Nucleic Acids Research, 1997, Vol. 25, No. 11 **2153–2160**

DNA binding and phasing analyses of Tn*5* transposase and a monomeric variant

Dona York and William S. Reznikoff*

Department of Biochemistry, 420 Henry Mall, University of Wisconsin–Madison, Madison, WI 53706, USA

Received January 21, 1997; Revised and Accepted April 10, 1997

ABSTRACT

Both full-length Tn*5* transposase and a COOH-terminal truncated monomeric form of the protein, Δ369, have been shown to specifically bind end sequences at comparable affinities. In addition, both proteins distort the target sequence in a similar manner, as determined by a circular permutation assay. In this study, ΔEK54, a derivative of Δ369 with a single amino acid substitution that significantly enhances binding activity, is used in further binding and bending studies along with full-length transposase. Phasing analysis has shown that distortion of the end sequences upon binding of full-length transposase and ΔEK54 protein is due in part to a protein-induced bend oriented towards the major groove. Because the center of transposase-induced bending maps to the extreme leftward end of the 19 bp consensus sequence, we examined the possibility that optimal protein binding requires additional upstream nucleotide contacts. Experiments presented here show that 9–10 nucleotides are needed upstream of +1 of the 19 bp sequence for efficient binding and this requirement can be met by either single-stranded or double-stranded DNA.

INTRODUCTION

The Tn*5* transposon is a composite transposon that contains two nearly identical insertion sequences, IS*50*R and IS*50*L, that are in inverse orientation to each other (Fig. 1). Each of these IS elements is capable of independent transposition. Three antibiotic resistance genes are centrally located with respect to IS*50*R and IS*50*L. Each IS*50* has two unique 19 bp end sequences that are in inverse orientation: outside end (OE) and inside end (IE) (1,2). Although the OE and IE differ in a number of nucleotide positions, both are specifically recognized by the Tn*5* transposase (Tnp) protein (3, for a review see 4). IS*50*R encodes P1 or Tnp and P2 or the inhibitor (Inh). The Inh protein is translated from the same reading frame as Tnp but lacks the first 55 amino acids (5). This protein is defective in transposition (6). Tnp has two opposing roles in transposition. As a *cis*-acting protein, Tnp is able to catalyze the transposition reaction. In contrast, Tnp acts primarily as an inhibitor of transposition *in trans* (7,8). IS*50*L encodes two nonfunctional proteins, P3 and P4 (9).

Although a number of host proteins have been proposed to play an auxiliary role in Tn*5* transposition, both the end sequences and the Tn*5* transposase are fundamental to the transposition process and play multifunctional roles. Two different segments within the 19 bp end sequences have been shown to be essential for wild type (Wt) Tnp binding while other nucleotides have been implicated in subsequent steps of the transposition reaction (3). The functional domains of the Tn*5* Tnp have been outlined based on mutational and deletion analyses. The specific DNA-binding domain lies in the NH_2-terminus of the protein. The Inh protein is unable to specifically recognize the end sequences (10). In addition, missense mutations in this region also eliminate Tnp binding (10,11). Furthermore, a number of hypertransposition mutants have been isolated that are a result of single amino acid substitutions in the NH_2-terminus of the protein (Zhou and Reznikoff, submitted). This hyperactive phenotype has been shown to be due in part to an increased OE DNA-binding affinity. The central portion of the Tnp protein shares amino acid sequence conservation with other members of the IS4 family of transposases (12). Members of the IS4 family contain two conserved regions which correspond to a domain that may be the active site for the cleavage and strand transfer reactions (13,14). A dimerization domain has been proposed based on deletion analysis of the COOH-terminus (10). Truncated forms of Tnp lacking 89 (Δ387) or 107 (Δ369) amino acids from the COOH-terminus display an altered ability to dimerize based on migration patterns in gel retardation assays. While Δ387 has been shown to bind the OE in both dimeric and monomeric forms (10), Δ369 protein binds in a monomeric form exclusively (10,15). Moreover, a single amino acid substitution at position 372 (Leu to Pro) also demonstrates the importance of this region in dimerization (16). Together these data suggest that major determinants for dimerization lie between residues 369 and 387.

A previous comparison of DNA-binding properties of Δ369 and full-length Tnp has demonstrated that binding affinities of the two proteins are similar (15). In addition, both proteins are able to severely distort the OE upon binding as determined by a circular permutation assay and the center of this distortion lies near the first nucleotide of the 19 bp consensus sequence. Two questions have emerged from these results and are addressed in this study: (i) what is the exact nature of this Tnp-induced distortion of the OE; and (ii) are additional nucleotide contacts needed upstream of +1 of the 19 bp OE sequence for optimal binding of Tnp? Using full-length Tnp and a monomeric variant of Tnp (ΔEK54) with enhanced

*To whom correspondence should be addressed. Tel: +1 608 262 3608; Fax: +1 608 262 3453; Email: reznikoff@biochem.wisc.edu

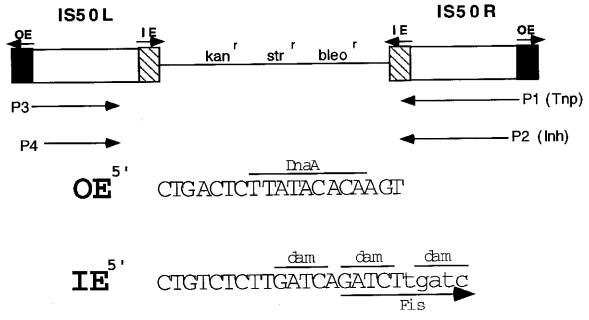

Figure 1. Tn*5* transposon. The top shows the organization of the Tn*5* transposon. The IS*50* elements, each contain an outside end (OE) and inside end (IE) whose orientations are indicated by arrows. Two nonfunctional protein products produced by IS*50*L (P3 and P4) are shown. The functional transposase (P1) and the inhibitor protein (P2) are indicated by arrows. The bottom of the figure shows the 19 bp sequences of the OE and IE. Consensus sequences of host protein binding sites within each sequence have been indicated by lines or arrows.

binding abilities (Zhou and Reznikoff, submitted), a circular permutation assay was performed with the OE followed by a phasing analysis. This analysis has shown that both proteins distort the OE similarly and this distortion is a result in part of a directed bend towards the major groove. The bending angle caused by Tnp binding differs for the two proteins by 12°. Because the center of this bend maps near the first nucleotide in the end consensus sequence, we investigated the possibility that additional nucleotide contacts are needed for optimal protein binding to the OE. This examination revealed that 9–10 nt are needed upstream of +1 while no additional contacts downstream of the 19 bp OE sequence are needed for optimal binding. Furthermore, these additional required upstream contacts can be met by either single-stranded or double-stranded DNA.

MATERIALS AND METHODS

Strains and plasmids

Escherichia coli strains DH5α and BL21 (DE3) pLysS were used for plasmid isolation and transposase overexpression, respectively. Plasmids pBR322 (17,18), pRZ9000 EK54 MA56 (Zhou and Reznikoff, submitted) and pRZ9012 (3) have been described previously.

Oligonucleotides

The following oligonucleotides (purchased from Research Genetics Inc.) were used in the construction of the OE phasing plasmids:

AT20 5′ CTAGACAAAAACGGGCAAAAACGGGCAAAAAGGTAC
3′ TGTTTTTGCCCGTTTTTGCCCGTTTTTC
*Age*I

AT22 5′ CTAGACAAAAACGGGCAAAAACGGGCAAAA**ACCGGT**AC
3′ TGTTTTTGCCCGTTTTTGCCCGTTTTTGGC

AT28 5′ CTAGACAAAAACGGGCAAAAACGGGCAAAAACGCGTGCCGGTAC
3′ TGTTTTTGCCCGTTTTTGCCCGTTTTTGCCCGCGGC

AT30 5′ CTAGACAAAAACGGGCAAAAACGGGCAAAAACCGGCCGGCCGGTAC
3′ TGTTTTTGCCCGTTTTTGCCCGTTTTTGGCCGGCCGGC

The following oligonucleotides (purchased from Research Genetics Inc.) were used in gel retardation and the single-stranded adjacent DNA assays. The 19 bp OE sequence is underlined. The single line indicates that the sequence is read from +1 to +19 in the 5′ to 3′ direction.

5′ TTCGAGCTCGGTACC CTGACTCTTATACACAAGT AGCGGATCCGCATGCA
3′ AAGCTCGAGCCATGG GACTGAGAATATGTGTTCA TCGCCTAGGCGTACGT

5′ GCTCGGTACC CTGACTCTTATACACAAGT
3′ CGAGCCATGG GACTGAGAATATGTGTTCA

5′ CTGACTCTTATACACAAGT
3′ GACTGAGAATATGTGTTCA

The following oligonucleotides were synthesized at the Biotechnology center, UW-Madison. The brackets indicate where aliquots were removed before the next base was added. Up refers to upstream of +1 of the consensus sequences and down refers to downstream of +19. The OE consensus sequence is underlined. A single line indicates that the consensus sequence of the OE and IE (from +1 to +19) is read in the 5′ to 3′ direction while a double line indicates the consensus is read +19 to +1 in the same direction.

OE_{up} 5′{TTCGAGCTCGGTACC C}TGACTCTTATACACAAGT AGCGGATCC-GCATGCA
OE_{down} 5′{TTGTGTATAAGA}GTCAG GGTACCGAGCTCGAA

The following are the primers (purchased at Research Genetics Inc.) used to extend the oligonucleotides synthesized above.

(A) 5′ TTCGAGCTCGGTACC; (B) 5′ TGCATGCGGATCCGCT

DNA construction

*Kpn*I/*Acc*65I

A 52 bp fragment : 5′AATTCGAGCTC**GGTACC**CTGACTCTTATACACAAG-TAGCGGATCCGCATGCA containing the OE (underlined, read +1 to +19 in the 5′ to 3′ direction) was cloned into the *Eco*RI/*Hin*dIII site of a derivative of pBR322 (cut with *Eco*RV/*Bsa*AI and religated). The *Eco*RI site of the pBR322 OE plasmid was subsequently changed to an *Xba*I site by a fill-in reaction (Klenow) followed by ligation of a linker containing an *Xba*I site. The *Hin*dIII site was changed to an *Nhe*I site by filling in (Klenow) and religating. This *Nhe*I site was subsequently changed to a *Sal*I site by again filling in with Klenow and ligating in a *Sal*I linker. This plasmid, designated pRZ9023, was used to construct the six phasing plasmids. pRZ9023 AT20, AT22, AT28 and AT30 were constructed by subcloning the oligonucleotides shown above into the *Xba*I/*Kpn*I site of pRZ9023. pRZ9023 AT20 was digested with *Acc*65I (bold), filled-in with Klenow and religated to create pRZ9023 AT24. pRZ9023 AT26 was constructed by an *Age*I digestion of pRZ9023 AT22 (see above) followed by a fill-in reaction and ligation. The *Xba*I/*Sal*I fragment from all six of these constructs was subsequently cloned into the *Xba*I/*Sal*I site of pBend 3.

Purification

ΔEK54 MA56 protein (lacking Inh protein) was purified as described previously (15). The homogeneity of purified ΔEK54 was determined to be ~85% by a coomassie-stained SDS–PAGE gel (not shown). Purified full-length Tnp MA56 (lacking Inh protein) was a gift from Maggie Zhou.

Gel retardation assay

The binding affinity of each protein was tested using annealed complementary 50 base oligonucleotides described above. Five micrograms of each complementary oligonucleotide, in a 100 µl volume, was heated to 70°C for 10 min and then cooled slowly to room temperature. Annealed oligonucleotides were labeled with γ-^{32}P in the presence of kinase. An aliquot of 0.5 ng of DNA was incubated in the presence of increasing amounts of purified full-length and ΔEK54 protein (0, 5, 25, 100, 400 and 800 nM) under conditions described previously (15). Reactions were electrophoresed on an 8% (29:1) nondenaturing 0.5× TBE polyacrylamide gel. Gels were run at 300 V for 2 h at 4°C, dried and exposed to film. The percentage of free and bound DNA was quantified using a Molecular Dynamics PhosphorImager.

Circular permutation assay

Gel retardation assays were performed with six circularly permutated 182 bp fragments generated from six different restriction enzyme digests (Fig. 3A) of pRZ9012 (3) a pBend 3 derivative containing the 19 bp OE consensus sequence. Each isolated restriction fragment was dephosphorylated with calf intestinal phosphatase and end-labeled with ^{32}P. Binding conditions were as described previously (3).

For data analyses, small variations in probe mobility were corrected for by dividing complex mobility by free probe mobility. All mobilities were then normalized to the fastest migrating complex and plotted as a function of the center position of each probe. The points are connected by a best fit of a cosine function (SigmaPlot). The centers of DNA distortion were determined as the points of maximum mobility from the best fit curve. The distortion angles (α_D) were determined from the formula:

$$A_{CP} = 1 - \cos(k\,\alpha_D/2) \text{ (ref. 19)}$$

A_{CP}, amplitude of the circular permutation function, reflects the difference between the relative mobility of the complexes with the minimum and maximum mobilities. *k* is a coefficient to adjust for electrophoretic conditions (temperature, gel composition and field strength). A set of standards containing two to six phased A–T tracts, obtained from Thompson and Landy (20; pJT170–*n*; *n* = 2–6), were electrophoresed under our assay conditions. From a best fit cosine function of mobility variations induced by the intrinsic bend standards as a function of bend angles (18° per A–T tract), a *k* value of 1.06 was obtained for our conditions.

Phasing analysis

The phasing analysis vectors: pRZ9023 AT20, AT22, AT24, AT26, AT28 and AT30 were constructed as described above. Each vector was digested with *Eco*RV, the fragment isolated (195–205 bp), dephosphorylated and end-labeled with ^{32}P. Conditions for gel retardation assays were as described for the circular permutation assay.

As with the data accumulated from the circular permutation assays, complex mobilities were corrected for variations in probe mobilities. The resulting relative mobilities of the complexes were then normalized to the average mobility of all the complexes and plotted as a function of the number of base pairs from the center of the Tnp-induced distortion (determined in the circular permutation assay) to the center of the intrinsic bend. Points were connected by a best fit of a cosine function using Sigma Plot. Estimation of the directed bend angle (α_B) was measured according to the formula:

$$\tan(k\alpha_B/2) = A_{PH}/2\tan(k\alpha_C/2) \text{ (ref. 19)}$$

The amplitude of the phasing function, A_{PH}, represents the difference in the relative mobilities of the slowest migrating species (in-phase) and fastest migrating species (out-of-phase) from the best fit curve. α_C represents the angle of the intrinsic bend which is 54° (three phased A–T tracts).

Minimal binding site determination

An oligonucleotide was synthesized (3′–5′) containing the OE sequence. This synthesis was interrupted at every base beyond +1 (see OE_{up} above) and an aliquot removed. This resulted in 15 oligonucleotides that differ only in the number of bases to the left of +1 from 1 to 15 bases. Using a ^{32}P-labeled primer (see B above) that is complementary to the 3′ end of the oligonucleotides, an extension reaction (95°C, 2 min; 42°C, 5 min; 72°C, 30 min) was performed using native *Pfu* DNA polymerase (Stratagene) to extend each primer. Each 100 µl reaction contained the following: 10 µl of *Pfu* DNA polymerase buffer (10×), 200 µM (final) of each dNTP, 100 ng of template oligonucleotide and 100 ng of labeled primer B. After extension, each mixture was purified using the QIAquick nucleotide removal kit from Qiagen following the manufacturer's instructions. About 1 ng of final product was used per gel retardation assay as described above. The downstream nucleotide binding requirements were determined in a similar manner. OE_{down}, shown above, was synthesized and an aliquot removed at every base within the brackets. Extension of these oligonucleotides and subsequent gel retardation analysis were performed as described above.

RESULTS

In vitro DNA binding activity of Tnp and ΔEK54

Previous studies have shown that Tnp protein forms two complexes, Complex I and Complex II, in the presence of OE DNA (21). The protein component of Complex I was found to represent a dimeric form of Tnp while Complex II represents binding of naturally-occurring proteolytic products of Tnp in monomeric form (15). A truncated variant of Tnp lacking the COOH-terminal 107 amino acids (Δ369) in complex with OE DNA was determined to be a monomer. Thus the major determinants for dimerization lie in the COOH-terminus of Tnp. While binding activities of both proteins are similar, the kinetics and concentration dependencies of each reaction differ significantly (15). The binding curve of Δ369 in complex with OE is representative of a simple, rapid, one step reaction. In contrast, the sigmoidal shape of the Tnp-OE binding curve and slow rate of complex formation indicates a more complex reaction. Based on previous observations (6,21), it has been proposed that formation of heterodimers through a subunit exchange is the key to this complexity (15).

A number of hypertransposition mutants have been isolated in Tnp, localized in the proposed DNA-binding domain of the protein. One of these mutants, EK54 with a glutamic acid to lysine substitution at amino acid residue 54, has been shown to display an increased OE binding activity *in vitro* which correlates to its increased transposition frequency *in vivo*. The monomeric form of Tnp, Δ369, containing the EK54 mutation also demonstrated an enhanced binding activity (Zhou and Reznikoff, submitted). In

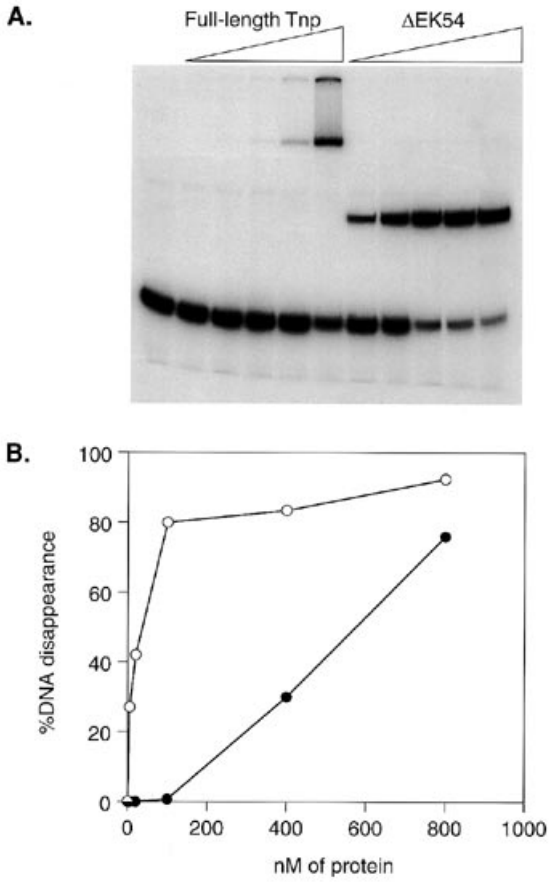

Figure 2. Binding analysis of full-length Tnp and ΔEK54. (**A**) Gel retardation assay of full-length Tnp (filled circles) and ΔEK54 (open circles) in the presence of the OE. The wedge above the lanes for directed and ΔEK54 represents increasing amounts of protein (5, 25, 100, 400 and 800 nM). (**B**) Disappearance of OE DNA (%) at each protein level was determined using a Molecular Dynamics PhosphorImager and plotted as a function of protein concentration.

this study, we have taken advantage of the enhanced binding activity of ΔEK54 to further investigate the molecular basis of DNA binding in a comparison study with full-length Tnp protein.

Binding affinity was tested using purified preparations of full-length Wt Tnp and ΔEK54 protein in a gel retardation assay (Fig. 2A). Increasing amounts of full-length Tnp protein and ΔEK54 were incubated with a 50 bp DNA fragment containing the 19 bp OE sequence and electrophoresed as described in Materials and Methods. Binding activities of each protein were approximated from binding curves generated from a plot of the % disappearance of free DNA as a function of protein concentration. Since protein concentration is in excess, the relative binding affinity (K_{obs}) can be estimated from the protein concentration in which 50% of OE DNA is in complex. Not surprisingly, the K_{obs} of ΔEK54 is significantly higher (~12-fold) than that of full-length Tnp protein. This result correlates with the increased transposition frequency *in vivo* and increased OE binding *in vitro* (Zhou and Reznikoff, submitted). In addition, the hyperbolic and sigmoidal shapes of the binding curves of ΔEK54 and full-length Tnp, respectively, were in agreement with those described previously for full-length Tnp protein and Δ369 (15).

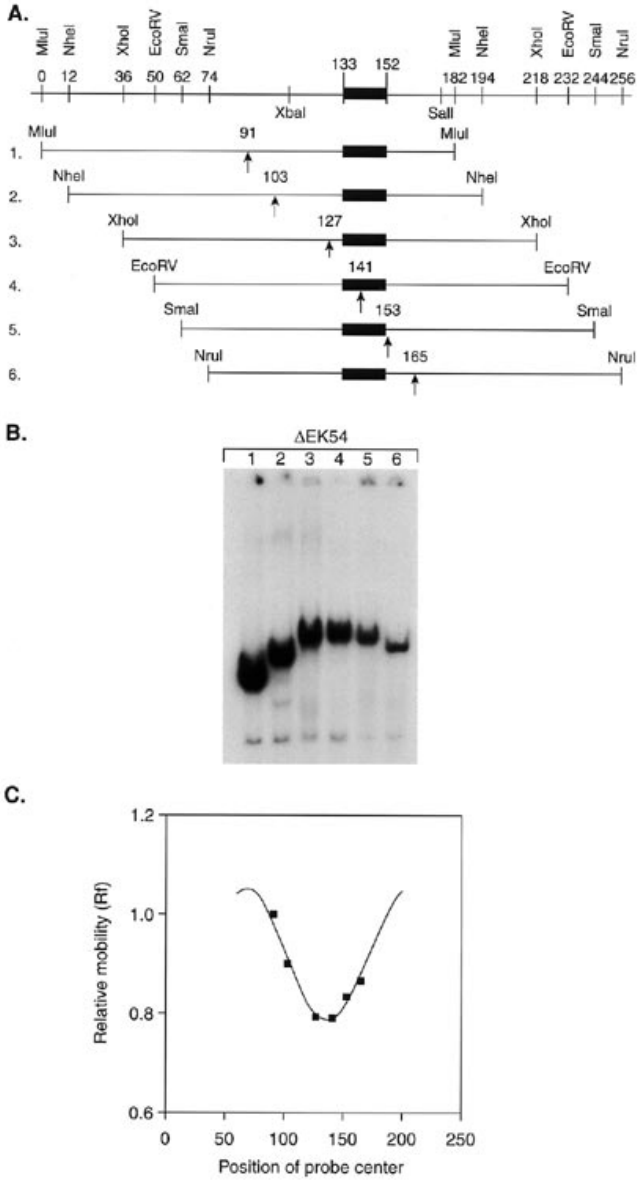

Figure 3. Circular permutation assay. (**A**) DNA probes used in OE bending experiments. The top line shows the 256 bp *Mlu*I/*Nru*I region of the pBend 3 derivative, pRZ9012 (15). The nucleotide positions of the 19 bp OE sequence (solid box) are indicated. Six circularly permuted probes (A–F) of equal size (182 bp), each containing the OE sequence, were generated by digestion with *Mlu*I, *Nhe*I, *Xho*I, *Eco*RV, *Sma*I and *Nru*I, respectively. The arrowheads mark the center of each probe. The nucleotide position of each center is also shown. (**B**) Gel retardation assay of ΔEK54 binding to each of the six bending probes (A–F) containing the OE: *Mlu*I/*Mlu*I, *Nhe*I/*Nhe*I, *Xho*I/*Xho*I, *Eco*RV/*Eco*RV, *Sma*I/*Sma*I and *Nru*I/*Nru*I. Each probe was incubated with 340 nM of ΔEK54 protein. Free and complexed DNA were separated on a 5% polyacrylamide gel at 300 V, 4°C for 2 h. (**C**) Mapping of the bending center of the OE fragment. The relative mobility, R_f (distance of migration of DNA–Tnp complex/distance of mobility of free DNA), of each bending probe was normalized by dividing by the R_f of the complex with the fastest mobility and plotted as a function of the position of the center in each probe as shown in (A). The plot is the average of four independent experiments with a standard deviation of each point <2%. The best fit of a cosine function allowed us to map the center of full-length Tnp-induced and ΔEK54-induced bending to position 134 (the 2nd nucleotide in the OE sequence).

Circular permutation and phasing analyses

Previous studies have shown that full-length Tnp protein and the truncated variant, Δ369, induce severe structural distortions upon binding to the OE (3,15) with the center of this distortion mapping near the first nucleotide of the 19 bp OE consensus sequence (15). Although a circular permutation assay is not a definitive test for protein-induced bending at a specific site, it is a classic indicator of general protein-induced perturbations of DNA including static bending, increased DNA flexibility and aberrant protein structures (22). From a quantitation of DNA distortion based on an empirical formula described by Kerppola and Curran (19; see Materials and Methods), the distortion angle (α_D) and location of this distortion was determined for full-length Tnp and ΔEK54 protein using a circular permutation assay. Six circularly permuted 182 bp fragments containing the 19 bp OE sequence at various positions (Fig. 3A) were used in a gel retardation analysis with full-length Tnp (data not shown) and ΔEK54 protein (Fig. 3B). The relative mobilities (relative distances of migration of the bound complexes divided by the migration distance of the respective free DNAs) were plotted as a function of the position of the center of each of the bending probes (Fig. 3A). Shown in Figure 3C is the result of the plot of ΔEK54. This curve allowed us to map the center of the ΔEK54-induced distortion, located at the lowest point of the curve, at position 134 which maps to the 2nd nucleotide in the 19 bp OE sequence. The same position was obtained from a plot of full-length Wt Tnp (data not shown) which is in good agreement with previous studies (3,15). Both proteins distort the OE DNA at an angle of ~80°.

A more direct method currently used to detect directed DNA bends is a phasing analysis. This is a specific method for the identification and analysis of DNA bends, based on a phase-dependent interaction between a protein-induced bend and an intrinsic DNA bend located on the same fragment. The intrinsically bent DNA used is a set of A–T tracts shown by the Crothers group (23) to bend DNA towards the minor groove at an angle of 18°. For these analyses, a DNA segment containing three A–T tracts (Fig. 4A) separated by 5 bp (center of the tracts separated by 1 helical turn) was cloned upstream of the OE. Between the A–T tracts and the OE, a variable length spacer (0, 2, 4, 6, 8 and 10 bp) was placed. Because the length of the spacer varies over one helical turn, the binding sites will be present on different faces of the DNA relative to the intrinsic bend. Each of these fragments were then used in a gel retardation assay. If the protein does not bend at the OE, all the Tnp complexes will migrate to approximately the same distance. If the protein does bend the DNA at its binding site, the mobility of complexes will vary such that it is slowest when the two bends cooperate to increase the overall extent of bending (in-phase) and fastest when the two bends counteract each other and basically cancel each other out (out-of-phase). Figure 4B shows the results of a gel retardation assay of fragments containing the OE complexed with full-length Tnp (left) and ΔEK54 (right), the results of which indicate that both proteins bend the OE sequence. The relative mobility of each OE complex was plotted as a function of the number of bp from the center of the A–T segment, shown by an arrow in Figure 4A, and the center of Tnp-induced bending at position +2. The highest point of the curve (Fig. 4C) represents the fastest migrating complex in Figure 4B, while the lowest point of the curve represents the slowest migrating or in-phase complex. At the lowest point of the curve, the distance between the center of

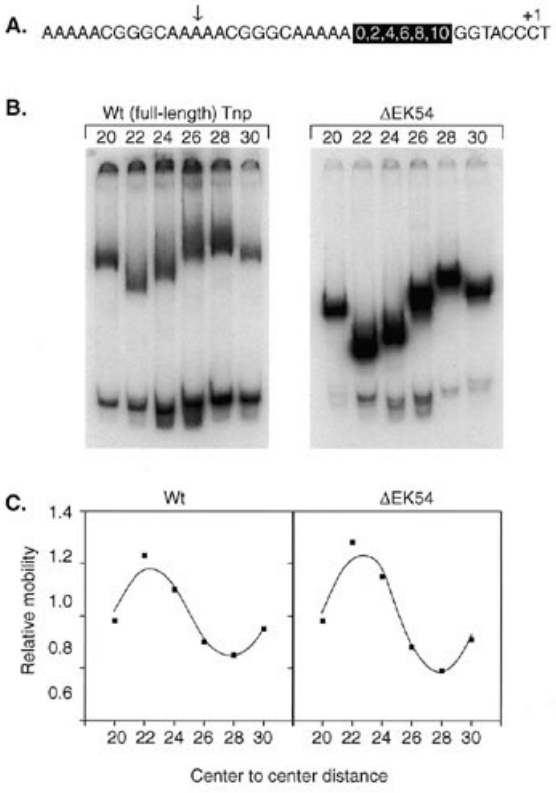

Figure 4 . Phasing analyses of the OE. (**A**) A three A–T tract segment of DNA, the centers of which are separated by one helical turn, was cloned upstream of the OE sequence. The first nucleotide (+1) is indicated with the rest of the sequence not shown. A variable length spacer, 0, 2, 4, 6, 8 and 10 bp was added in the position of the rectangle. (**B**) Each of these fragments were used in a mobility shift assay with full-length Tnp (far left panel) and ΔEK54 (right panel). (**C**) The relative mobility of each complex (migration distance of complex divided by migration distance of free DNA) was determined and normalized by dividing by the average relative mobility of all the complexes. The average relative mobility of six independent experiments (standard deviation of each point <1.5%) was plotted as a function of the distance from the center of the A–T tract segment as shown by the arrow in (A) and the center of the Tnp-induced bend at position +2, as determined from the circular permutation assay. The points were connected by the best fit of a cosine function.

Tnp-induced bend and the center of the intrinsic bend is 28 bp or 2.5 helical turns. Therefore, Tnp bends towards the major groove. The directed bend angle (α_B), calculated as described by Kerppola and Curran (19; see Materials and Methods) was found to be 36° for full-length Tnp and 48° for ΔEK54 protein. The directed bend angle of the monomeric form of Tnp, Δ369, was found to be comparable to that of the full-length protein (data not shown). The significance of these results will be addressed in the Discussion.

OE sequence requirements for efficient Tnp binding

The overall center of the bend maps to about position two of the OE. Such a bend center suggests the possibility of additional contacts between Tnp and neighboring vector DNA. Non-specific contacts between the γδ transposase and end-neighboring DNA has also been suggested based on hydroxyl radical protection footprinting (24).

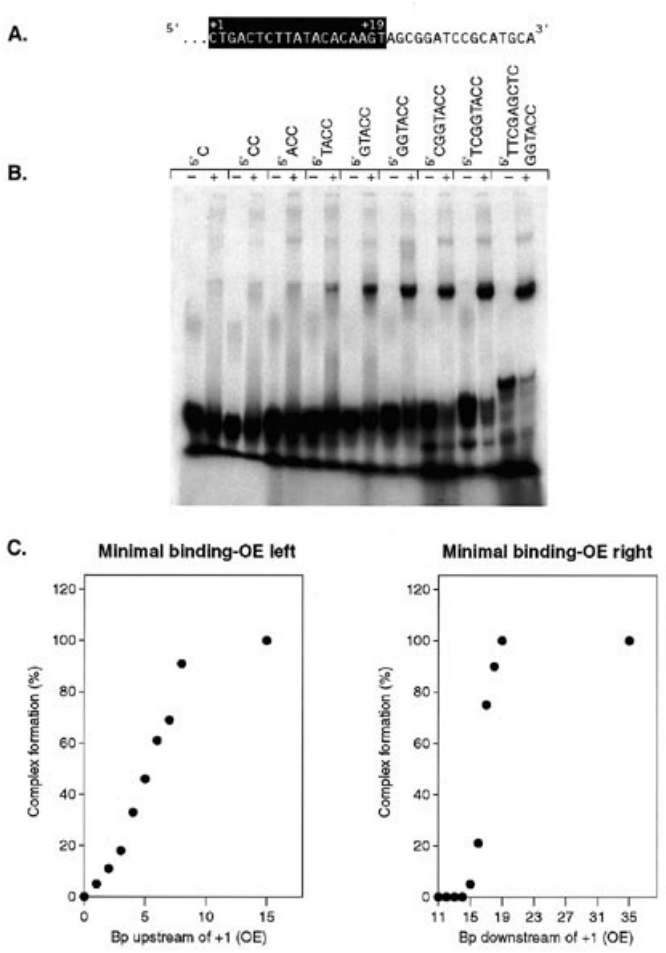

Figure 5. OE sequence requirements for binding of ΔEK54. (**A**) The sequence of the synthesized oligonucleotide containing the 19 bp OE is shown. The position of +1 and +19 of the OE are indicated. The three dots indicates the direction of additional bases added upstream of the OE sequence. (**B**) Gel retardation of ΔEK54 with OE fragments. As each additional base was added to the end of the oligonucleotide shown in (A), an aliquot was removed, resulting in 15 oligonucleotides which differ on the 5′ end by 1–15 additional bases, as indicated above each set of lanes on the gel. The recessed 3′ OH ends of these oligonucleotides were extended (see Materials and Methods), and the double-stranded fragments were used in a gel retardation assay with ΔEK54 (340 nM). (**C**) Binding efficiencies of extended fragments. The % of free and complexed DNA was measured on a PhosphorImager. The binding efficiencies of each fragment were normalized to that of the fragment containing 15 bp upstream of +1 (arbitrarily set to 100%) and plotted as a function of the number of base pairs upstream of +1, as shown in the left panel. Nucleotide requirements downstream of +1 were also determined in a similar manner and the results of the plot are shown in the right panel.

To address this question, an oligonucleotide was synthesized (3′–5′) containing the OE sequence. This synthesis was interrupted at every base beyond +1 (Fig. 5A) and an aliquot removed. This resulted in 15 oligonucleotides that differ only in the number of bases to the left of +1 from 1 to 15 bases. Using a ^{32}P-labeled primer that is complementary to the 3′ end of the oligonucleotides, each template was extended to yield blunt-ended double-stranded fragments that differ in length from 1 to 15 bp upstream of +1. Fragments containing 1–8 bp to the left of +1 and the longest fragment (15 bp upstream of +1) were used in a gel retardation assay with ΔEK54 (Fig. 5B). The same fragments were also used in a binding assay with full-length Tnp (data not shown). This same technique was used to determine the rightmost end of the OE. Each fragment was analyzed for binding efficiencies normalized to the binding efficiency of the full length fragment (arbitrarily set at 100%) and graphed (Fig. 5C). The results show that at least 9–10 bp of adjoining upstream DNA are required for efficient binding of ΔEK54 to the OE. However, no adjacent DNA is required for optimal binding downstream of +19. The same results were seen with full-length Tnp (data not shown).

Essential upstream nucleotide contacts: single-stranded or double-stranded?

Based on the results from the minimal binding site determination, complementary 29 base oligonucleotides containing the 19 bases of the OE sequence and 10 additional bases upstream of +1 were annealed as described under Materials and Methods and used in a comparison gel retardation assay with a 50 bp fragment. We also decided to test the possibility that single-stranded DNA upstream of +1 can suffice to facilitate Tnp–OE binding. Figure 6A shows the templates used in this assay. Preparation of templates is described in the figure legend. In lane 1, a 50 bp fragment containing the OE was incubated in the absence (C) or presence

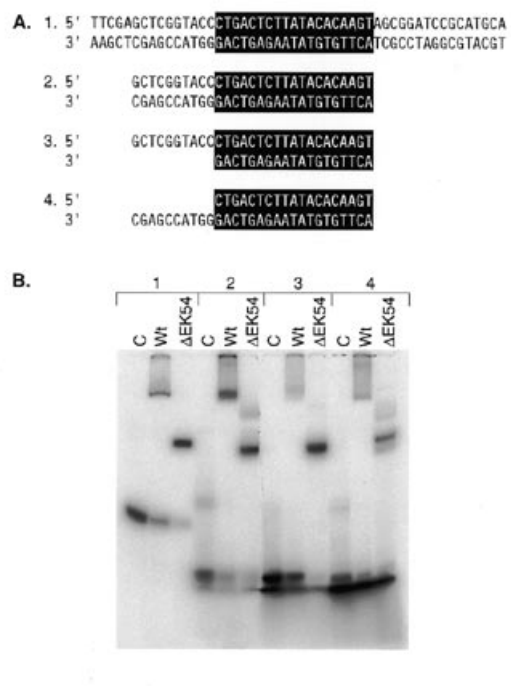

Figure 6. Tnp binding to the OE with single-stranded tails. Complementary oligonucleotides, each containing the OE consensus (rectangle) and shown in (**A**) 1, 50 top + 50 bottom; 2, 29 top + 29 bottom; 3, 29 top + 19 bottom; and 4, 19 top + 29 bottom were annealed as follows. Equal molar amounts were heated to 65°C for 10 min and allowed to cool slowly to room temperature. 100 ng of these templates were then labeled by a kinase reaction using [γ-^{32}P]ATP. Unincorporated nucleotides were removed by chromatography through a Sephadex-50 column. Fragments (1 ng) were then used in a gel retardation assay either in the absence (C) of protein or in the presence of either full-length Tnp protein (750 nM) or ΔEK54 (300 nM) and run on a 9% gel, at 300 V, 4°C for 2.5 h. A Molecular Dynamics PhosphorImager was used to determine binding efficiencies.

of full-length Tnp (750 nM) and ΔEK54 (300 nM). Not surprisingly, both proteins bind efficiently to this DNA. Binding efficiency of the 50 bp fragment was arbitrarily set to 100% for comparison studies with the other templates. Lanes 2 show the results with the 29 bp fragment. Both proteins bind this template with similar efficiency as the 50 bp fragment. No binding is observed to the 19 bp OE fragment or to a single-stranded oligonucleotide containing the OE (data not shown). However, when either the top strand of the 29mer oligonucleotide and the complementary strand of the 19mer are annealed or the top strand of the 19mer oligonucleotide is annealed to the bottom strand of the 29mer oligonucleotide and these substrates are used as Tnp binding targets, binding is restored (lanes 8 and 9). Wt Tnp binding, however, does appear to be somewhat less efficient for the 5′ overhang when compared to ΔEK54. Together, these results suggest that the contacts that Tnp protein needs upstream of +1 of the OE for optimal binding can be met by either single-stranded or double-stranded DNA.

DISCUSSION

Transposition is a type of recombination involving the movement of a specific DNA sequence, called a transposable element, to a new location on the same or different DNA molecule. Transposable elements are found in virtually all living organisms. In addition to their ability to translocate, these elements also generate deletions, inversions and chromosomal fusions. Their ability to facilitate genomic restructuring strongly implicates their role as an evolutionary tool. In addition, transposable elements have an equally important role in medical research. Many bacterial transposable elements (transposons) carry antibiotic resistance genes. Therefore, the act of transposition is crucial to the dissemination of antibiotic resistance among bacteria. More importantly, the mechanism of insertion of retroviral DNA, including HIV-1, is fundamentally similar to that of bacterial transposons (for a general review of transposable elements, see 25).

Mechanistically prokaryotic transposition and the integration of a number of other DNA elements including HIV-1 are closely related and basically follow the same steps. The first dedicated step is the site-specific binding of the transposase or integrase to DNA end sequences. Tn*5* transposase has been shown *in vitro* to bind specifically to both the OE and IE (3,6,26). In the second step of transposition, a higher-ordered protein–DNA complex is formed through protein oligomerization that brings the two ends together. In the case of Tn*5* transposition, it has been proposed that the monomeric form of Tnp is the active form of the protein which is able to bind (and bend) to one end and mediate synapse formation through dimerization with a second Tnp monomer bound at the other end. After synapse formation, either a single-stranded or double-stranded cleavage, as in the case of Tn*5* (Goryshin and Reznikoff, in preparation), occurs to yield a highly reactive 3′ OH. Bending of the end sequence would help direct the Tnp-induced cleavage precisely at +1. Tnp complexed to the OE binds to target DNA. The 3′ OH of the transposon, in a concerted reaction, attacks the 5′ PO_4 of the target and the two ends join in a strand exchange reaction. This strand exchange product is resolved through a simple 'cut and paste' mechanism as proposed for Tn*7* (27), Tn*10* (28,29) and Tn*5* (30). The gap generated by the double-strand cut (9 bases in the case of Tn*5*, 31–33) is repaired by host machinery resulting in a short duplication.

The K_{obs} of ΔEK54 is significantly higher (~12-fold) than that of full-length Tnp protein, as demonstrated in the gel retardation assay. This result correlates with the increased transposition frequency *in vivo* and increased OE binding *in vitro* (Zhou and Reznikoff, submitted).

Transposase proteins from a number of other systems have been shown to distort DNA upon binding (34–36). In this study, we have shown that Tnp, as well as ΔEK54, distort the OE DNA upon binding as determined by a circular permutation assay (3,15). This distortion is due in part to a directed bend as demonstrated by the phasing analysis. The directed bend angle, measured as described by Kerppola and Curran (19), is ~36° for full-length Tnp and 48° for ΔEK54. Both bends are oriented towards the major groove.

Although the distortion angle of both proteins are similar there is an apparent difference in the directed bend angle of full-length and ΔEK54. The smaller static (spatially oriented) bend angle of full-length Tnp suggests that the protein exerts more of a non-directional alteration of the OE in comparison to ΔEK54. The shallower bend angle of full-length Tnp could be related to the oligomeric state of the protein. Full-length Tnp protein binds to the OE sequence in dimeric form while the ΔEK54 in complex with the OE is in monomeric form (15). Therefore, the dimerization interface of the full-length Tnp protein may affect DNA bending. This model has been proposed to explain the lack of bending by a small subclass of bZIP homodimers: CREB,

ATF1 and ATF2 (37). In particular, while ATF2 can bend DNA in heterodimers formed with Fos, Jun or Fra2, homodimers do not induce bending. However, the results of the phasing analyses with Δ369 protein, the Wt monomeric form of Tnp (15) argues against this model. Both proteins were shown to have similar directed bend angles. Therefore, we believe that the EK54 mutation is responsible for the increase in the bend angle presented in this study. Recent studies by Zhou and Reznikoff (submitted) suggest that the EK54 variant of Tnp possesses a somewhat modified OE sequence specificity when compared to Wt Tnp in that additional base pairs participate in the binding reaction. This change in nucleotide contacts may explain the difference in both the binding affinity and the bending ability of the ΔEK54 protein.

The center of the static bend for both proteins maps near the first nucleotide of the 19 bp OE fragment. This result compliments observations made from a missing nucleoside experiment (3). Removal of nucleosides at positions 1–5 enhance Tnp binding. Because it has been demonstrated that Tnp bends the DNA in this region of the OE, we interpret this result as follows. Removal of these nucleosides relieve the structural constraints of an intact DNA backbone. This reduces the energy involved in bending resulting in more stabilized protein binding.

The location of the Tnp-induced bend center near the first nucleotide in the 19 bp OE sequence is most interesting when considered in the light of the subsequent cleavage and strand exchange reactions. Results from a recently developed *in vitro* transposition system have unambiguously shown that both the 3′ OH and the 5′ cleavage occur precisely at the +1/–1 boundary, (Goryshin and Reznikoff, in preparation). Therefore, the bend near +1 may actually aid in directing the Tnp-mediated cleavage precisely at the end of the OE.

In addition, the location of the bend center near the extreme left end of the OE strongly suggests that additional nucleotide contacts are needed for optimal Tnp binding. Minimal binding site determination, presented in this study, revealed that 9–10 nt are needed upstream of +1 while no additional contacts are needed downstream (+19) for efficient Tnp binding. Furthermore, these additional required contacts can be either single-stranded or double-stranded.

The absence of any detectable specific binding of Tnp to the 19 bp OE sequence *in vitro* is somewhat of a paradox in the context of the proposed model for transposition of Tn*5*. According to the model, initial Tnp contact would be at the double-stranded OE sequence. After synaptic complex formation, Tnp would cleave at the 3′ and 5′ at the boundary between –1 and +1. From the *in vitro* studies we would predict that the Tnp–OE complex would be unstable and the protein should fall off. However, this would result in abortive transposition. By what mechanism could the protein stay in complex with the donor DNA in order to mediate subsequent target cleavage and strand transfer? The ability of Tnp protein to remain in complex after donor cleavage must require either that prior synapse (oligomerization) formation stabilizes the Tnp-cleaved OE complex through an allosteric event or that a concerted reaction between cleavage and strand transfer occurs. Recent *in vitro* transposition results from Goryshin and Reznikoff (manuscript in preparation) indicate that the donor cleavage and strand transfer reactions are not concerted; therefore synapse formation, specifically oligomerization, probably stabilizes the Tnp-cleaved OE complex through an allosteric change in Tnp.

ACKNOWLEDGEMENTS

We wish to thank Chris Scherrer at the Biotechnology Center for preparation of the oligonucleotides used in the minimal binding determination experiment. This work was supported by NIH grant GM50692 and a gift from the Promega Corporation. D.Y. was supported by NCI training grant T32 CA09075. W.S.R. is the Evelyn Mercer Professor of Biochemistry and Molecular Biology.

REFERENCES

1 Johnson, R. C. and Reznikoff, W. S. (1983) *Nature* **304**, 280–282.
2 Sasakawa, C., Carle, G. F. and Berg, D. E. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 7293–7297.
3 Jilk, R. A., York, D. and Reznikoff, W. S. (1996) *J. Bacteriol.* **178**, 1671–1679.
4 Reznikoff, W. S. (1993) *Annu. Rev. Microbiol.* 945–963.
5 Krebs, M. P. and Reznikoff, W. S. (1986) *J. Mol. Biol.* **192**, 781–791.
6 de la Cruz, N. B., Weinreich, M. D., Wiegand, T. W., Krebs, M. P. and Reznikoff, W. S. (1993) *J. Bacteriol.* **175**, 6932–6938.
7 Isberg, R. R., Lazaar, A. L. and Syvanen, M. (1982) *Cell* **30**, 883–892.
8 Johnson, R. C., Yin, J. C. and Reznikoff, W. S. (1982) *Cell* **30**, 873–882.
9 Rothstein, S. J. and Reznikoff, W. S. (1981) *Cell* **23**, 191–199.
10 Weinreich, M. D., Mahnke, B. L. and Reznikoff, W. S. (1994) *J. Mol. Biol.* **241**, 166–177.
11 Johnson, R. C. and Reznikoff, W. S. (1984) *J. Mol. Biol.* **177**, 645–661.
12 Rezsohazy, R., Hallet, B., Delcour, J. and Mahillon, J. (1993) *Mol. Microbiol.* **9**, 1283–1295.
13 Engelman, A., Mizuuchi, K. and Craigie, R. (1991) *Cell* **67**, 1211–1221.
14 Katz, R. A., Mack, J. P., Merkel, G., Kulkosky, J., Ge, Z., Leis, J. and Skalka, A. M. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 6741–6745.
15 York, D. and Reznikoff, W. S. (1996) *Nucleic Acids Res.* **24**, 3790–3796.
16 Weinreich, M. D., Gasch, A. and Reznikoff, W. S. (1994) *Genes Dev.* **8**, 2363–2374.
17 Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L. and Boyer, H. W. (1977) *Gene* **2**, 95–113.
18 Sutcliffe, J. G. (1979) *Cold Spring Harb. Symp. Quant. Biol.* **2**, 77–90.
19 Kerppola, T. K. and Curran, T. (1991) *Science* **254**, 1210–1214.
20 Thompson, J. F. and Landy, A. (1988) *Nucleic Acids Res.* **16**, 9687–9705.
21 Wiegand, T. W. and Reznikoff, W. S. (1994) *J. Mol. Biol.* **235**, 486–495.
22 Kerppola, T. K. and Curran, T. (1991) *Cell* **66**, 317–326.
23 Zinkel, S. S. and Crothers, D. M. (1987) *Nature* **328**, 178–181.
24 Wiater, L. A. and Grindley, N. D. (1991) *J. Biol. Chem.* **266**, 1841–1849.
25 Berg, D. E. and Howe, M. M. (1989) *Mobile DNA* (Am. Soc. for Microbiology, Washington DC).
26 Makris, J. C., Nordmann, P. L., and Reznikoff, W. S. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2224–2228.
27 Bainton, R., Gamas, P. and Craig, N. L. (1991) *Cell* **65**, 805–816.
28 Benjamin, H. W. and Kleckner, N. (1989) *Cell* **59**, 373–383.
29 Morisato, D. and Kleckner, N. (1984) *Cell* **39**, 181–190.
30 Berg, D. E. (1989) in Berg, D. E. and Howe, M. M. (eds), *Mobile DNA.* American Society for Microbiology, Washington DC, pp. 184–210.
31 Berg, D. E., Johnsrud, L., McDivitt, L., Ramabhadran, R. and Hirschel, B. J. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 2632–2635.
32 Nag, D. K., DasGupta, U., Adelt, G. and Berg, D. E. (1985) *Gene* **34**, 17–26.
33 Schaller, H. (1979) *Cold Spring Harb. Symp. Quant. Biol.* 401–408.
34 Arciszewska, L. K. and Craig, N. L. (1991) *Nucleic Acids Res.* **19,** 5021–5029.
35 Derbyshire, K. M. and Grindley, N. D. (1992) *EMBO J.* **11**, 3449–3455.
36 Kuo, C. F., Zou, A. H., Jayaram, M., Getzoff, E. and Harshey, R. (1991) *EMBO J.* **10**, 1585–1591.
37 Kerppola, T. K. and Curran, T. (1993) *Mol. Cell. Biol.* **13**, 5479–5489.