to changes in promoter to enhancer distance in the 1000–5000 bp range[30].

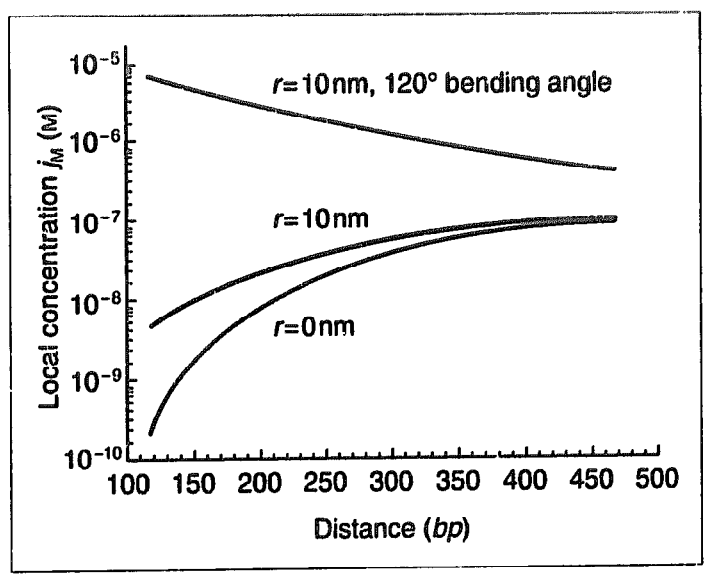

Figure 4
Local concentration j_M for short linear DNA fragments. Comparison of j_M at $r = 0$ (red), at $r = 10$ nm (blue), and at $r = 10$ nm (green), with a central bending angle of 120°. The data were derived from Brownian dynamics simulation. The simulations were performed using a statistical segment length of 100 nm, at 25°C and in 0.1 M NaCl.

Table I. Effect of the DNA conformation on the local concentration increase caused by DNA looping[a]

	Local concentration (j_M) in mol l^{-1}				
	Separation distance in base pairs				
DNA conformation	150 (0)[b]	150 (10)[b]	500 (0–10)[b]	1000 (0–20)[b]	10 000 (0–20)[b]
Linear[c]	2×10^{-9}	1×10^{-8}	1×10^{-7}	8×10^{-8}	3×10^{-9}
Linear, curved[d]	1×10^{-8}	5×10^{-6}	4×10^{-7}	1×10^{-7}	3×10^{-9}
Relaxed circle[e]	1×10^{-9}	1×10^{-8}	1×10^{-7}	6×10^{-8}	–
Superhelical[f]	1×10^{-6}	1×10^{-6}	5×10^{-6}	5×10^{-6}	–
Superhelical, curved[g]	–	–	7×10^{-5}	4×10^{-5}	–

[a]The local concentration (in mol l^{-1}) of one end or site of a DNA fragment in the proximity of the other end or site is given for separation distances of 150 bp, 500 bp, 1000 bp and 10 000 bp. For fragments of 500 bp or longer there is no significant difference between the concentration if the ends are at $r = 0$ (DNA cyclization) or at $r = 20$ nm (for protein–protein interactions).
[b]The numbers in brackets indicate the end-to-end or site-to-site distances (r, in nm; see Fig. 2) to which the given concentration values refer.
[c]Values calculated according to Eqns 2 and 3, see text.
[d]The curvature was located in the center of the fragment with a DNA bending angle of 120°.
[e]Values are given for a circle of 3.5 kb and are taken from Ref. 31.
[f]Values are given for a circle of 2.7 kb (Ref. 32) or 3.5 kb (Ref. 31), and a superhelical density of $\sigma = -0.06$.
[g]Values are given for a circle of 2.7 kb and a superhelical density of $\sigma = -0.06$ (Ref. 32). The bending angle is 120° and is located at the center between the two sites. Similar values are obtained at a bending angle of 90° (see Fig. 5).

with increasing lengths of intervening DNA; thus, at a separation distance of about 500 bp this bend results only in an approximately fourfold increase in the local concentration (j_M) as compared to an equivalent straight fragment. However, j_M increases to a value of 5×10^{-6} M at 150 bp, the separation at which upstream elements are generally found, and this value of j_M is 50 times higher than at the 500 bp peak of the plot of Fig. 3 for unbent intervening DNA. Thus, for strongly bent DNA fragments a shorter separation distance (100–200 bp) is optimal for interactions between activator proteins and RNA polymerase at the promoter. In this context, it is also noteworthy that the general transcription factor known as the TATA-box-binding protein, or TBP, introduces an 80° bend into the DNA at a position located about 30 bp upstream of the transcription start[33]. This bend, especially in conjunction with additional intrinsic DNA curvature, is likely to facilitate interactions between the upstream elements and the promoter, similar to the effects of the 120° bend shown here.

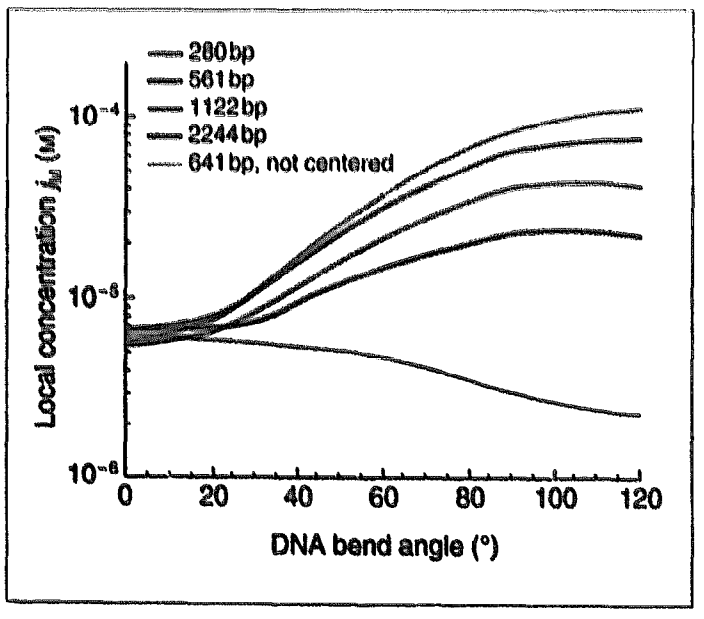

Figure 5
Local concentration in superhelical DNA and the effect of curvature. The concentration of one DNA segment in a sphere of 10 nm around a second segment is plotted versus the bending angle for a 2.7 kb plasmid with a superhelical density of $\sigma = -0.06$. Data and figure are adopted from Ref. 32. Separation distances were 280 bp, 561 bp, 1122 bp, 2244 bp, and 641 bp. For all separation distances except 641 bp, the bend was centered between the two sites. For the 641 bp distance the bend was located 140 bp away from one site and 561 bp away from the second site. The data were obtained from Monte-Carlo simulations with a statistical segment length of 100 nm, at 25°C and 0.1 M NaCl.

Topologically constrained DNA

The DNA around a promoter can often be constrained into the equivalent of a closed circular DNA loop by, for example, the effects of neighboring nucleosomes, thus permitting supercoiling of the intervening DNA. Such supercoiling introduces additional constraints and opportunities on the values of j_M that can be introduced by DNA looping, and can be modeled by considering the distances between two DNA segments (protein-binding sites) located at varying separations on a covalently closed circular DNA molecule.

At a given separation in base pairs, two segments in a covalently closed circular DNA will be closer to one another than two segments in a linear DNA of the same length. The enhancement of j_M caused by the ring closure has been calculated for the freely jointed chain model[34], and plays a role only when the separation distance is of the order of half the length of the DNA circle (Table I). On the other hand, the effect is much more pronounced when the circular DNA is internally twisted to form a superhelix. The enhancement of the local concentration of one segment relative to another owing to supercoiling was first treated theoretically by Vologodskii *et al.*[31] Using Monte-Carlo calculations, local concentrations in the range of 10^{-6} M were found for a 3.5 kb DNA circle with a superhelical density of $\sigma = -0.06$ and with separation distances of 150–1200 bp (Table I). This represents an enhancement of two orders of magnitude over the corresponding values for relaxed circular DNA.

The computations in Ref. 31 assume homogeneous bending and twisting elasticity along the DNA chain, and a straight equilibrium conformation. It has been argued that sequence-specific DNA bends may influence the conformation of superhelical DNA over long distances by determining the positions of the ends of the interwound structure. Thus, since the loops located at the 'ends' of the supercoiled DNA ring are more strongly bent than the rest of the structure, forming them at the position of a permanent bend will be thermodynamically more favorable than forming them elsewhere. Experimental evidence for this effect has been provided by Laundon and Griffith[35], who showed that curved segments can orient the structure of superhelical DNA, and by Kremer and co-workers[36], who demonstrated a global decrease of structural fluctuations in superhelical DNA by the insertion of a curved sequence.

In general, sequence inhomogeneities are expected to influence the global structure of a superhelical DNA. Thus, sequences of increased flexibility, permanent curvature, or associated with

DNA-bending proteins, are more likely to be found in places where the bending strain of the superhelix structure is high, i.e. at the loops located at the ends of supercoiled DNA molecules or at branch points in supercoiled structures. A quantitative treatment of this effect has been presented by Klenin and co-workers[32], using a Monte-Carlo model similar to that of Vologodskii *et al.*[31], with appropriate modifications to accommodate the curved sequences. The end result is that a curved DNA segment inserted in a superhelix will lead to a strong increase in the local concentration of one site in the vicinity of the other for a pair of DNA segments that are located symmetrically with respect to the curved segment, and to a decrease in local concentrations otherwise.

An example of this effect is shown in Fig. 5. We see that two points located at a separation of 280 bp in a 2.7 kb superhelical DNA of superhelical density $\sigma = -0.06$ will be characterized by a j_M value of 5×10^{-6} M when the DNA between them is homogeneous. By contrast, when a 120° bend is inserted halfway between the two segments, the local concentration increases to 10^{-4} M. For two segments located asymmetrically with respect to the bend, j_M decreases by a factor of three when the segments are 140 and 561 bp from the curved segment, respectively.

It is therefore very tempting to speculate that local bending might serve to 'steer' such action-at-a-distance effects in DNA. One unresolved issue is the timescale in which such global rearrangements of DNA structure might take place. The results of Kremer and co-workers[36] suggest that the presence of intrinsic bends can change the kinetics of intramolecular rearrangements within the DNA. Brownian dynamics simulations show that a superhelical structure relaxes to conformational equilibrium in characteristic times of milliseconds or more. Similar timescales are expected for the rearrangement of branches or end-loops after the introduction of a bend. Since such timescales are comparable to the rates of enzymatic processes acting on DNA, we can expect that the kinetics of specific DNA–protein interactions will also be influenced by such structural rearrangements of superhelical DNA.

Concluding remarks

The influence on the local concentrations of separated DNA sites with different conformations of intervening DNA are summarized in Table I. The data presented suggest that, in most cases, mere DNA connectivity between two sites will not be very effective in promoting protein–protein interactions in an equilibrium sense. We show, especially for proteins bound at upstream element sites (at distances of 100–200 bp along the DNA), that intrinsic or protein-induced DNA curvature, or superhelical torsion, can serve to impose constraints on local DNA conformation to create specific three-dimensional arrangements that favor such equilibrium 'action at a distance'. We show also that looping does not significantly increase the equilibrium concentrations of individual activator proteins bound to enhancer sites (located at distances of 1000–5000 bp from the promoter) above the free-solution concentrations that might be expected to be present. In addition, specific kinking or bending of the intervening DNA at these distances is unlikely to change the local protein concentrations induced by looping at the promoter very much. How then might enhancers, if they indeed work by raising the local equilibrium concentrations of proteins, promote formation of the active transcription complex at the promoter?

For long range interactions (i.e. 1000 bp or greater), superhelicity, especially in conjunction with curvature located symmetrically relative to the two sites, can be very effective in bringing two sites close to one another. The organization of the DNA by nucleoprotein complexes is likely to lead to similar results (reviewed in Ref. 37). In addition, enhancers could facilitate the formation of activator complexes in an equilibrium sense by binding protein components as pre-organized subassemblies, thus giving these complexes a significant equilibrium advantage over individual protein subunits arriving at the promoter free in solution. This idea is consistent with the fact that enhancers often contain several spaced repeats of protein-binding sites that act synergistically (for example, see Ref. 38).

The notion that a particular DNA arrangement is required for long-range interactions in general is interesting, both with respect to the regulation and to the specificity of biological processes such as transcription. One class of enhancers might, for example, not be involved in direct interaction with the transcription complex, but rather could stabilize a certain DNA conformation to potentiate interactions between the promoter and other enhancer sequences, as discussed by others[37,39]. The stereospecific requirements for long-range interactions might also explain why the effect of transcriptional enhancers is usually restricted to a single promoter, even if several other promoters are located nearby.

Acknowledgements

This work was supported in part by a salary grant from the Boehringer Ingelheim Fonds while K. R. was a postdoctoral fellow at the University of Oregon, and in part by DFG grant La500/4-1 and a startup grant by the Verein zur Förderung der Krebsforschung (to J. L.) and USPHS grants GM-15792 and 29158 (to P. H. vH.). P. H. vH. is an American Cancer Society Research Professor of Chemistry.

References

1 von Hippel, P. H., Yager, T. D. and Gill, S. C. (1992) in *Transcriptional Regulation* (Vol. 1) (McKnight, S. L. and Yamamoto, K. R., eds), pp. 179–201, Cold Spring Harbor Laboratory Press
2 von Hippel, P. H. and Berg, O. G. (1989) in *Protein–Nucleic Acid Interaction* (Saenger, W. and Heinemann, U., eds), pp. 1–18, Macmillan Press
3 Gralla, J. D. (1991) *Cell* 66, 415–418
4 Schleif, R. (1992) *Annu. Rev. Biochem.* 61, 199–223
5 Weiss, D. S. *et al.* (1992) in *Transcriptional Regulation* (Vol. 2) (McKnight, S. L. and Yamamoto, K. R., eds), pp. 667–694, Cold Spring Harbor Laboratory Press
6 Hori, R. and Carey, M. (1994) *Curr. Opin. Genet. Dev.* 4, 236–244
7 Koleske, A. J. and Young, R. A. (1995) *Trends Biochem. Sci.* 20, 113–116
8 Dunaway, M. and Dröge, P. (1989) *Nature* 341, 657–659
9 Müller, H-P., Sogo, J. M. and Schaffner, W. (1989) *Cell* 58, 767–777
10 Grunstein, M. *et al.* (1992) in *Transcriptional Regulation* (Vol. 2) (McKnight, S. L. and Yamamoto, K. R., eds), pp. 1295–1315, Cold Spring Harbor Laboratory Press
11 Tinker, R. L., Williams, K. P., Kassavetis, G. A. and Geiduschek, E. P. (1994) *Cell* 77, 225–237
12 Wang, J. C. and Giaever, G. N. (1988) *Science* 240, 300–304
13 Jacobson, H. and Stockmayer, W. H. (1950) *J. Chem. Phys.* 18, 1600–1606
14 Flory, P. J. (1969) *Statistical Mechanics of Chain Molecules*, Wiley
15 Shore, D., Langowski, J. and Baldwin, R. L. (1981) *Proc. Natl Acad. Sci. USA* 78, 4833–4837
16 Shore, D. and Baldwin, R. L. (1983) *J. Mol. Biol.* 179, 957–981
17 Shimada, J. and Yamakawa, H. (1984) *Macromolecules* 17, 689–698
18 Levene, S. D. and Crothers, D. M. (1986) *J. Mol. Biol.* 189, 61–72
19 Hagerman, P. J. and Ramadevi, V. A. (1990) *J. Mol. Biol.* 212, 351–362
20 Hagerman, P. J. and Ramadevi, V. A. (1990) *J. Mol. Biol.* 212, 363–376

21 Crothers, D. M., Drak, J., Kahn, J. D. and Levene, S. D. (1992) *Methods Enzymol.* 212, 3–29
22 Bellomy, G. R. and Record, M. T. (1990) *Prog. Nucleic Acids Res.* 39, 81–128
23 Haykinson, M. J. and Johnson, R. C. (1993) *EMBO J.* 12, 2503–2512
24 Law, S. M., Bellomy, G. R., Schlax, P. J. and Record, M. T. (1993) *J. Mol. Biol.* 230, 161–173
25 Vologodskii, A. V. and Cozzarelli, N. R. (1994) *Curr. Opin. Struct. Biol.* 4, 372–375
26 Hagerman, P. J. (1988) *Annu. Rev. Biophys. Biophys. Chem.* 17, 265–286
27 Inners, L. D. and Felsenfeld, G. (1970) *J. Mol. Biol.* 50, 373–389
28 Das, A. (1993) *Annu. Rev. Biochem.* 62, 893–930
29 Kratky, O. and Porod, G. (1949) *Rec. Trav. Chim.* 68, 1106–1113
30 Schreiber, E. and Schaffner, W. (1989) *Somatic Cell Mol. Genet.* 15, 591–603
31 Vologodskii, A. V. *et al.* (1992) *J. Mol. Biol.* 227, 1224–1243
32 Klenin, K., Frank-Kamenetskii, M. D. and Langowski, J. (1995) *Biophys. J.* 68, 81–88
33 Kim, J. L., Nikolov, D. B. and Burley, S. K. (1993) *Nature* 365, 521–527
34 Bloomfield, V. A., Crothers, D. M. and Tinoco, J. I. (1974) *Physical Chemistry of Nucleic Acids*, Harper & Row
35 Laundon, C. H. and Griffith, J. D. (1988) *Cell* 52, 545–549
36 Kremer, W., Klenin, K., Diekmann, S. and Langowski, J. (1993) *EMBO J.* 12, 4407–4412
37 Wolffe, A. P. (1994) *Curr. Opin. Genet. Dev.* 4, 245–254
38 Chang, C. and Gralla, J. D. (1993) *Mol. Cell. Biol.* 13, 7469–7475
39 Tjian, R. and Maniatis, T. (1994) *Cell* 77, 5–8

The G-box: a ubiquitous regulatory DNA element in plants bound by the GBF family of bZIP proteins

Anne E. Menkens, Ulrike Schindler and Anthony R. Cashmore

The G-box (CACGTG) is a ubiquitous, *cis*-acting DNA regulatory element found in plant genomes. Proteins known as G-box factors (GBFs) bind to G-boxes in a context-specific manner, mediating a wide variety of gene expression patterns. We suggest that, as for many biological systems, different combinations of these common elements can lead to diversity and specificity in the regulation of plant gene expression.

THE G-BOX IS a hexameric motif, CACGTG, found in many diverse plant genes. This sequence functions as a *cis*-acting promoter element, and was first characterized in early studies on the 5′-noncoding region of the light-regulated ribulose 1,5-bisphosphate small subunit (*RBCS*) genes[1]. In these studies, three conserved sequences were identified which we designated boxes L, I and G. At that time, our thinking was that these elements may be unique to light-regulated genes and would play an important role in their regulation. While our later studies demonstrated the essential role of the G- and I-boxes in light-regulated gene expression[2], many groups have shown that the G-box sequence resides in the promoters of many genes that are switched on in response to quite diverse stimulatory pathways. The analysis of the G-box as a regulatory element has been complemented by the isolation of a family of bZIP proteins that specifically interact with this element. In the case of *Arabidopsis*, we have referred to these proteins as G-box-binding factors (GBFs), and in this article we use this term to include the homologous bZIP proteins described for other plant species.

A. E. Menkens is at the Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104, USA; **U. Schindler** is at Tularik Incorporated, South San Francisco, CA 94080, USA; and **A. R. Cashmore** is at the Leidy Laboratory of Biology, Philadelphia, PA 19104-6018, USA.

G-box function in regulated gene expression

G-box elements have been demonstrated to be essential functional components of many stimulus-responsive promoters. This has been done by fusing either deleted or mutated versions of various plant promoters to reporter genes, and assaying for transcriptional activity in either transient assays or transgenic plants. Representative examples of promoters used in such studies, and the regulatory elements that were identified are listed in Table I. Comparison of these results show that the G-box plays a role in the responsiveness of plant promoters to light[2,3], anaerobiosis[4], *p*-coumaric acid[5], and hormones such as abscisic acid[6], ethylene[7] and methyl jasmonate[8]. In these examples, deletion or mutation of the G-box dramatically reduces the overall promoter activity and/or renders the promoter unresponsive to a given stimulus. Importantly, in each of these promoters, the G-box resides in a unique DNA context and additional elements are critical to the appropriate response. For example, mutation of the flanking I-box sequences in the *Arabidopsis rbcS-1A* gene rendered the promoter unresponsive to light[2].

In addition to studies involving G-box sequences in the context of the native promoter, the regulatory properties of chimeric reporter constructs, obtained by fusing G-box sequences to inactive truncated promoters, have been examined[9]. Interestingly, isolated multimers of either the G-box in the context of a palindromic octamer (CCACGTGG) or the G-box-like element characteristic of the *rab-16A* gene (motif 1, GTACGTGG) mediate distinct responses when fused to a truncated cauliflower mosaic virus (CaMV) promoter and assayed for activity in transgenic tobacco plants. Whereas the former sequence directed expression preferentially in roots, motif 1 directed a developmentally regulated expression pattern in seeds[9].

The diverse expression properties mediated by promoters containing identical G-box sequences clearly demonstrate that the functional properties of this element vary according to the promoter context in which it resides. How could the surrounding DNA sequences influence the property