Uracil interference, a rapid and general method for defining protein-DNA interactions involving the 5-methyl group of thymines: The GCN4-DNA complex

William T.Pu and Kevin Struhl*

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA

ABSTRACT

We describe a novel uracil interference method for examining protein contacts with the 5-methyl group of thymines. The protein of interest is incubated with target DNA containing randomly distributed deoxyuracil substitutions that is generated by carrying out the polymerase chain reaction in the presence of a mixture of TTP and dUTP. After separating DNA-protein complexes away from unbound DNA, the locations of deoxyuracil residues that either do or do not interfere with DNA-binding are determined by cleavage with uracil-N-glycosylase followed by piperidine. Using this uracil interference assay, we show that the methyl groups of the four core thymines, but not the two peripheral thymines, of the optimal binding site (ATG-ACTCAT) are important for high affinity binding of GCN4. Similar, but not identical, results are obtained using $KMnO_4$ interference, another method used for studying protein-DNA interactions involving thymine residues. These observations strongly suggest that GCN4 directly contacts the 5-methyl groups of the four core thymines that lie in the major groove of the target DNA. Besides providing specific structural information about protein-DNA complexes, uracil interference should also be useful for identifying DNA-binding proteins and their target sites in eukaryotic promoter regions.

INTRODUCTION

GCN4 is a member of the AP-1 family of eukaryotic transcription factors that also includes the Jun and Fos oncoproteins (1). GCN4 contains a bZIP DNA-binding domain (2) and binds its DNA target sequence as a homodimer (3). The optimal target sequence (ATGACTCAT) is composed of two overlapping, nonequivalent half-sites (ATGAC and ATGAG) (4–6). The protein-DNA interactions that define the binding specificity of GCN4 and other bZIP proteins are poorly understood. One specific model of the bZIP protein-DNA complex, termed the induced fork, predicts that four residues which are highly conserved among bZIP

proteins make direct, sequence-specific contacts with DNA (7). Two of these residues are alanines, a residue that is limited to making hydrophobic contacts with the thymine 5-methyl group. Thus, we wished to determine if any thymine 5-methyl groups within the GCN4 recognition sequence were important for high-affinity DNA-binding by GCN4.

The thymine methyl group is important for the high-affinity binding of many proteins to their recognition sites (8). A variety of methods have been developed to study contacts between proteins and the 5-methyl groups of thymines. The first such method relied on the fact that 5-bromodeoxyuracil substituted sites are protected from UV-induced strand scission by proteins which lie in close proximity to the 5-bromodeoxyuracil residue (9). The role of the thymine methyl group has also been studied by synthesizing oligonucleotides in which the methyl group on individual thymines is replaced by other functional groups, such as hydrogen (deoxyuracil), or bromine (5-bromodeoxyuracil) (10). More recently, interference methods in which hydrazine is used to remove the thymine base and open the sugar ring (11), or in which $KMnO_4$ is used to modify the thymine ring (12), have also been described. However, none of these methods are entirely satisfactory because they are not specific for the 5-methyl group of thymine (e.g. base elimination, $KMnO_4$ modification, protection from UV-strand scission) and/or they involve the expensive synthesis of many oligonucleotides with substitutions in specific positions.

In this paper, we describe a simple and general uracil interference method that is specific for examining protein contacts with the thymine 5-methyl group but does not require *de novo* synthesis of any uracil-substituted oligonucleotides (Figure 1). Template DNA (either an oligonucleotide or a restriction fragment) is amplified using the polymerase chain reaction (PCR) in the presence of a mixture of TTP and dUTP, thereby producing products in which deoxyuracil is substituted for thymine on both strands at a frequency dependent upon the dUTP:TTP ratio. The resulting collection of DNA molecules is incubated with the protein of interest. DNA molecules containing deoxyuracil substitutions that do not interfere with protein binding are then selected by purifying the DNA-protein complex away from

* To whom correspondence should be addressed

unbound DNA. The locations of these noninterfering deoxyuracil residues are determined by digesting the purified DNA with uracil-N-glycosylase followed by piperidine and analyzing the products on a denaturing polyacrylamide gel. Uracil-N-glycosylase specifically cleaves uracil bases from DNA, leaving apyrimidinic sites that are susceptible to cleavage by piperidine.

We use this uracil interference assay to show that the methyl groups of the four core thymines, but not the two peripheral thymines, of the open binding site (ATGACTCAT) are important for high affinity binding of GCN4. The results are similar, but not identical, to those obtained using $KMnO_4$ interference, another method used for studying protein-DNA interactions involving the thymine 5-methyl group.

MATERIALS

DNAs and protein

Synthetic oligonucleotides used in this study (Figure 2) were purified by electrophoresis in denaturing polyacrylamide gels. GCN4p, a 58 residue peptide containing the entire DNA-binding domain, was purified from an overproducing E. coli strain (13). Full-length GCN4 was also synthesized in vitro using a wheat germ extract as described previously (14).

Uracil interference assay

Thymine was randomly replaced by deoxyuracil in duplex DNA molecules containing the optimal GCN4 binding site (ATGAC-TCAT) by PCR amplification. Specifically, 0.2 pmol of OP-1, 20 pmol of unlabeled primer, and 20 pmol of ^{32}P 5'-end-labeled
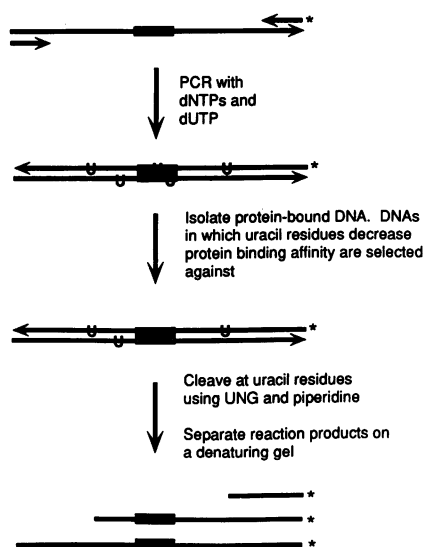

Figure 1. Uracil Interference. An oligonucleotide or a restriction fragment (long arrow) containing a protein binding site (hatched box) is amplified by PCR using one unlabeled primer (short arrow) and one 5'-labeled primer (*) in the presence of dGTP, dATP, dCTP, dTTP and dUTP, producing reaction products in which deoxyuracil is randomly substituted for thymine on both strands. This collection of DNA molecules is incubated with the protein of interest, and DNA molecules containing deoxyuracil substitutions that do not interfere with protein binding are selected by purifying the DNA-protein complex away from unbound DNA. The resulting DNA is cleaved at uracil residues using uracil-N-glycosylase followed by piperidine, and the reaction products are separated on a denaturing polyacrylamide gel.

primer were combined in 100 μl PCR buffer (50 mM KCl, 3.5 mM $MgCl_2$, 10 mM Tris—HCl pH 8.3, 0.01% gelatin) containing 200 μM each dNTP, 50 μM dUTP, and 5 U Taq DNA polymerase and amplified for 8 PCR cycles (94°C 1 min, 40°C 2 min, 72°C 2 min). The reaction products were electrophoretically separated on an 8% native polyacrylamide, and the full length PCR product was recovered.

The purified PCR product was incubated in DNA-binding buffer (20 mM Tris pH 7.4, 50 mM KCl, 1 mM EDTA, 220 μg/ml poly dI-dC) either with 5 pmol GCN4p (45 μl reaction volume) or with 5 μl in vitro synthesized protein (60 μl reaction volume), and the protein-DNA complex was separated from unbound DNA by electrophoresis as described previously (15). DNA was recovered from the protein-DNA complex and digested for 1 hour with 1 U uracil-N-glycosylase (Perkin Elmer-Cetus) at 37°C in 60 μl PCR buffer, ethanol precipitated, and treated with 1 M piperidine for 30 min at 90°C. The sample was lyophilized, resuspended in formamide, and electrophoretically separated on a 12% denaturing polyacrylamide gel. As a control, a portion of the PCR product was digested with uracil-N-glycosylase and then treated with piperidine to determine the degree of uracil substitution at each position in the unselected DNA population.

$KMnO_4$ interference

This assay was performed essentially as described previously (12). The duplex substrates containing the optimal GCN4 site were obtained by annealing two complementary oligonucleotides (OP-1 and OP-2). To create a $KMnO_4$ modified duplex labelled at the 5'-end of the top strand, OP-1 labelled at the 5'-end with ^{32}P in 5 μl of a 30 mM Tris—HCl buffer (pH 8) was denatured at 90°C for 1 min, chilled on ice, and mixed with 20 μl of a 0.25 mM $KMnO_4$ solution. After 10 min at 20°C, the reactions were stopped by mixing with 50 μl stop buffer (1.5 M sodium acetate pH 7.0, 1 M 2-mercaptoethanol). DNA was ethanol precipitated twice and annealed to the bottom strand oligonucleotide OP-2. To create a $KMnO_4$ modified duplex labelled at the 5'-end of the bottom strand, ^{32}P 5'-end labelled OP-2 was modified with $KMnO_4$ in the same manner and then annealed to OP-1.

$KMnO_4$ modified DNA was incubated with either 5 pmol GCN4p or 3 μl in vitro synthesized protein in 45 μl binding buffer, and the protein-DNA complex was purified as described above. DNA recovered from this complex was incubated with

A

OP-1   5'-GGGCTTTCTGCTCTGTCATCGAATTCCGGATGACTCAT
        TTTTTGGATCCATGGTCATAGCTGTT-3'

OP-2   5'-AACAGCTATGACCATGGATCCAAAAAATGAGTCATCCG
        GAATTCGATGACAGAGCAGAAAGCCC-3'

Primer-1   5'-GGGCTTTCTGCTCTGTCATC-3'

Primer-2   5'-AACAGCTATGACCATGG-3'

B   5'- A T G A C T C A T -3'
        3'- T A C T G A G T A -5'
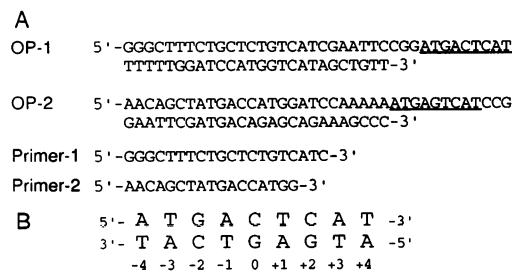   -4 -3 -2 -1 0 +1 +2 +3 +4

Figure 2. (a) Synthetic oligonucleotides used in this study. OP-1 and OP-2 are complementary oligonucleotides containing the optimal GCN4 binding site (underlined) and are referred to as the top and bottom strands, respectively. Primers 1 and 2 are used to amplify OP-1 by PCR. (b) The optimal GCN4 binding site is shown, with the numbering system (5) indicated below.

1 M piperidine for 30 min at 90°C, lyophilized, resuspended in formamide, and electrophoretically separated in a 12% denaturing polyacrylamide gel. A portion of the modified DNA was treated with piperidine to determine the degree of modification at each position in the unselected DNA population.

RESULTS

GCN4 contacts to thymines as revealed by the uracil interference assay

We used a novel uracil interference assay (Figure 1) to study the role of the thymine 5-methyl group in DNA-binding by GCN4p. As shown in Figure 3, uracil substitution of only a subset of thymine residues within the GCN4 target site interfere with GCN4p binding. This indicates that the assay specifically identifies thymines which contribute to the protein binding.

It is clear that the four core thymine methyl groups of the optimal binding site (positions ± 3 and ± 1) contribute strongly to DNA-binding by GCN4p. Quantitative analysis (not shown) indicates that the symmetrically equivalent thymine methyl groups in each half site contribute similarly to GCN4p binding, with at most a two-fold difference observed for uracil substitution of T_{-3} compared to T_{+3}. In contrast, the methyl groups on the peripheral thymines (position ± 4) are much less important for GCN4p binding. Uracil substitution of T_{-4} has no detectable effect, whereas uracil substitution of T_{+4} does reduce DNA binding to a small extent. These observations are consistent with and extend the results obtained by saturation mutagenesis (4) and random selection (5) experiments (see Discussion).
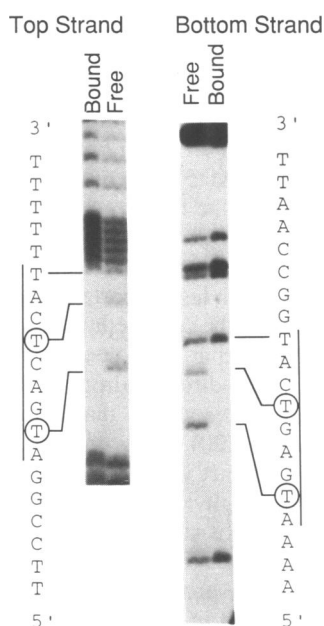
The uracil and KMnO$_4$ interference assays are not identical

KMnO$_4$ attacks the C5-C6 double bond of thymines in single-stranded DNA in a glycolization reaction followed by oxidation to carboxyclic acid and/or aldehyde products and ring opening (12). A previous study suggested that this modification interferes with steroid hormone receptor binding only at positions where the protein forms an intimate contact with the thymine 5-methyl group. Specifically, thymines at which substitution by deoxyuracil interfered with binding by steroid hormone receptor corresponded to those thymines at which KMnO$_4$ modification interfered with DNA-binding (12).

To determine if this correlation was upheld in the case of another DNA-binding protein, we studied the interaction of GCN4p with DNA using the KMnO$_4$ interference method (Figure 4). Although modification of any of the four core thymines interferes with DNA binding, the results are subtly different from those obtained using uracil interference. First, KMnO$_4$ modification of the peripheral thymine of the bottom strand (T_{-4}) results in a small, but detectable interference with DNA-binding. Second, KMnO$_4$ modifications appear to have less effect at thymines ± 3 as compared to thymines ± 1.

The discrepancy between the uracil and KMnO$_4$ interference assays is more pronounced when *in vitro* synthesized full-length GCN4 is used instead of *E. coli* derived GCN4p (Figure 5). Although results for the top strand were consistent between the two methods (not shown), the KMnO$_4$ method identifies only one thymine (T_{-1}) on the bottom strand which is critical for DNA binding. In contrast, the uracil method identifies both core thymines on the bottom strand as critical for DNA binding, and quantitative analysis (not shown) suggests that T_{+3} contributes slightly more than T_{-1}. Thus, uracil substitution does not interfere with the same protein-DNA interactions as KMnO$_4$ modification. Presumably, the uracil interference assay is more
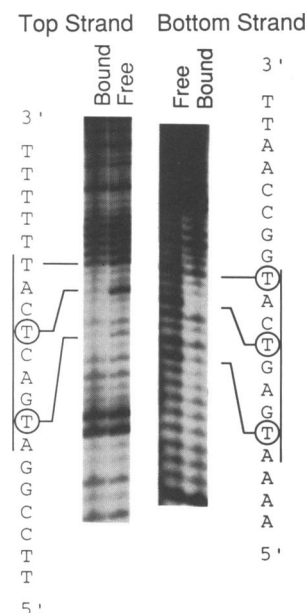

Figure 3. The uracil interference assay detects thymine 5-methyl groups that are important for binding by GCN4p. Uracil-substituted DNAs labeled at the 5'-end of the top or bottom strand were incubated with GCN4p, and protein-DNA complexes purified. The bound and input (free) DNAs were cleaved at uracil residues, and the products were analyzed on a denaturing polyacrylamide gel. The sequences of the top and bottom strands are shown, with the optimal GCN4 binding site marked by a bar. The bands on the gel corresponding to thymines within the binding site are indicated. Positions in which uracil substitution for thymine interferes with protein binding are circled.

Figure 4. The KMnO$_4$ interference assay detects thymines that are important for binding by GCN4p. KMnO$_4$ modified DNAs in which the top or bottom strand was 5'-end-labeled were incubated with GCN4p, and protein-DNA complexes were purified. After piperidine cleavage, the products were analyzed on a denaturing polyacrylamide gel. Symbols are as in Figure 3.
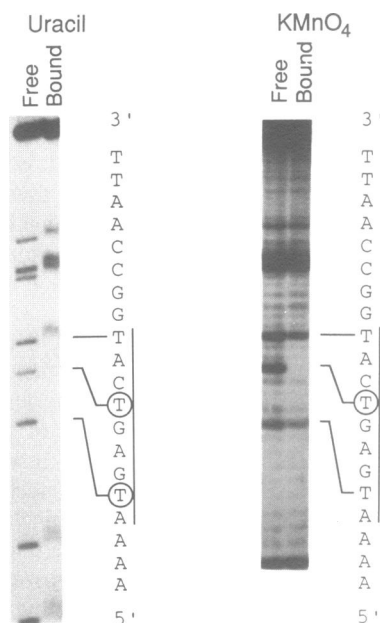
Figure 5. The uracil and KMnO₄ interference assays are not equivalent The interaction of full length GCN4 with the bottom strand of its target site was examined using either the uracil interference assay or the KMnO₄ interference assay. Symbols are as in Figure 3.

specific for protein interactions with the thymine 5-methyl group, because deoxyuracil substitution selectively replaces this methyl group with a hydrogen, while KMnO₄ modification introduces more extensive modifications but leaves the 5-methyl group intact.

The above results also suggest that the full-length protein and the isolated DNA-binding domain interact with the target site in a subtly different manner. However, GCN4p was purified from *E. coli* while full-length GCN4 was synthesized *in vitro*, raising the possibility that the difference is an artifact related to the translation extract and/or the purification procedure. We disfavor this artifactual possibility because mock *in vitro* translation reactions not programmed with GCN4 mRNA do not generate products that bind the target DNA fragment used here (not shown), and because the uracil method yields similar results with GCN4p and *in vitro* synthesized GCN4. In this regard, regions outside the bZIP domains of Fos and Jun influence the affinity of DNA-binding and the magnitude of the protein-induced, DNA bend angle (16−18), although effects on DNA contacts were not detected.

DISCUSSION

Uracil interference is a convenient method for detecting contacts with the 5-methyl group of thymines

Synthetic uracil-substituted oligonucleotides have been used to detect protein interactions with specific 5-methyl groups of thymine residues in DNA target sites (8, 10, 19). The uracil interference method described here is theoretically equivalent to analyzing a collection of synthetic oligonucleotides in which individual thymine residues are replaced by uracil. However, the method does not require the synthesis of a different oligonucleotide for each position to be studied, and hence should drastically reduce the expense and the time for carrying out a complete analysis. Moreover, uracil interference permits the

simultaneous examination of all of the thymines in an internally controlled experiment and it can be applied to DNA fragments as large as hundreds of base pairs. Consequently, the technique should also be extremely useful as an initial biochemical characterization of eukaryotic promoter regions, specifically for the identification of DNA-binding proteins and their target sequences. In this sense, the method is analogous to methylation interference using dimethyl sulfate, a reagent that probes major groove interactions with the N7 position of guanines and to a lesser extent the minor groove interaction with the N3 position of adenines. By providing complementary structural information, uracil interference can become part of the standard characterization of DNA-binding proteins, and it should have particular value for studying protein interactions to dA:dT rich binding sites.

In contrast to previous DNA-binding studies involving the glucocorticoid receptor (12), our analysis of GCN4 indicates that the results of the KMnO₄ interference assay are not identical to those of the uracil interference assay. It is likely that the uracil assay is more specific than the KMnO₄ assay for detecting protein interactions with the thymine methyl group, since substitution of uracil for thymine precisely replaces the thymine methyl group with hydrogen. In contrast, KMnO₄ modification markedly alters the structure of thymine but leaves the 5-methyl group intact. It is surprising that replacement of T$_{+3}$ with deoxyuracil markedly reduced binding by GCN4, whereas modification of this residue by KMnO₄ had only a small effect (Figure 5). Evidently, in this case the more drastic KMnO₄ modification does not interfere with the protein-thymine methyl interaction.

Interactions between GCN4 and DNA

Saturation mutagenesis and random selection experiments demonstrated that the GCN4 binding site is ATGACTCAT with the central 7 bp being critical and the outer 2 bp being less important (4, 5). From these observations, we suggested that GCN4 directly contacts all 7 central bp, but did not determine the specific residues involved (5). The results presented here show that the methyl groups on the four core thymine residues in the binding site are crucial for high-affinity DNA-binding by GCN4p, whereas the methyls on the two peripheral thymines of the binding site (position ±4) are much less important. Although we cannot rule out the possibility that the core methyl groups indirectly affect protein binding by influencing DNA conformation, we favor the hypothesis that they make direct hydrophobic interactions with GCN4p. NMR studies have shown that uracil has little effect on global DNA structure (20), and a previous study of Jun-Fos heterodimers, which interact with the same DNA sequences as GCN4, found the same methyl groups to be important for DNA-binding (19). In the Jun-Fos study, bromouracil substitutions at the same positions inhibited DNA-binding less markedly than uracil substitutions, thus providing further evidence against conformational effects and in favor of direct interactions. The strong evidence that GCN4 directly interacts with the thymine 5-methyl groups is consistent with and supportive of the proposal (7) that the conserved alanines in bZIP proteins (ala238 and ala239 in GCN4) directly contact the GCN4 binding site. Preliminary results suggest that a change of ala238 to tyr238 alters DNA-binding specificity at the ±3 position (D.Tzamarias and K.Struhl, unpublished), thus leading to the prediction of a hydrophobic interaction between the methyl groups of ala238 and the ±3 thymines.

The GCN4 target sequence ATGACTCAT is inherently asymmetric, since it is composed of two overlapping and nonequivalent half-sites, ATGAC and ATGAG (3, 5), with ATG-AC the preferred half-site (6). The asymmetric interference pattern obtained with the $KMnO_4$ probe demonstrates that GCN4 makes nonequivalent contacts with each half-site. The observed pattern cannot be simply due to asymmetry of individual protein-DNA complexes, since it is produced on a population of protein-DNA complexes. Interestingly, mutagenesis experiments suggest that the left half-site contributes more to GCN4 binding than the right half-site (5, 6), and the thymine identified by the $KMnO_4$ method as being relevant to DNA-binding lies in the left half-site. Previous analysis of the Jun-Fos heterodimer also indicated that the contribution of the thymine methyl groups in each half-site was markedly asymmetric, suggesting that the heterodimer interacts asymmetrically with DNA (19). Unlike the situation with GCN4, this asymmetry was observed in uracil substitution experiments, perhaps because binding by the Jun-Fos heterodimer causes a significant bend in the DNA (18), whereas binding by GCN4 does not (13, 21).

ACKNOWLEDGEMENTS

REFERENCES

1. Curran,T. and Franza,B.J. (1988) *Cell* **55**, 395–397.
2. Landschulz,W.H., Johnson,P.F. and McKnight,S.L. (1988) *Science* **240**, 1759–1764.
3. Hope,I.A. and Struhl,K. (1987) *EMBO J.* **6**, 2781–2784.
4. Hill,D.E., Hope,I.A., Macke,J.P. and Struhl,K. (1986) *Science* **234**, 451–457.
5. Oliphant,A.R., Brandl,C.J. and Struhl,K. (1989) *Mol. Cell. Biol.* **9**, 2944–2949.
6. Sellers,J.W., Vincent,A.C. and Struhl,K. (1990) *Mol. Cell. Biol.* **10**, 5077–5086.
7. O'Neil,K.T., Hoess,R.H. and DeGrado,W.F. (1990) *Science* **249**, 774–778.
8. Ivarie,R. (1987) *Nucleic Acids Res.* **15**, 9975–9983.
9. Ogata,R. and Gilbert,W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 4973–4976.
10. Goeddel,D.V., Yansura,D.G. and Caruthers,M.H. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3579–3582.
11. Brunelle,A. and Schleif,R.F. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6673–6676.
12. Truss,M., Chalepakis,G. and Beato,M. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 7180–7184.
13. Weiss,M.A., Ellenberger,T., Wobbe,C.R., Lee,J.P., Harrison,S.C. and Struhl,K. (1990) *Nature* **347**, 575–578.
14. Hope,I.A. and Struhl,K. (1985) *Cell* **43**, 177–188.
15. Pu,W.T. and Struhl,K. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 6901–6905.
16. Abate,C., Luk,D. and Curran,T. (1991) *Mol. Cell. Biol.* **11**, 3624–3632.
17. Cohen,D.R. and Curran,T. (1990) *Oncogene* **5**, 929–939.
18. Kerppola,T.K. and Curran,T. (1991) *Cell* **66**, 317–326.
19. Risse,G., Jooss,K., Neuberg,M., Bruller,H.-J. and Muller,R. (1989) *EMBO J.* **8**, 3825–3832.
20. Delort,A.-M., Neumann,J.M., Molko,D., Herve,M., Teoule,R. and Dinh,S.T. (1985) *Nucleic Acids Res.* **13**, 3343–3355.
21. Gartenberg,M.R., Ampe,C., Steitz,T.A. and Crothers,D.M. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6034–6038.