doi:10.1016/S0022-2836(03)00720-4 J. Mol. Biol. (2003) 331, 89–99

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

A Protein–DNA Binding Mechanism Proceeds Through Multi-state or Two-state Parallel Pathways

Diego U. Ferreiro and Gonzalo de Prat-Gay*

Instituto Leloir, Facultad de Ciencias Exactas y Naturales Universidad de Buenos Aires and CONICET. Patricias Argentinas 435, (1405) Buenos Aires, Argentina

*Corresponding author

The DNA-binding mechanism of the dimeric C-terminal domain of the papillomavirus E2 protein with its specific DNA target was investigated and shown to proceed through two parallel pathways. A sequential multi-step reaction is initiated by the diffusion-controlled formation of an encounter complex, with no evidence of base sequence discrimination capacity. Following a substantial conformational rearrangement of the protein, a solvent exclusion step leading to the formation of a final protein–DNA complex was identified. This last step involves the largest burial of surface area from the interface and involves the consolidation of the direct readout of the DNA bases. Double-jump stopped-flow experiments allowed us to characterize the sequence of events and demonstrated that a fast-formed consolidated complex can take place through a parallel route. We present the simplest model for the overall mechanism with a description of all the intermediate species in energetic terms.

Keywords: DNA-binding; HPV-E2; protein–DNA; recognition; stopped-flow

Introduction

In the same way that unfolded proteins need not sample all possible conformations for each residue as challenged by the Levinthal paradox, gene regulatory proteins do not undergo a random three-dimensional diffusion process in order to reach their DNA targets.[1] A two-step mechanism has been proposed in which a first non-specific-binding step is followed by protein translocation events along the DNA chain such as inter-segment transfer, protein sliding or "hopping" through microscopic dissociation events and one-dimensional diffusion.[1–4]

As previously established for the *Escherichia coli* Lac repressor, DNA-binding rates can be faster than what it is expected for a diffusion-controlled reaction.[2] In addition to non-specific binding and facilitated translocation processes, association rate enhancements beyond diffusion limits can be explained by spatially arranged and complementary electrostatic fields that may funnel the positively charged DNA-binding domain into the negatively charged DNA polymer,[3,5] earlier described as "oriented diffusion".[2]

Several protein–DNA interactions have been described since then, and a general mechanism was proposed,[6] involving: (i) conformational changes in either protein or DNA, or non-specific binding of the protein to DNA, prior to the elementary bimolecular step, (ii) a bimolecular step where an initial complex is formed at the diffusion–collision rate or slower, depending on orientation effects, and (iii) local and/or global conformational changes in the encounter complex, leading to a consolidated and stable species.

A group of transcription regulatory proteins present segments within their DNA-binding domains that range from globally[7] to locally disordered, and that must undergo substantial folding coupled to DNA binding,[8] leading to complex binding mechanisms. Another group of proteins forms either monomers or weak dimers in solution, and dimerization can take place after an initial binding event by the monomer.[9]

What are the essential sequences of steps during and beyond the elementary bimolecular event that lead to the formation of a stable specific protein–DNA complex? A unique example of a detailed characterization of a protein–DNA recognition mechanism is the TATA box-binding protein (TBP) to DNA.[10] In this work, thermodynamic, kinetics and structural changes were analyzed along the reaction coordinate. TBP is a particular case of protein–DNA recognition where a substantial

Abbreviations used: EBNA, Epstein-Barr nuclear antigen 1; ANS, 1-anilino-8-naphtalene sulphonate.

E-mail address of the corresponding author: gpratgay@leloir.org.ar

deformation of the DNA takes place.[10,11] We now address this problem using the DNA-binding C-terminal domain of the E2 transcriptional regulator from human papillomavirus (HPV16-E2C)[12,13] and a small DNA duplex containing its target site, as a model protein–DNA recognition system. This domain forms a highly stable and intertwined dimeric β-barrel structure,[14] only shared by the EBNA1 C-terminal domain.[15] Although a short-lived non-native monomer can be found on its kinetic folding pathway,[16] HPV16-E2C forms a tight dimer with a dissociation constant of 1×10^{-12} M^{-1}.[17,18] HPV16-E2C is a stable dimer under our experimental conditions and the DNA target used is a small duplex, thus minimizing sliding or other coupled diffusion facilitating mechanisms, and simplifying the number of possible events to be analyzed. Using stopped-flow spectroscopic techniques, we describe the kinetic mechanism of interaction of this dimeric DNA-binding domain to specific and non-specific DNA targets. We show that this reaction can proceed through either multi-state or two-state parallel pathways, and describe the intermediates in energetic terms.

Results and Discussion

Two diffusion controlled bimolecular events in excess of protein

We used a fluorescein-modified site 35 18mer DNA duplex target previously described,[17] and carried out kinetic experiments of its interaction with E2C. Under pseudo first-order conditions with excess of E2C domain, a large FITC fluorescence change was observed in stopped-flow experiments, and the data fitted to a double exponential process (Figure 1(A)). Both observed rates are dependent on the protein concentration and thus correspond to bimolecular association events, with k_{on} of 0.6×10^9 M^{-1} s^{-1} and 1.4×10^9 M^{-1} s^{-1}, respectively (Figure 1(B)). These magnitudes are in the range of the calculated diffusion limit for an electrostatic assisted collision between macromolecules of these sizes.[6,19] In the presence of sucrose or glycerol, the association rates are slowed down with a magnitude expected for a diffusion-limited process for both bimolecular reactions (not shown).

The amplitude of the slower association phase, $k_{on}1$, disappears as the protein concentration is increased (Figure 1(C)) and, at the highest protein concentrations tested, the observed rate fits to a single exponential event (Figure 1(C)). This is indicative that a competition between reactants, and therefore reaction pathways, is taking place so that when DNA is limiting only the faster binding occurs. Since there is no evidence of higher order oligomers formed by E2C (even when tested up to millimolar concentrations), and dissociation to monomers is in the sub-picomolar range,[18] this result suggests that there is a pre-equilibrium of

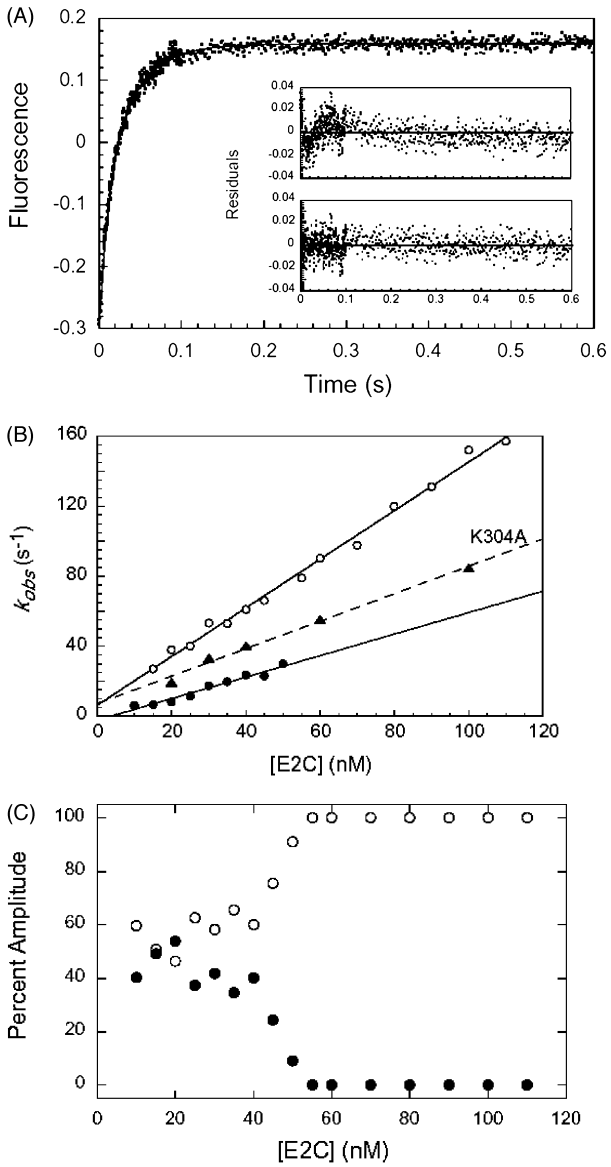

Figure 1. Kinetic association curves between E2C and its cognate site 35 DNA. (A). Typical stopped-flow fluorescence trace at 5 nM FITC-site 35 and 40 nM E2C. The line represents the best fit of the data points to the fluorescence changes. The top and bottom insets show the residuals to single or double-exponentials, respectively. (B). Pseudo first-order plot for the relaxation times of both phases: $k_{on}1$ (filled circles), $k_{on}2$ (empty circles). Above 60 nM [E2C], the data fit to one exponential. The filled triangles show the results obtained in a different experiment for the site-directed mutant (K304A), where only one binding event was detected (see the text). (C). Relative amplitudes corresponding to $k_{on}1$ (filled circles) and $k_{on}2$ (empty circles) shown in (B).

DNA-binding conformers in the dimeric domain, that might correspond to the two different "binding modes" described in the equilibrium-binding mechanism.[17]

The extrapolated off-rate for the faster-binding event, $k_{off}2$, is 6 s^{-1}, with an estimated encounter K_D of 5 nM. The disappearance of the slower bimolecular event precludes the determination of an

accurate $k_{off}1$, but it nevertheless appears to be below 1 s^{-1}. It should be pointed out that these off-rates represent the dissociation of an encounter complex and not the overall dissociation rate of a consolidated complex (see below).

When we analyzed a mutant protein K304A at the equilibrium, we found that only one "binding mode" was present under the same experimental conditions. This mutation is located at the C-terminal end of the highly dynamic DNA recognition helix, which suggests that this residue is somehow involved in the stabilization of the binding helix. (D.U.F. & G.P.G., unpublished results). If there is a correspondence between "binding modes", found at equilibrium binding experiments, and the conformers suggested, the K304A mutant should affect the equilibrium between conformers. Indeed, in a stopped-flow experiment under pseudo first-order conditions, only one bimolecular-binding event was observed (Figure 1(B), black triangles), with a second order rate constant of 0.7×10^9 M^{-1} s^{-1}, similar to the slowest of the binding events observed for the wild-type protein. Since the encounter rate is not affected by mutation, it appears that there is no sequence discrimination at this early event. A single-binding event is observed throughout the concentration range, suggesting that there is no competition between protein species as was observed for the wild-type protein. The fact that the mutant presents a single-binding event supports the hypothesis that an equilibrium between conformers is the basis for the two observed rates, and that this equilibrium could be shifted by a single mutation.

Kinetic binding to non-cognate and non-specific DNA sites

E2 proteins bind to the consensus palindromic sequence ACCGN4CGGT, where N4 is a four-base spacer, which is not directly contacted by E2,[14] although HPV16 E2C was shown to display a higher affinity for AT-rich spacers.[20,21] We have measured the binding affinity in solution of the HPV16 E2C domain for a bovine papillomavirus (site BOV) derived site used in crystallographic studies[14] containing a CG for AA replacement in the spacer region of the duplex and its affinity is three orders of magnitude lower than the cognate HPV16 site 35.[17] Stopped-flow experiments in excess of protein over the FITC-labeled BOV oligonucleotide, show a major bimolecular event with an association rate similar to that observed for the cognate site 35, and another minor and much slower bimolecular event (Table 1).

We investigated the kinetic-binding parameters of non-specific DNA duplexes of different characteristics. One of the duplexes, ISET35, contains the same base composition of site 35 but in a random sequence. When measured in a pseudo first-order experiment, two binding events were observed with rates similar to those of site BOV (Table 1). The fact that the $k_{on}2$ event is as fast as the specific DNA duplex suggests that the non-specific binding is also diffusion-controlled. The off-rate of one of these two events cannot be calculated accurately, but the faster associating event clearly shows an extremely fast off-rate of 107 s^{-1} (Table 1).

The other type of DNA duplexes that we investigated is the binding site of EBNA1, the origin-binding site of the Epstein-Barr nuclear antigen 1.[15] Under pseudo first-order conditions, two binding events were observed, and these are four orders of magnitude slower than those for the specific duplex described above (Table 1). Since E2 does not recognize any DNA sequence in this oligonucleotide,[17] and given that a non-specific random duplex binds very fast, the behavior towards the EBNA conserved site could be explained by orientation restrictions in this DNA duplex, something that would not occur in a flexible random sequence.[6] In addition, this sequence would not present the pre-curvature observed in E2 sites.[22]

Dissociation

In order to investigate the dissociation rate of the consolidated protein–DNA complex, we added E2C to FITC-modified site 35 DNA at a 1:1 ratio, and a fluorescence change was observed upon complex formation. After 20 minutes incubation, a 50-fold excess of non-fluoresceinated site 35 DNA was added, and a decrease in fluorescence to initial values indicated complete dissociation (Figure 2).

Table 1. Kinetic parameters characterizing the encounter complex formation and the dissociation events between E2C and different oligonucleotides

Oligo	$k_{on}2$ (M^{-1} s^{-1})	$k_{on}1$ (M^{-1} s^{-1})	k_{-1} (s^{-1})	$k_{off}2$ (s^{-1})	$k_{off}1$ (s^{-1}) [a]	$k_{off}2$ (s^{-1})[a]
Site 35[b]	1.4×10^9	0.6×10^9	6.3	ND	0.018	0.004
Site BOV[c]	1.2×10^9	6.0×10^5	ND	0.07	0.037	0.009
Site EBNA[d]	3.6×10^5	1.8×10^5	0.12	0.032	0.57	0.027
Iset35[e]	1.5×10^9	1.1×10^6	107	ND	1.73	0.34

[a] Measured from the dissociation of the consolidated complex. $k_{-3} = k_{off1}$.
[b] HPV-16 cognate: TCA ACCG ATTT CGGT TAC (recognition sequence underlined).
[c] BPV-1 cognate: CCG ACCG ACGT CGGT CGG (recognition sequence underlined).
[d] Epstein-Barr nuclear antigen 1 cognate: GGG TAG CAT ATG CTA CCC.
[e] Randomized site35: TAC TTG ACA GGT CCA TGT.

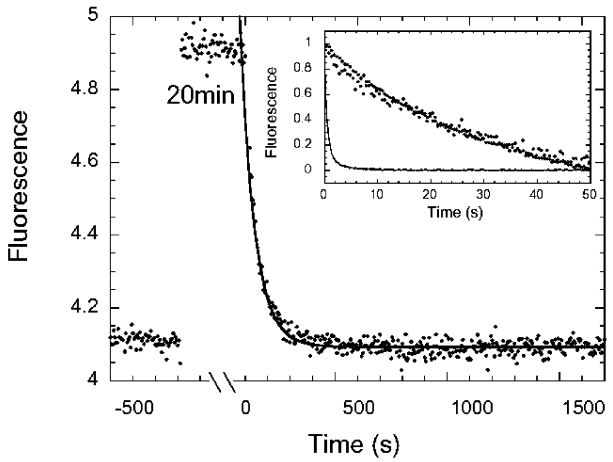

Figure 2. Dissociation kinetics between E2C and different DNA targets. 50 nM FITC-site 35 oligonucleotide was mixed with 50 nM E2C and, after 20 minutes, the complex was outcompeted with a 50-fold molar excess of unlabelled site 35 as the FITC fluorescence was recorded. Inset: stopped-flow traces of similar experiments conducted with different oligonucleotides: site BOV (filled circles); site EBNA (empty circles); iset35 (line). All the traces fitted to double-exponential decays, with the k_{off} rates listed in Table 1.

For a more accurate determination of the off-rates, in particular for the faster non-specific DNAs, the experiment was carried out using stopped-flow. The site 35 DNA dissociation curve fits to a double exponential decay, with a rate of 0.004 s^{-1}, corresponding to 70% of the amplitude and a 0.018 s^{-1} phase accounting for 30% of the amplitude (Table 1). Neither the rates nor the relative amplitudes of these phases are affected by varying the concentration of the non-fluoresceinated DNA (not shown). This fact rules out the possibility that the dissociation process under investigation is mediated by facilitated diffusion mechanisms such as displacement or intersegment transfer.[1] The dissociation events of all the DNAs tested exhibited biphasic behavior and the rates were greatly accelerated for the non-specific duplexes, EBNA1 and ISET, and only doubled for the BOV site. These results point at the fast rate of dissociation as the basis for the lower affinity of the non-specific sites. However, it is not possible to directly correlate the phases observed for the specific and non-specific oligonucleotides, since the changes in fluorescent signals are not necessarily the same.

A preferential-binding event in excess of DNA is followed by rearrangements

The binding of E2C to DNA can also be reported by changes in fluorescence of the tryptophan residues facing the center of the intertwined β-barrel.[17] We carried out a pseudo first-order experiment, now with the DNA in excess over the fixed E2C concentration, monitoring an increase in tryptophan fluorescence. The kinetic profiles conformed

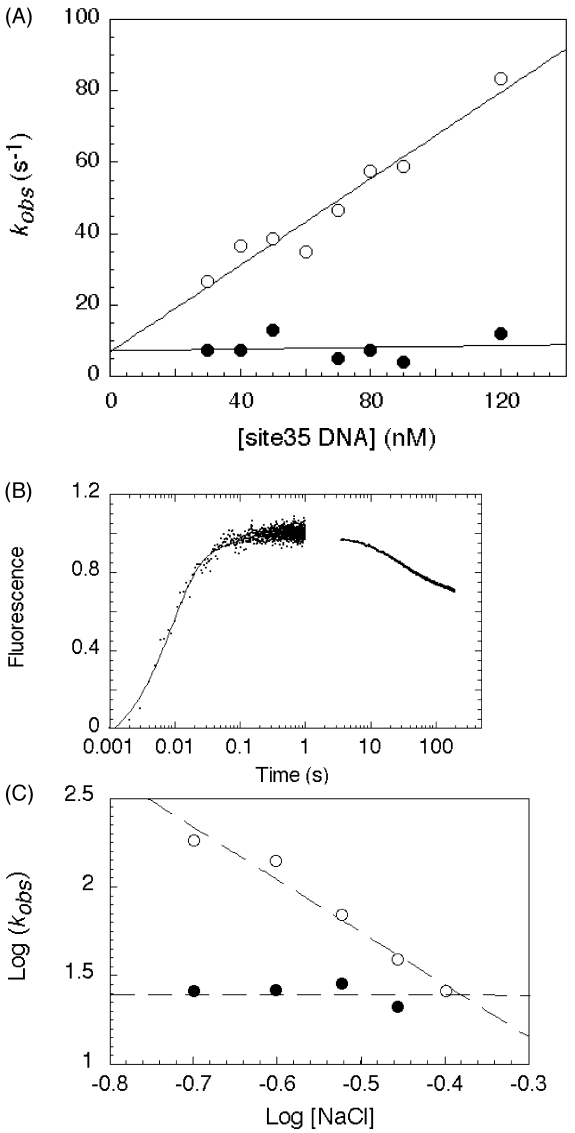

Figure 3. Kinetics of the association events in excess of DNA. (A). Pseudo first-order plot of the association between 25 nM E2C with different concentrations of site 35 DNA. The Trp fluorescence change was recorded and the data fitted to a double-exponential process: $k_{on}1$ (empty circles), k_2 (filled circles). (B). Fluorescence stopped-flow kinetic trace recorded in two time bases (1 second and 200 seconds) after mixing 300 nM E2c and 300 nM site35. (C). Double-log plot of the salt concentration dependence for the same rates described in (A) at 25 nM E2C and 70 nM site 35.

to a single concentration-dependent phase with a k_{on} of 0.6×10^9 M^{-1} s^{-1} (Figure 3(A)), in agreement with the slower binding mode ($k_{on}1$) observed in excess of protein (Figure 1(A)). In addition, a concentration-independent phase is observed, with a rate $k_2 = 8$ s^{-1} (see Figure 7 for overall rate constant nomenclature), which we assign to a conformational rearrangement of the protein molecule. The off-rate of the binding event cannot be obtained accurately enough from extrapolation (Figure 3(A)) but appears to be within experimental error of that obtained in the pseudo first-order

experiment in excess of protein. The faster binding event observed with FITC-DNA experiments might not be detected because of the lower sensitivity of this intrinsic probe as opposed to fluorescein.

The co-crystal structure of E2C and DNA shows that an extensive contact network is formed at the binding interface,[14,22] and it has been suggested that large conformational changes are prerequisite for tight binding.[23,24] It is unlikely that this complex interface is fully formed in less than 100 ms, the $t_{1/2}$ of the 8 s^{-1} phase. Searching for slower phases, we carried out experiments at higher concentrations of protein and DNA so as to accelerate and separate the bimolecular event. Effectively, the association and the first-order rearrangement of 8 s^{-1} are separable from a slower phase found in a different time range (Figure 3(B)). This additional phase of 0.04 s^{-1} does not depend on protein concentration and shows a decrease in fluorescence, as opposed to the two faster phases.

The single bimolecular event is highly dependent on salt concentration (Figure 3(C)). The slope of the double log plot indicates that four to five Na^+ counterions are displaced from the binding interface.[6] The first-order phase k_2 described is not affected by salt, suggesting that no exposed ionic interactions are involved in this event (Figure 3(C)).

The specific interaction pathway involves a solvent exclusion and conformational rearrangement step

The formation of a final consolidated protein–DNA complex must include a precise set of interactions together with solvent removal from the interface to a significant extent.[14,22] Since there is no direct means for observing such a phenomenon, we made use of the fluorescent probe 1-anilino-8-naphtalene sulphonate (ANS). E2C was shown to bind ANS[16] and bisANS in at least two sites, one of them being displaced by specific DNA.[24] Due to its small size, ANS could be potentially trapped at the encounter complex interface together with the solvent. E2C was mixed with near to saturation amounts of ANS and equilibrated, yielding an increase in ANS fluorescence, as expected. Upon mixing with site 35 DNA, a decrease in fluorescence was indeed observed, following a single exponential decay with a rate of 0.04($\pm$0.008) s^{-1}. This rate was independent of either ANS or protein concentration (not shown). Since the off-rate of ANS spontaneously dissociating from the protein is much slower (0.003 s^{-1}), we interpret that the observed rate reflects the exclusion of solvent from the interface. Pseudo first-order experiments in the presence of identical amounts of ANS show that the dye does not affect the on-rate of the elementary bimolecular step at all (not shown). One of the first-order phases reported by tryptophan fluorescence in kinetic-binding experiments with site 35 DNA has an identical rate (Figure 3(B)), strongly suggesting that a concomitant protein conformational change takes place together with the solvent exclusion. While the related site BOV DNA displaces the ANS in a similar time frame with a rate of 0.032($\pm$0.007) s^{-1}, the non-specific ISET and EBNA DNAs are not capable of displacing the dye, even at concentrations tenfold their K_Ds (not shown). These results indicate that, although non-specific DNAs are able to bind as fast as the specific duplex to form the encounter complexes, these complexes do not undergo a subsequent rearrangement involving solvent exclusion in the way the specific event does, in agreement with the proposed models of non-specific binding to DNA.[6,25]

A parallel pathway leading to a fast-formed consolidated E2C-DNA complex

In order to discriminate whether the observed dissociation events (Figure 2) are parallel or sequential reactions, and to obtain further information on the binding mechanism, we carried out a double-jump association-dissociation experiment. Using a stopped-flow fluorimeter, E2C was mixed with equimolar amounts of FITC site 35 DNA at 100 nM concentration, where the elementary bimolecular step has a half-life of 12 ms. At different delay times ranging from 0.01 seconds to 100 seconds, the complex was mixed with a 25-fold excess of unmodified DNA and the fluorescence change upon dissociation was monitored as in Figure 2. Surprisingly, we found that 70% of the amplitude corresponded to two slow phases of 0.0037 s^{-1} and 0.0011 s^{-1} (Figure 4(A)), highly coincident with the phases observed in the dissociation of the consolidated complex (Figure 3, Table 1). However, these phases are already present at 10 ms after the initial mixing of protein and DNA. While these amplitudes remain constant throughout the time range, two phases of 1 s^{-1} and 0.07 s^{-1} are transient and almost disappear at 2 seconds and 20 seconds, respectively (Figure 4(B)). We interpret that these phases are sequential, where the amplitude of the 1 s^{-1} phase disappears first and is highly coincident with the off-rate obtained from extrapolation in pseudo first-order experiments, strongly suggesting that it corresponds to the actual dissociation of the encounter complex. We assign the 0.07 s^{-1} rate to the reverse reaction of the first rearrangement observed in the pathway after the encounter complex. The presence of dissociation phases at the earliest delay times with rates coincident with those corresponding to dissociation of the stable complex, indicate that a significant amount of consolidated complex is formed without going through defined structural rearrangements, i.e. they proceed through a parallel reaction pathway. Furthermore, the fact that both phases show the same relative amplitudes throughout the time range indicates that the equilibrium between the final complexes is in fast exchange.

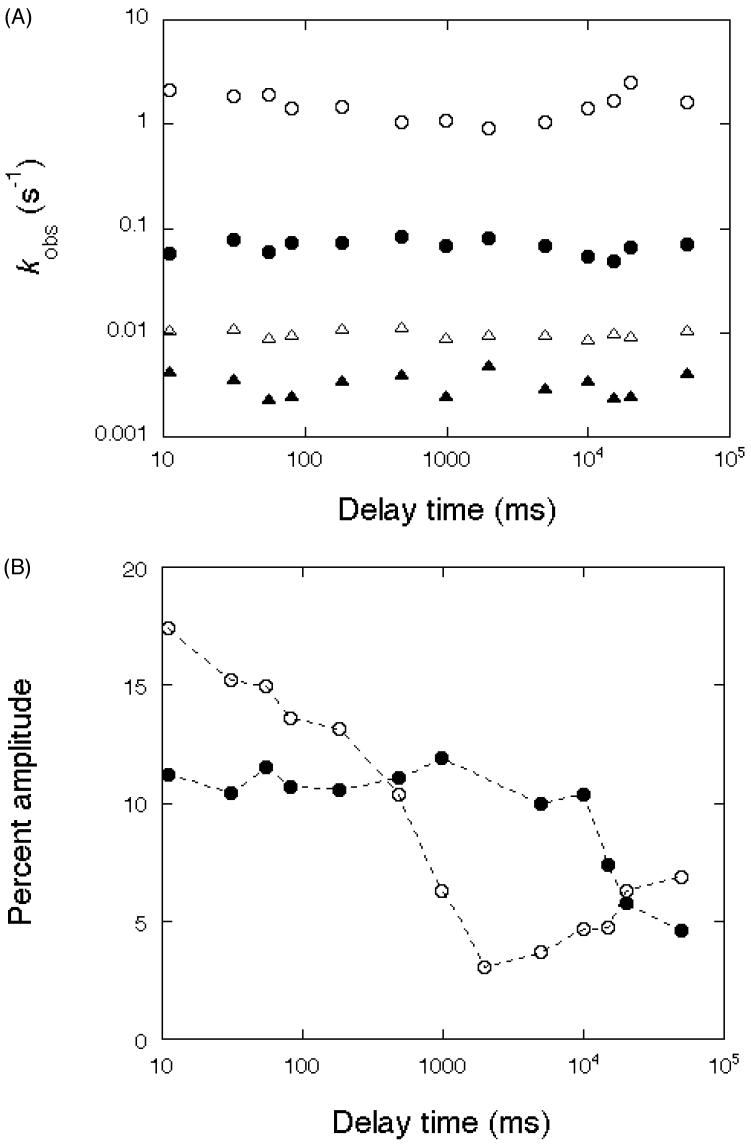

Figure 4. Double-jump association–dissociation experiment. 100 nM E2C was mixed with 100 nM FITC-site 35 and after the "delay time", the complex was competed with 2.5 μM site 35. After the second mix, dissociation fluorescence traces were collected at different time frames and the data fitted to exponential decays. (A). Recovered $k_{obs}1$, (empty circles), 2 (filled circles), 3 (empty triangles), 4 (filled triangles). (B). Corresponding amplitudes of phases 1 and 2. The amplitudes corresponding to phases 3 and 4 remained constant (see the text).

Similar double-jump association–dissociation experiments were performed with site BOV DNA. The resulting profiles show three phases: (i) a fast event with a $k_{obs} = 2\,s^{-1}$ that disappears with time, which we ascribe to the dissociation of the encounter complex E2C-site BOV; (ii) a phase with a rate of $0.008\,s^{-1}$ present throughout the time range, ascribable to the major dissociation phase of the consolidated E2C-site BOV complex (Table 1) which, as in the site 35 DNA, is present from the shortest delay time after mixing. A slow phase with a rate constant of $0.004\,s^{-1}$, coincident with one of the phases observed in the dissociation of the consolidated E2C-site BOV complex (Table 1), tends to disappear as the time of mixing progresses, suggesting that it corresponds to dissociation of a complex having undergone a conformational rearrangement after the encounter complex in the direction towards the formation of the final complex.

In the case of the non-specific iset35 DNA duplex, two phases were present from the earliest delay times with rates of $2.0\,s^{-1}$ and $0.5\,s^{-1}$, respectively, coincident with the phases described in the dissociation of the final complex (Table 1). Thus, even in the non-specific interaction, there are parallel pathways one of which passes through an encounter complex that rearranges and the other through the direct formation of a final complex, presumably because of a highly favorable conformation of both reactants, even if these conformations are loose, locally disordered or highly flexible.[26]

Temperature dependence of the encounter, rearrangement and dissociation events

Analysis of the temperature dependence of the rates of interaction between macromolecules is essential for the understanding of the nature of the steps involved in a kinetic mechanism.[6] We chose to monitor changes in tryptophan fluorescence in excess of DNA, since there is only one observable path. Pseudo first-order experiments at

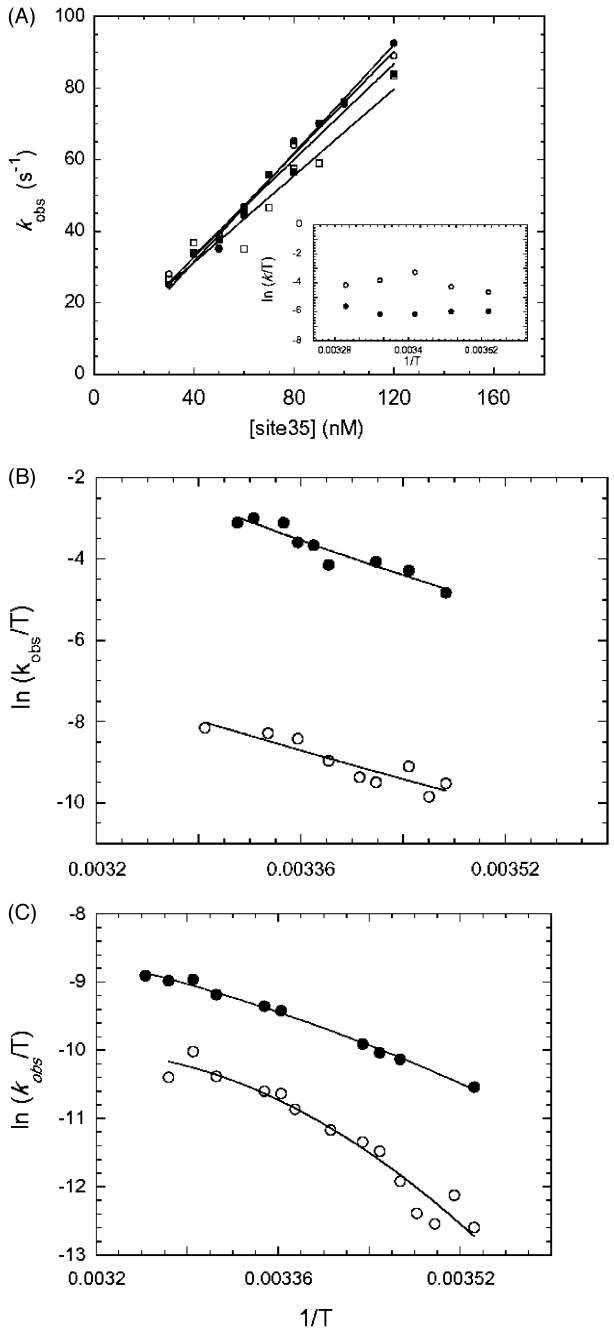

Figure 5. Temperature dependence of the association, rearrangement and dissociation reactions. (A). Encounter complex formation energetics probed by pseudo first-order curves at 25 nM fixed E2C concentration at different temperatures: 10 °C (filled circles), 15 °C (empty circles), 20 °C (filled squares), or 25 °C (empty squares). The binding was followed by the E2C Trp fluorescence change as in Figure 3. Inset: Eyring plot of the extracted $k_{on}1$ (filled circles) and k_{-1} (empty circles) for the encounter complex formation. (B). Eyring plots of the concentration-independent phases were conducted at 100 nM E2C and 300 nM site 35 between 10 °C and 30 °C; k_2 (filled circles); k_3 (empty circles). (C). Dissociation of the consolidated E2C–DNA complex was monitored as in Figure 2 at different temperatures. The data fitted to double-exponential decays through all the temperature range, $k_{off}1$ (filled circles), $k_{off}2$ (empty circles).

different temperatures (Figure 5(A)) show a negligible effect on either the on-rate or the off-rate corresponding to the encounter complex (Figure 5(A), inset). The lack of change precludes an accurate determination of thermodynamic activation parameters. In pseudo first-order conditions at a single DNA concentration in excess over E2C, temperature dependence data show a linear Eyring plot with a small activation energy for the encounter complex (Table 2). Thus, a negligible effect of temperature on the elementary association step is in concordance with an entropically driven encounter,[27] consistent with the counterion release evidence obtained (Figure 3(C)). Thus, temperature effects appear to be restricted to those related to effects on the diffusion properties of the macromolecules.[6] If one considers surface area as the unique source of $\Delta Cp^{\ddagger}$,[28,29] the lack of curvature in either reactions suggests a minimal change in the exposure of surface area to the solvent from either the free reactants or encounter complex to their transition state ensembles (TSE).

The rearrangement rates increase with increasing temperature, indicating positive activation energies (Figure 5(B), Table 2). The free energy of activation for both reactions is high (between 15 kcal mol^{-1} and 20 kcal mol^{-1}) which, compared with the overall binding free energy (−13.2 kcal mol^{-1}[17]), yields deep energetic wells for the intermediate steps. The activation energy of the first rearrangement (k_2), has both unfavorable enthalpic and favorable entropic contributions. The positive $\Delta S^{\ddagger}$ can arise from increased conformational freedom in the TSE, but needs to be further investigated. Similarly, at this stage, it is not possible to dissect the contribution of each macromolecule, protein or DNA, to the thermodynamic activation parameters. It has been shown that both E2 and its specific site present some degree of structural plasticity and each of them induces mutual conformational changes upon binding.[21,23,24]

The major contribution to the activation free energy of the second rearrangement, k_3, is enthalpic, which can be assigned to hydrogen bonding and van der Waals interactions formed at the interface. For these close contacts to be formed, hydration shells of both macromolecules must be re-ordered and, as the ANS-solvent displacement experiment showed, this occurs in the same time frame as k_3. Taken together, these results suggest that the "direct readout" of the DNA bases would be consolidated during this event.

There is no appreciable $\Delta Cp^{\ddagger}$ component in any of the rearrangement steps, something somehow unexpected, since a substantial negative $\Delta Cp^{\ddagger}$ was detected for E2C-site 35 interaction by calorimetric studies (1.5 kcal mol^{-1} K^{-1},[30]). This could indicate that the $\Delta Cp^{\ddagger}$ component is present in subsequent rearrangement processes that are not reported by a tryptophan fluorescence change. However, in the TBP-DNA-binding mechanism, ΔCp was shown to arise from conformational equilibria and no single step with non-zero ΔCp was evidenced.[11]

Table 2. Thermodynamic parameters for the activation energies extracted from the temperature dependence of the reaction rates

	$k_{on}1$	k_2	k_3	$k_{off}2$	$k_{off}1$
$\Delta G_0^{\ddagger}$ (kcal mol^{-1})	4.8 ± 1.1	16.1 ± 3.6	19.2 ± 3.2	19.6 ± 0.1	20.4 ± 0.5
$\Delta H_0^{\ddagger}$ (kcal mol^{-1})	1.5 ± 0.3	21.5 ± 2.6	18.0 ± 3.0	11.0 ± 1.3	15.4 ± 7
$\Delta S_0^{\ddagger}$ (cal K mol^{-1})	−11.1 ± 1	18.1 ± 2.5	−4.0 ± 1.2	−28.9 ± 1.3	−17.0 ± 7
$\Delta Cp^{\ddagger}$ (kcal mol^{-1} K^{-1})	ND	ND	ND	0.26 ± 0.1	0.98 ± 0.4

Alternatively, since we are probing the energetics in going from the free molecules in solution, through the intermediates and their TSE to the final complex, we must consider that the $\Delta Cp^{\ddagger}$ component might be evidenced in the reverse direction, the dissociation reaction.

To probe the temperature dependence of the dissociation kinetics we made use of FITC-modified site 35. At all the temperatures tested, two fluorescent events describing the parallel dissociation pathways were detected (Figure 5(C)). Both reactions have high activation free energies, in the same range of the forward k_3, providing evidence of being the rate-limiting step for both dissociation pathways (Table 2). The activation energies have entropic and enthalpic contributions, which is what is expected for the breakdown of the large interface contact network and the overall entropies related to the process. A large negative change in heat capacity was observed in one of the phases (Figure 5(C), Table 2), which confirms that the burial of surface area to the solvent takes place in the last rearrangement step, together with the exclusion of solvent from the interface. Clearly, a large $\Delta Cp^{\ddagger}$ change can only be observed in the dissociation reaction.

Conclusions

We carried out an initial dissection of the kinetic DNA-binding mechanism of a dimeric viral transcriptional regulator, the E2 C-terminal domain from human papillomavirus, as a model for a specific DNA-binding protein. In excess of protein, two parallel-binding events were detected upon binding either to specific or non-specific oligonucleotides. Since these events were shown to be diffusion-limited, we suggest that the protein exists in at least two different conformers prior to DNA binding. Supporting this hypothesis, a mutant protein that presented a single "binding mode" when tested under equilibrium conditions displayed only one kinetic event, suggesting that the equilibrium between conformers is linked to alternative "binding modes", and that this can be shifted by a single amino acid replacement. In the presence of excess DNA, only one binding event can be discerned, and two slower first-order phases were evidenced.

Double-jump association–dissociation experiments provide valuable information on various aspects of the mechanism. First, they indicate that both first-order phases are subsequent to an intermediate encounter complex. Second, they provide a direct measurement of the off-rate of the encounter complex. Third, and most strikingly, they show that a consolidated protein–DNA complex is already present at 10 ms after the initial interaction of the macromolecules. This constitutes direct evidence for two parallel pathways, one through an intermediate encounter complex that undergoes slow rearrangements, and the other that leads directly to the final complex in a diffusion-controlled manner. Since the respective fluorescence changes need not be proportional to the actual amount of molecules that proceed through each pathway, we cannot make an unequivocal estimate of the respective populations. Finally, the presence of the two phases in the dissociation of the consolidated complex at the earliest association times, strongly suggests that they are parallel dissociation events. Therefore, the overall mechanism appears to proceed through parallel pathways in both directions.

Two slow rearrangement phases evidenced by tryptophan fluorescence suggest that they involve substantial changes in the conformation of the protein, since the two tryptophan residues are buried in the central barrel constituting the dimer interface. The slower rate of 0.04 s^{-1} is highly coincident with the dissociation of trapped ANS from the DNA contacting region, indicative of solvent exclusion from the interface, as the final step in the consolidation of the stable protein–DNA complex. Figure 6 shows a scheme of the simplest model compatible with the results described.

Based on the early occurrence of the two dissociation phases observed we have proposed that they are parallel dissociation events, corresponding to two conformers of the consolidated complex undergoing fast exchange, suggesting that the solution geometries of the DNA-bound conformers are structurally related. However, more experiments are required to test whether one of the phases could be an intermediate for dissociation, in which case, even if there are conformer ensembles in the final complex, these would dissociate through a single route.

In addition, it should be noted that the two second-order phases observed in the pseudo first-order experiments probed by FITC-modified DNA could correspond to the two parallel association reactions, one leading to the intermediate encounter complex and the other to a fast-formed final complex. A prerequisite of this fast-formed complex is

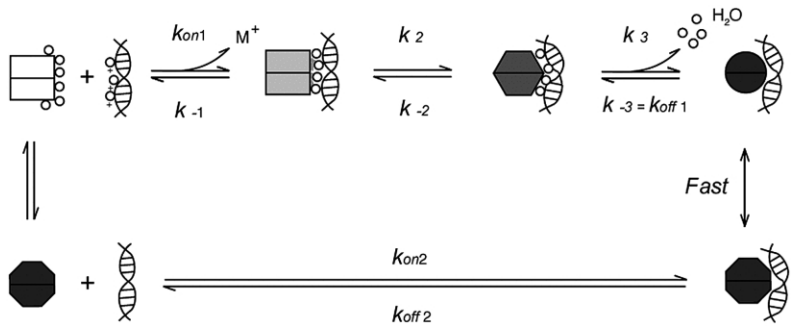

Figure 6. E2C–DNA interaction scheme. E2C exists in two populations of conformers, each of which binds DNA through different pathways. Rate constant values are listed in Table 1.

a population of highly favorable conformational state of the binding partners. The fast route either does not produce a tryptophan fluorescence change or both second-order events cannot be discriminated by this probe. The $k_{off}2/k_{on}2$ ratio yields an estimated K_D of 1.5×10^{-10} M, in excellent agreement with the equilibrium K_D (2×10^{-10} M),[17] further supporting the proposition that it is a consolidated complex. Further, the estimation of the K_D from the product of the rate constants of the multi-step pathway ($k_{-1}k_{-2}k_{-3}/k_{on1}k_2k_3$) yields a value of 0.04 nM compared with 0.2 nM obtained from equilibrium, which is an astonishing agreement considering the nature of the reaction pathway involved.

The inability of the non-specific duplexes to displace ANS from the interface suggests that more solvent remains in non-specific complexes, and these do not undergo conformational rearrangements and docking as the specific complex does. As was earlier shown for the *Eco* RI-DNA interaction,[31] water is differentially sequestered at the interface between specific or non-specific complexes, providing an additional source of free energy for specific recognition.[32]

From the various kinetic and thermodynamic data obtained in this and in previous work,[17] we can start building a picture of the energetics of the different steps of the specific multi-state mechanism (Figure 7). The initial barrier appears to be entropic, judged by the negative $\Delta S^{\ddagger}$. However, since a positive ΔS should be expected for counterion release as the unique entropy source, it is clear that there are other contributions that outbalance this effect yielding the negative ΔS for the TSE. We know that there is no major solvent release at this stage, therefore, a possible source is the entropy loss of the two dynamic interaction surfaces from both macromolecules. A more unlikely explanation is that the counterion release takes place after the initial collision TSE. The encounter complex is stabilized by electrostatic interactions and, remarkably, the energetic difference for this stage is almost the same for all the oligonucleotides tested. This suggests that "sliding" of the transcription factor along the DNA molecule could be occurring over this isopotential energetic surface. If the oligonucleotide contains an E2 site, the binding proceeds through the conformational changes described. Finally, as the solvent is excluded from

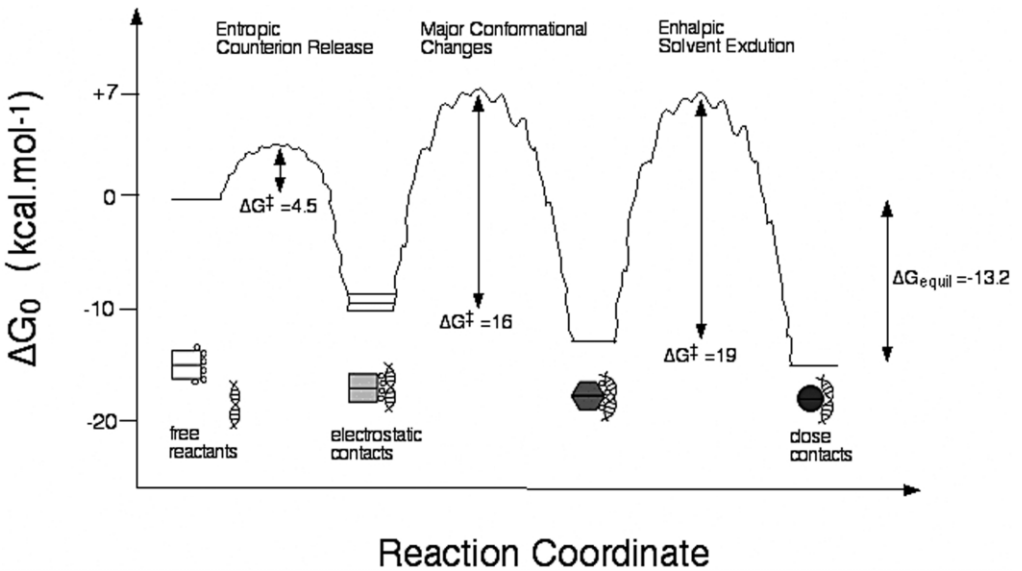

Figure 7. Graphical representation of the thermodynamic changes associated with E2C–DNA specific interaction through the two-intermediate pathway. The values correspond to 25 °C and 1 M standard state for free energy of activation, listed in Table 2. The free energy levels where deduced from equilibrium and kinetic measurements. For the encounter complex, the lines represent the free energy levels for specific and non-specific oligonucleotides.

the interface, the "direct readout" of the DNA bases and the amino acid side-chains takes place.

There appears to be little if any curvature on the Eyring plots, suggesting a small $\Delta Cp^{\ddagger}$ component and thus a minimal change in exposure of surface area, and solvent appears to remain throughout the initial steps. The burial of solvent-exposed area is triggered during the last rearrangement step, and is only detected through the temperature dependence of the reverse rate ($k_{off1} = k_{-3}$).

Protein–DNA interactions involve several coupled processes and, by using a fairly simple system, we were able to begin answering general mechanistic and energetic questions often quite difficult to address in more complex systems. We were able to determine a sequential set of steps and also demonstrated that, alternatively, a protein–DNA recognition reaction can proceed through a parallel pathway leading to a rapidly formed consolidated complex. Further mutagenesis analysis and molecular dynamics simulations will contribute to the dissection of such an elementary reaction in the biological world.

Materials and Methods

Chemicals, protein and DNA

All solutions were prepared with distilled and deionized (Milli-Q plus) water. All chemicals were of analytical grade, purchased from ICN (Aurora, Ohio, USA). All experiments were carried out in 25 mM BisTris–HCl buffer (pH 7.0), 200 mM NaCl, 1 mM DTT, and filtered through 0.45 μm membranes prior to usage. The 81 amino acid residue carboxyl-terminal domain of the HPV-16 E2 was recombinantly expressed, purified and stocked as described.[17,18] Protein concentration was determined spectrophotometrically using a molar extinction coefficient $\varepsilon_{280} = 41,920\ M^{-1}\ cm^{-1}$ (dimer). All synthetic oligonucleotides were purchased from IDT (Coranville, CT, USA). Where stated, an FITC moiety was introduced by a six-carbon linker to the 5′ end of one of the strands during synthesis. Double-stranded oligonucleotides where annealed as described.[17]

Stopped-flow kinetics

All fluorescence stopped-flow kinetic experiments were performed using a SX.18MV stopped-flow apparatus (Applied Photophysics, Leatherhead, UK). All concentrations reported are those resulting from mixing equal volumes of each syringe at 25(±0.1) °C, unless otherwise stated. The reactions were monitored using the fluorescence of the FITC moiety of the modified oligos or the Trp fluorescence of the protein. Excitation was set to 490 nm or 280 nm and the total fluorescence was collected through cut-off filters (Schott, PA, USA). Five to ten kinetic traces were collected and averaged for each concentration point. The data were fitted to extract the rates and amplitudes using non-linear least-squares fitting software provided by the manufacturer, using single or double-exponential equations. Double-jump association–dissociation experiments were performed mixing a 0.5:0.5 volume of protein and FITC-DNA and, after the "delay time", the reaction sample was subsequently mixed with a volume of non-fluoresceinated DNA, and the fluorescent decay was measured.

Thermodynamic analysis

Thermodynamic parameters for the TSE corresponding to the initial reaction steps were extracted from pseudo first-order curves at different temperatures. The TSE energetics of the subsequent intermediates were done at 100 nM E2C and 300 nM DNA between 10 °C and 30 °C. Thermodynamic parameters were extracted from Eyring plots (ln (k_{obs}/T) *versus* $1/T$), fitting the data points to the equations previously described.[33]

Acknowledgements

D.U.F. holds a PhD scholarship from CONICET. This work was supported by the Wellcome Trust, CRIG OIA U41 RG27994. The authors thank the support of ANPCyT, Fundación Bunge y Born and Fundación Antorchas. We are indebted to Cathy Royer for helpful criticisms to the manuscript.

References

1. Berg, O. G., Winter, R. B. & von Hippel, P. H. (1981). Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry*, **20**, 6929–6948.
2. Riggs, A. D., Bourgeois, S. & Cohn, M. (1970). The lac repressor–operator interaction. 3. Kinetic studies. *J. Mol. Biol.* **53**, 401–417.
3. von Hippel, P. H. & Berg, O. G. (1989). Facilitated target location in biological systems. *J. Biol. Chem.* **264**, 675–678.
4. Winter, R. B., Berg, O. G. & von Hippel, P. H. (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.
5. Sharp, K., Fine, R. & Honig, B. (1987). Computer simulations of the diffusion of a substrate to an active site of an enzyme. *Science*, **236**, 1460–1463.
6. Record, M. T., Jr, Ha, J. H. & Fisher, M. A. (1991). Analysis of equilibrium and kinetic measurements to determine thermodynamic origins of stability and specificity and mechanism of formation of site-specific complexes between proteins and helical DNA. *Methods Enzymol.* **208**, 291–343.
7. Wright, P. E. & Dyson, H. J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293**, 321–331.
8. Spolar, R. S. & Record, M. T., Jr (1994). Coupling of local folding to site-specific binding of proteins to DNA. *Science*, **263**, 777–784.
9. Kohler, J. J. & Schepartz, A. (2001). Kinetic studies of Fos·Jun·DNA complex formation: DNA binding prior to dimerization. *Biochemistry*, **40**, 130–142.
10. Powell, R. M., Parkhurst, K. M., Brenowitz, M. & Parkhurst, L. J. (2001). Marked stepwise differences within a common kinetic mechanism characterize TATA-binding protein interactions with two consensus promoters. *J. Biol. Chem.* **276**, 29782–29791.
11. Powell, R. M., Parkhurst, K. M. & Parkhurst, L. J.

(2002). Comparison of TATA-binding protein recognition of a variant and consensus DNA promoters. *J. Biol. Chem.* **277**, 7776–7784.
12. Hegde, R. S. (2002). The papillomavirus E2 proteins: structure, function, and biology. *Annu. Rev. Biophys. Biomol. Struct.* **31**, 343–360.
13. McBride, A., Romanczuk, H. & Howley, P. (1991). The papillomavirus E2 regulatory proteins. *J. Biol. Chem.* **266**, 18411–18414.
14. Hegde, R., Grossman, S., Laimins, L. & Sigler, P. (1992). Crystal structure at 1.7 Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. *Nature*, **359**, 505–512.
15. Bochkarev, A., Barwell, J. A., Pfuetzner, R. A., Bochkareva, E., Frappier, L. & Edwards, A. M. (1996). Crystal structure of the DNA-binding domain of the Epstein-Barr virus origin-binding protein EBNA1, bound to DNA. *Cell*, **84**, 791–800.
16. Mok, Y. K., Bycroft, M. & de Prat-Gay, G. (1996). The dimeric DNA binding domain of the human papillomavirus E2 protein folds through a monomeric intermediate which cannot be native-like. *Nature Struct. Biol.* **3**, 711–717.
17. Ferreiro, D. U., Lima, L. M., Nadra, A. D., Alonso, L. G., Goldbaum, F. A. & de Prat-Gay, G. (2000). Distinctive cognate sequence discrimination, bound DNA conformation, and binding modes in the E2 C-terminal domains from prototype human and bovine papillomaviruses. *Biochemistry*, **39**, 14692–14701.
18. Mok, Y. K., de Prat Gay, G., Butler, P. J. & Bycroft, M. (1996). Equilibrium dissociation and unfolding of the dimeric human papillomavirus strain-16 E2 DNA-binding domain. *Protein Sci.* **5**, 310–319.
19. Goeddel, D. V., Yansura, D. G. & Caruthers, M. H. (1977). Binding of synthetic lactose operator DNAs to lactose represessors. *Proc. Natl Acad. Sci. USA*, **74**, 3292–3296.
20. Bedrosian, C. L. & Bastia, D. (1990). The DNA-binding domain of HPV-16 E2 protein interaction with the viral enhancer: protein-induced DNA bending and role of the nonconserved core sequence in binding site affinity. *Virology*, **174**, 557–575.
21. Hines, C. S., Meghoo, C., Shetty, S., Biburger, M., Brenowitz, M. & Hegde, R. S. (1998). DNA structure and flexibility in the sequence-specific binding of papillomavirus E2 proteins. *J. Mol. Biol.* **276**, 809–818.
22. Kim, S. S., Tam, J. K., Wang, A. F. & Hegde, R. S. (2000). The structural basis of DNA target discrimination by papillomavirus E2 proteins. *J. Biol. Chem.* **275**, 31245–31254.
23. Hegde, R. S., Wang, A. F., Kim, S. S. & Schapira, M. (1998). Subunit rearrangement accompanies sequence-specific DNA binding by the bovine papillomavirus-1 E2 protein. *J. Mol. Biol.* **276**, 797–808.
24. Lima, L. M. & de Prat-Gay, G. (1997). Conformational changes and stabilization induced by ligand binding in the DNA-binding domain of the E2 protein from human papillomavirus. *J. Biol. Chem.* **272**, 19295–19303.
25. Jen-Jacobson, L., Engler, L. E. & Jacobson, L. A. (2000). Structural and thermodynamic strategies for site-specific DNA binding proteins. *Struct. Fold. Des.* **8**, 1015–1023.
26. Shoemaker, B. A., Portman, J. J. & Wolynes, P. G. (2000). Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc. Natl Acad. Sci. USA*, **97**, 8868–8873.
27. Oda, M. & Nakamura, H. (2000). Thermodynamic and kinetic analyses for understanding sequence-specific DNA recognition. *Genes Cells*, **5**, 319–326.
28. Livingstone, J. R., Spolar, R. S. & Record, M. T., Jr (1991). Contribution to the thermodynamics of protein folding from the reduction in water-accessible nonpolar surface area. *Biochemistry*, **30**, 4237–4244.
29. Sturtevant, J. M. (1977). Heat capacity and entropy changes in processes involving proteins. *Proc. Natl Acad. Sci. USA*, **74**, 2236–2240.
30. Di Pietro, S. M., Centeno, J. M., Cerutti, M. L., Lodeiro, M. F., Ferreiro, D. U., Alonso, L. G. *et al.* (2003). Specific antibody-DNA interaction: a novel strategy for tight DNA recognition. *Biochemistry*, **42**, 6218–6227.
31. Sidorova, N. Y. & Rau, D. C. (1996). Differences in water release for the binding of EcoRI to specific and nonspecific DNA sequences. *Proc. Natl Acad. Sci. USA*, **93**, 12272–12277.
32. Sidorova, N. Y. & Rau, D. C. (2001). Linkage of EcoRI dissociation from its specific DNA recognition site to water activity, salt concentration, and pH: separating their roles in specific and non-specific binding. *J. Mol. Biol.* **310**, 801–816.
33. Chen, X. & Matthews, C. R. (1994). Thermodynamic properties of the transition state for the rate-limiting step in the folding of the alpha subunit of tryptophan synthase. *Biochemistry*, **33**, 6356–6362.

Edited by P. Wright

(Received 21 February 2003; received in revised form 4 June 2003; accepted 4 June 2003)