Binding site selection analysis of protein–DNA interactions via solid phase sequencing of oligonucleotide mixtures

Joseph A.Gogos[1], George Tzertzinis[1] and Fotis C.Kafatos[1,2,*]
[1]Department of Cellular and Developmental Biology, The Biological Laboratories, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA and [2]Institute of Molecular Biology and Biotechnology, Research Center of Crete, PO Box 1527, Heraklion 71110, Crete, Greece

Received January 14, 1991; Accepted March 1, 1991

ABSTRACT

By combining the concept of degenerate oligonucleotide mutagenesis (1,2,3,4) and the convenience of solid phase chemical DNA sequencing (5), we have developed a rapid procedure for determining the specificity of DNA-binding proteins *in vitro*. Starting with a degenerate oligonucleotide mixture, the technique assays for alternative nucleotides in fractions that are bound or non-bound to the protein of interest. In contrast to previous approaches using degenerate oligonucleotides, it does not involve cloning but rather employs direct sequencing of the oligonucleotide mixtures after attachment to a solid support. Solid state processing obviates the need for both DNA extractions from polyacrylamide gels and time-consuming ethanol precipitations. Because of its convenience and sensitivity, this binding site selection analysis is well suited to determining rapidly the sequence preference of DNA-binding proteins that are available in small amounts, and complements well established approaches like methylation interference or missing contact assays. The solid phase reaction protocol we propose can also improve these latter approaches.

INTRODUCTION

A wealth of information on the nature of protein-DNA interactions has been obtained using biochemical approaches that assay either for protection of bound DNA, or for interference in binding, combined with mutational analysis. Although mutational studies were based initially on individual oligonucleotides, their power was greatly enhanced by using degenerate oligonucleotide mixtures, from which particular sequences could be selected by protein binding (1,2,3,4). Similar selection procedures have been used for studies of RNA specificity (6,7). However, protocols which involve cloning of the oligonucleotides before and after protein binding are laborious: a large number of clones must be obtained and sequenced to yield significant information. To make selection procedures convenient for everyday practice, we have experimented with direct sequencing of oligonucleotide mixtures. In this case, the 'acceptability' of a mutant base to the protein might be evaluated from the intensity of the corresponding band in the autoradiogram of a single sequencing gel, rather than from the statistics of binary representation (presence or absence) among multiple individual clones. Indeed, we were successful in developing a strategy for binding site selection analysis (BSSA), which starts with a low affinity site and detects in the protein-bound oligonucleotide mixture the increased abundance of bases that improve the binding affinity.

Pooled sequencing is greatly facilitated by a protocol for immobilizing the selected and non-selected oligonucleotides on a solid support compatible with direct chemical sequencing.

MATERIALS AND METHODS

Oligonucleotides were synthesized by the phosphoramidite method with an automated oligonucleotide synthesizer and were purified by denaturing polyacrylamide gel electrophoresis. For the degenerate positions (typically three per oligonucleotide), mixtures of phosphoramidites were used, usually containing 70% wild type base and 10% each of the three other bases. The proportions of non-wild type bases were dictated by two kinds of considerations. The upper limit of these proportions (and the limit of three mutated positions per oligonucleotide) was imposed mainly by the desire to have a high frequency of *singly* mutated oligonucleotides in the mixture in order to simplify the final analysis. The lower limit was dictated by the background noise of the sequencing reactions (see below). Oligonucleotides were phosphorylated at high specific activity with polynucleotide kinase according to standard protocols (8). The mixture was converted to perfectly matched duplexes by second strand synthesis using the Klenow fragment of DNA polymerase I, or T7 DNA polymerase. Annealing of the primer was performed in 3× polymerase buffer (30 mM Tris pH 7.4, 150 mM NaCl, 15 mM DTT) by heating at 95° for 3 min and slow cooling to room temperature. Extension was performed in 1×buffer at 37° after adding dNTPs at a final concentration of 0.2 mM. Double

* To whom correspondence should be addressed at Harvard University.

stranded products were separated from excess template or unextended primer by non-denaturing gel electrophoresis and were purified from acrylamide by crush elution. We used a 12 nucleotide primer, but primers as small as 6 nucleotides are acceptable, maximizing the proportion of the oligonucleotide available for mutation.

Protein-binding reactions were performed at 4°, in 4 mM Tris-HCl pH 7.5, 60 mM NaCl, 5% glycerol, 0.5 mM DTT, 1 mM EDTA, 0.3 μg/μl of poly(dI-dC).poly(dI-dC). In the case of CF2, crude bacterial extracts containing recombinant protein were used and 0.1 mM $ZnCl_2$ was included in the reaction mixture; a control binding reaction, using non-recombinant bacterial extracts, was run to verify the absence of non-specific complexes with the same mobility as the specific CF2 complex. After binding, the mixture was subjected to mobility shift analysis using non-denaturing gel electrophoresis (9). After exposing the gel on film for 6 to 12 hr at 4°, the free and protein-bound DNA bands were cut out and embedded in an agarose slab (1% in 1×TBE buffer), and electrophoretically transferred (30 min to 1 hr) to a small piece of Hybond M-G paper (Amersham) placed in a slot ahead of the DNA in the direction of its migration. The DNA is retained quantitatively on the paper.

Following electrophoretic transfer, the DNA was steam denatured *in situ* by exposure of the still wet paper to boiling water vapor for 2 min, the paper was left to air dry at 4°, washed for 30 sec by immersing in 50 ml of double distilled water to remove the TBE buffer, blotted on Whatman paper to remove excess liquid, dehydrated by washing for 30 sec in 95% ethanol and allowed to air dry for 2−3 min. Finally, the paper was cut into three pieces, each of which was transferred into a separate tube. Solid phase chemical sequencing was performed at room temperature directly on the DNA immobilized on Hybond M-G paper, using $KMnO_4$ for T, NH_2OH for C and DMS for G (5). For the *potassium permanganate* reaction, 1 ml of double distilled water was added to the tube along with 30 μl of a 2 mg/ml $KMnO_4$ stock solution. After 10 min the liquid was replenished and the reaction was continued for 10 min. For the *hydroxylamine* reaction, 1 ml of a 275 mg/ml (pH 6.0) hydroxylamine stock solution (prepared fresh each week) was added to the tube and incubated for 20 min. For the *DMS* reaction, 1 ml of ammonium formate buffer (50 mM, pH 3.5) was added to the tube along with 7 μl of DMS and the reaction was left to proceed for 6 min (formate buffer should be at room temperature before addition of DMS to avoid precipitation of the chemical). These conditions are preferred because of their convenience (all are at room temperature), and because background considerations are not crucial when looking only for enrichment of minor bases (see below). If desired, slightly lower background can be obtained with less convenient but optimized reaction conditions (10).

Reactions were terminated simply by removing the paper and rinsing it with double distilled water. After another round of rinsing and dehydration in ethanol, the paper was immersed in 1 M piperidine (100 μl) and incubated at 90° for 30 min. During this incubation DNA was eluted by the piperidine, which also cleaved a certain proportion of the modified bases. The paper was then removed and piperidine was eliminated by repeated cycles of freezing and lyophilization (in vacuo centrifugation, Speedvac). The size of the Hybond M-G paper was kept to a minimum to minimize the amount of residue that is formed by paper degradation during the piperidine reaction and persists after lyophilization. When larger paper size could not be avoided, the sample was centrifuged after removal of the paper, and the supernatant was transferred to a new tube. Piperidine and soluble degradation products (probably tetraethylammonium salts) were then removed by ethanol precipitation of the DNA in the presence of 0.3 M sodium acetate and carrier DNA. Phenol extraction at this stage is detrimental, because DNA is partitioned mainly in phenol in the presence of these salts.

RESULTS AND DISCUSSION

Outline and features of the BSSA procedure

The technique is best suited to identifying DNA sequences that bind to a certain protein with high relative affinity. The scheme we follow most frequently (Fig. 1) begins with a sequence encompassing a *low* affinity binding site. Synthetic oligonucleotides of this sequence are prepared, in which mutations are introduced at three positions by including low concentrations of the three non-wild type bases at the corresponding step of synthesis. A mixture of exact duplexes representing both the wild type and mutated sequences is then prepared, by synthesizing the complementary strands in a primer extension reaction. Single end labelling is achieved using end labelled primer or template. The duplexes are then purified by non-denaturing gel electrophoresis and are used in excess as substrate in a binding reaction with the protein of interest. Protein-bound DNA is fractionated from free oligonucleotides by mobility shift in a non-denaturing gel (9). The bound and free fractions are excised from the gel and electrophoretically transferred, rapidly and quantitatively, to Hybond M-G paper, which is compatible with Maxam-Gilbert chemistry. Alternatively the bound fraction may be compared to the starting oligonucleotide mixture.

After the DNA is transferred to paper, it is denatured *in situ* by exposure to boiling water vapor. Following drying, the paper is subjected directly to the base-specific modification and cleavage reactions (see Materials and Methods). One attractive feature of this procedure is its extreme simplicity, which exceeds that of the usual interference (11,12) or missing contact (13) assays. Indeed, these latter procedures can be simplified by adopting the same solid phase reaction protocol that we use for BSSA. Another attractive feature of our procedure is its high sensitivity, which results from the quantitative transfer of the DNA to Hybond M-G paper and its retention there, practically without losses, until the final step of piperidine cleavage; this method of recovering DNA is much more efficient than any method of DNA extraction from acrylamide and precipitation from solution.

In theory two sequencing reactions, one specific for a given purine and the other for the non-complementary pyrimidine, should suffice for following the fate of all four bases when applied to both strands. In practice we use three different reactions for both labelled strands. For cross-checking the results it would be desirable to have available four reactions, but no A-specific protocol has yet been described with Hybond M-G.

The technique is most discriminant when the DNA probe is in huge excess over the protein. Of course these are also the conditions for highest sensitivity with respect to protein availability; this feature, plus the maximization of DNA recovery on the solid support, make the procedure well suited for studying the specificity of binding activities in crude nuclear extracts, before protein purification or cloning. Under DNA excess, the relative proportions of the various sequences in the bound fraction are independent of the amount of protein and depend only on the concentrations of the sequences and their respective affinity constants. If saturating amounts of protein are used, no detectable

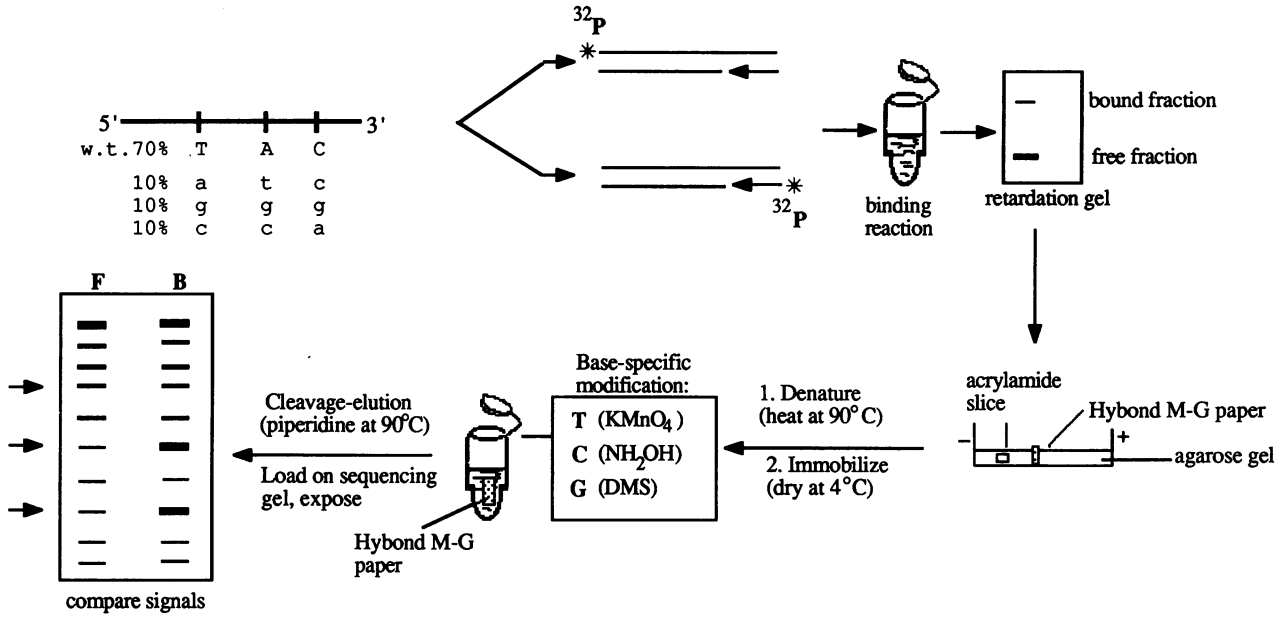

Figure 1. Schematic of experimental strategy. Typically the procedure utilizes synthetic oligonucleotides of a low affinity sequence that are mutagenized in three positions by including low concentrations (10% each) of the other three (non wild type) bases at the corresponding steps of the synthesis (bases shown in lower case letters). The complementary strands are then synthesized by primer extension using the Klenow or T7 DNA polymerase. Single end labelling is achieved using end-labelled primer or template (asterisk). The mixture of double-stranded products is used in excess as a substrate in a binding reaction with a protein of interest. Protein-bound and free oligonucleotides are separated by non-denaturing electrophoresis, and gel slices that include the bands of interest are cut out and embedded in a layer of agarose (1% 1×TBE). The DNA from each gel slice is electrophoretically transferred (30 min to 1 hr) to a piece of Hybond M-G paper placed in a slot in front of it. Double stranded DNA is denatured *in situ* by placing the still wet piece of paper over boiling water vapor for 2 min, immobilized by drying at 4° and subjected directly to base specific modification and cleavage reactions, as described in Materials and Methods. For each of the three reactions, the consequences of each base substitution are evaluated after gel electrophoresis and autoradiography, by comparing band intensities in the bound and free fractions.

differences between bound and non-bound DNAs are expected, especially since the major DNA sequence permits binding (albeit weakly).

The inherent background noise of sequencing reactions makes use of a low affinity binding site convenient as the starting point for the procedure. We have found that it is practical to detect selection reliably both for and against the major (wild type) base, but only for and *not* against the minor (non-wild type) bases, if the latter are each present at less than approximately 10–15% of the total. The lowest threshold for detecting negative selection depends on the nature of the major base, whose non-specific modification causes most of the background, and also on the sequence context (data not shown).

Alternative BSSA schemes are also feasible. For example, if only a single position is varied per oligonucleotide and all four bases are used in equal proportions, it is feasible to detect selection both for and against each of the bases, thus permitting the use of a high affinity site as the starting point. Selection analysis of a single position at a time is necessary in this case, to avoid unacceptably high proportions of doubly mutant sequences. A second possible scheme is to start with a consensus sequence and design degenerate oligonucleotides with the most frequently used bases constant and several surrounding positions varied randomly. The feasibility of such a scheme was recently established independently by Blackwell and Weintraub (14), who developed a binding site selection protocol involving reiterative protein binding of oligonucleotide mixtures and PCR amplification, followed by single strand dideoxynucleotide sequencing. In our hands Maxam-Gilbert sequencing has lower background and is more reliable than the dideoxynucleotide method when used to sequence oligonucleotide mixtures (data not shown).

BSSA on a cloned factor with permissive binding properties

As an example of the applications of our procedure, we report results obtained with a cloned *Drosophila* DNA binding factor, CF2 (chorion factor 2). This zinc finger-containing factor was cloned by binding site screening of an expression cDNA library with a probe derived from the *Drosophila s15* chorion gene promoter (15). Gel retardation, DNase I protection and methylation interference studies have revealed that CF2 expressed in bacteria has multiple binding sites in the *Drosophila s15* (15) and *s36* (P. Tolias, pers. commun.) chorion genes, as well as in several moth chorion gene promoters which are capable of expression in transgenic *Drosophila* (S.A. Mitsialis and J.A.Gogos, unpublished observations). Interestingly, several of the sites that show weak but reliable CF2 binding bear little sequence similarity to the high affinity *s15* site that was used to clone CF2 initially. A typical example is a double stranded oligonucleotide from the upstream region of the *s15* promoter, which contains a low affinity CF2 site as revealed by competition binding experiments (data not shown), but has little similarity to the downstream high affinity site (Fig. 2a). Thus, CF2 appears to be a member of the interesting class of proteins with permissive DNA binding properties, a class that includes the HAP2 yeast factor (16).

An oligonucleotide was designed based on the low affinity segment of *s15*, with degenerate mutations at three positions that are non-identical to the high affinity CF2 site. Fig. 2a summarizes

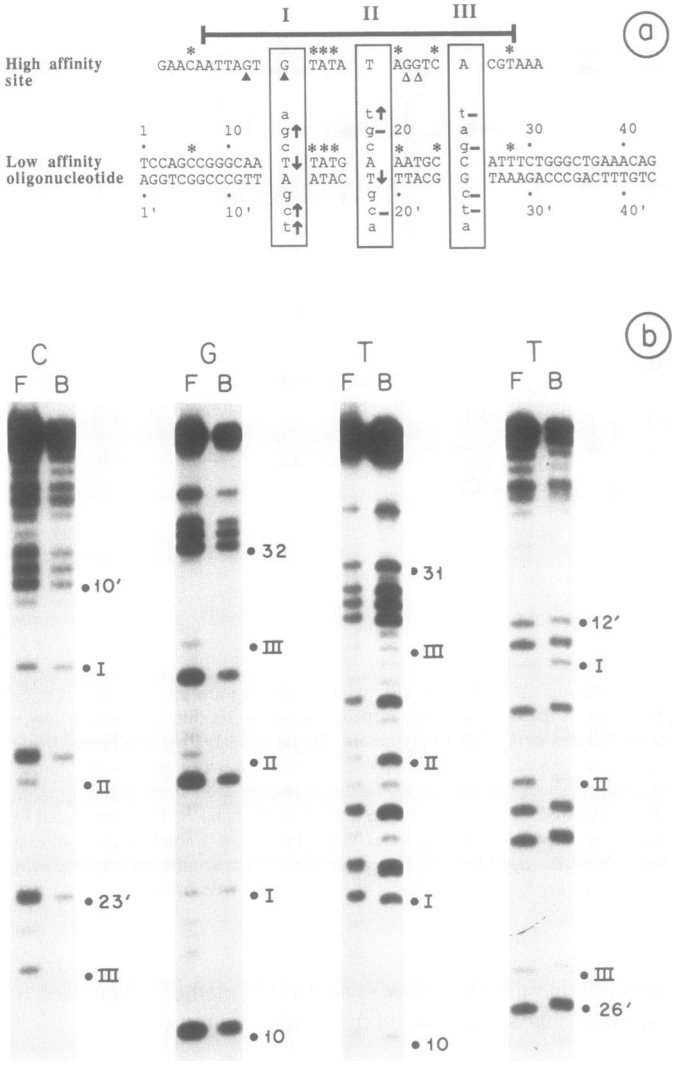

Figure 2. BSS analysis using cloned chorion factor 2 (CF2). **(a)**. Sequence alignment between a high affinity CF2-binding site and the degenerate oligonucleotide used for BSS analysis. Methylation interference and footprint results on the high affinity site (Shea et al. 1990) are indicated as follows: bar (footprint), filled triangle (complete interference), open triangle (partial interference). Asterisks indicate identity between high and low affinity sites. Although alternative alignments are possible, this alignment maximizes identities in the center of the footprint. The low affinity oligonucleotide is shown in double stranded form, and minority bases at degenerate positions (numbered I, II, III) are indicated in lower case. Up and down arrows indicate observed selection for or against the corresponding bases. Absence of detectable selection is indicated by (−). **(b)**. Autoradiogram of sequencing gels showing the results summarized in **a**. Bases are numbered on the right, exactly as in **a**. Plain numbers indicate bases of the positive strand, and primed numbers indicate bases of the negative strand. **C**, **G** and **T** indicate reactions that detect the corresponding bases. **F** is free DNA and **B** is protein-bound DNA.

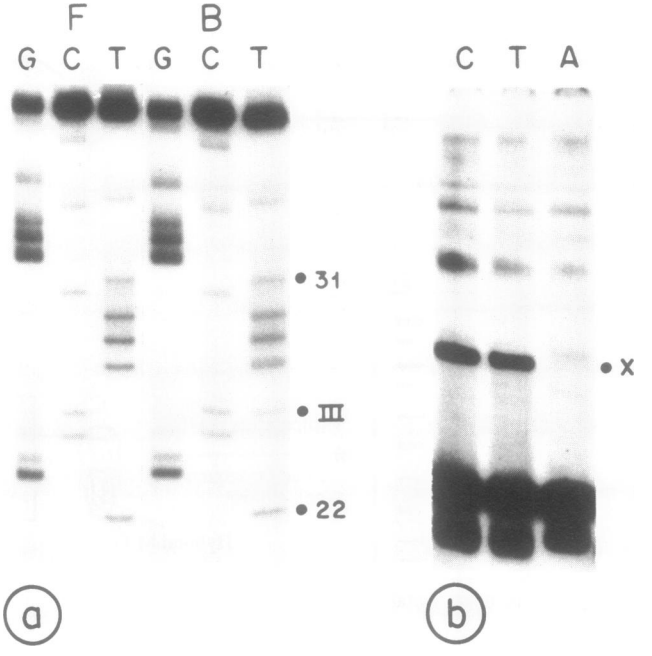

Figure 3. BSS analysis using crude nuclear extract. **(a)**. Autoradiogram of sequencing gel showing the results of BSS analysis with complex X (see text); numbering and other symbols exactly as in Figure **1a**. The oligonucleotide was practically identical to that used for Fig. 2, except that only two sites were matched (bases 16 and 25; see Fig. 2a and the minor bases were 7.5% of the total. **(b)**. Gel retardation experiment using singly mutated oligonucleotides having C, T or A at position III, as indicated. Complex X is indicated by a dot.

the design and the conclusions of the experiment, and Fig. 2b shows typical results. At position I, where the high affinity site has a G in the plus strand scored as a contact residue by methylation intereference (15), and the low affinity site has T, clear selection was evident in favor of the minor G base and against the major T. Selection was also scored in favor of the minor C and T bases in the negative strand. At position II, where the high affinity site has T and the low affinity one has A in the plus strand, selection was noted for the minor T base, but not for G; in the negative strand selection was evident against the major T base, but C appeared to be indifferent. Note that in both positions the decreased intensity of the wild type base served as an internal control against potential sequencing artifacts, and confirmed the positive selection of certain minor bases. It is notable that in both cases mutations that make the low affinity site more similar to the high affinity sequence were selected positively. At position III, no selection was evident within the sensitivity of the assay; either this position is indifferent in terms of binding affinity, or selection operates only against one or two minor bases.

In conclusion, our BSSA experiments identified two positions that are important for CF2 binding and a third one that appears not to be important. In the former two positions increased similarity to the high affinity site resulted in greater binding, as expected, thus validating the technique. In one of the important sites an additional favorable substitution was identified. These results could not have been obtained with typical methylation interference experiments using the low affinity oligonucleotide.

BSSA on a crude nuclear extract

As mentioned above, our procedure is well suited to the mutational analysis of binding site affinities for proteins that are present in small amounts, e.g. in nuclear extracts. A case in point is the factor(s) responsible for complex X, which has not as yet been cloned but is prominent in mobility shift experiments using the *s15* chorion promoter in conjunction with crude ovarian nuclear extracts of *Drosophila* (15). Since this complex is formed on the same oligonucleotide that binds CF2 with low affinity (see Fig. 2a), we investigated how the two sites are related.

The results of Fig 3a indicate that, in contrast to the apparent indifference of CF2 to mutations in position III, complex X selects for the minor T base in the same position. We performed follow-up experiments testing for complex X formation with singly mutated oligonucleotides, and verified that both T and C are acceptable at position III, but A is not (Fig. 3b). Thus, even before cloning the complex X factor,this series of experiments permitted us to document that the CF2 and X binding sites are non-identical, although they may be overlapping.

In conclusion, the technique described here should be a useful complement to the recently described PCR method of binding site selection analysis (14). The PCR method is preferable when highly redundant oligonucleotides are to be used, for example in an exploratory search for an unknown binding site (17); it also has obvious advantages for sensitivity and detection of minor differences in affinity. Our method is attractive because of its extreme simplicity and convenience. It is best suited for checking the behavior of variants of a known sequence, such as may be encountered in homologous genes of related organisms. Furthermore, it is an informative and convenient complement to methylation interference or missing contact analysis: all of these techniques use similar methods (Maxam-Gilbert sequencing, potentially on the same solid support), and thus can be incorporated in an integrated experimental approach, while providing different types of information on protein-DNA interactions.

ACKNOWLEDGEMENTS

We thank A.Mitsialis for useful suggestions, A.Khoury for reviewing the manuscript, E.Fenerjian for secretarial assistance and B.Klumpar for photography. The work was supported by an NIH grant to FCK and fellowships from the Gerondelis Foundation to JAG and GT.

REFERENCES

1. Horwitz, M.S.Z. and Loeb, L.A. (1986) Proc. Natl. Acad. Sci. USA **83**, 7405–7409.
2. Hill, D.E., Oliphant, A.R., Struhl, K. (1987) Methods in Enzymology **155**, 558–568.
3. Oliphant, A.R., Struhl, K. (1987) Methods in Enzymology **155**, 568–582.
4. Singer, V.L., Wobbe, C.R. and Struhl, K. (1990) Genes Dev. **4**, 636–645.
5. Rosenthal, A., Jung, R. and Hunger, H.D. (1987) Methods in Enzymology **155**, 301–331.
6. Tuerk, C. and Gold, L. (1990) Science **249**, 505–510.
7. Green, R., Ellington, A.D., Szostak, J.W. (1990) Nature **347**, 406–408.
8. Sambrook, J., Fritsch, E.F., Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
9. Fried, M. and Crothers, D. (1981) Nucleic Acids Res. **8**, 6505–6525.
10. Gogos, J.A., Karayiorgou, M., Aburatani, H. and Kafatos, F.C. (1990) Nucleic Acid Res. **18**, 6807–6814.
11. Siebenlist, V. and Gilbert, W. (1980) Proc. Natl. Acad. Sci. USA **77**, 122–126.
12. Truss, M., Chalepakis, G., and Beato, M. (1990) Proc. Natl. Acad. Sci. USA **87**, 7180–7184.
13. Brunelle, A. and Schleif, R. (1987) Proc. Natl. Acad. Sci. USA **84**, 6673–6676.
14. Blackwell, T.K. and Weintraub, H. (1990) Science **250**, 1104–1110.
15. Shea, M.J., King, D.L., Conboy, M.J., Mariani, B.D. and Kafatos, F.C. (1990) Genes Dev. **4**, 1128–1140.
16. Pfeifer, K., Prezant, T. and Guarente, L. (1987) Cell **49**, 19–27.
17. Blackwell, T.K., Kretzner, L., Blackwood, E.M., Eisenman, R.N. and Weintraub, H. (1990) Science **250**, 1149–1151.