Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 1247–1250, February 1991
Genetics

Transposon-facilitated DNA sequencing

($\gamma\delta$/polymerase chain reaction)

MICHAEL STRATHMANN, BRUCE A. HAMILTON, CAROL A. MAYEDA, MELVIN I. SIMON, ELLIOT M. MEYEROWITZ, AND MICHAEL J. PALAZZOLO

California Institute of Technology, Division of Biology, 147-75, Pasadena, CA 91125

Communicated by E. B. Lewis, September 26, 1990

ABSTRACT **We describe here a transposon-based DNA sequencing strategy that allows the introduction of sequencing priming sites throughout a target sequence by bacterial mating. A miniplasmid was designed to select against transposon insertions into the vector. Sites of transposon insertion are mapped by the polymerase chain reaction with bacterial overnight cultures providing the templates. A small set of plasmids with transposons spaced several hundred base pairs apart can then be sequenced. Sequencing primers corresponding to the transposon ends allow sequencing in both directions. Thus, the entire sequence of both strands can be easily determined.**

One of the major problems in DNA sequence analysis of large or even moderately sized fragments is how to position unsequenced regions next to known priming sites. A variety of techniques have been developed for this purpose including random shotgun subcloning, unidirectional deletions and subcloning, and the continued synthesis of additional oligodeoxynucleotide primers (1–4). These methods are expensive or require many molecular manipulations.

A number of strategies employ bacterial transposons to generate priming sites within a target DNA sequence (5–10). Several criteria exist for an efficient transposon-based sequencing strategy: (*i*) Mobilization of the transposon must be relatively simple. (*ii*) Selection for transposon insertions into the plasmid as opposed to the bacterial chromosome must be efficient. (*iii*) The transposon must insert into the target sequence and not into the plasmid vector. (*iv*) The transposition sites must be easily mapped to minimize the number of required sequencing reactions. In this paper we describe a transposon-based strategy that meets these criteria.

We employ $\gamma\delta$, which belongs to the Tn*3* family of transposons (11) and which has been used previously in transposon-facilitated strategies (8, 20). The members of this family contain 38-base-pair (bp) terminal inverted repeats and transpose by a replicative mechanism. Donor and target sequences are joined in an intermediate structure termed a cointegrate. The cointegrate, which contains two copies of the transposon, is rapidly resolved by a site-specific recombination system. The resolvase is encoded by the transposon and acts at the 120-bp *res* site located within the mobile element.

$\gamma\delta$ is present on the F factor. Consequently, transposition to a plasmid transiently fuses the F factor and plasmid in a cointegrated structure. This cointegrate can be transferred to a recipient cell by conjugation. Resolution of this structure in the recipient yields the F factor and the plasmid each with a single $\gamma\delta$ insertion (Fig. 1).

This paper describes the use of conjugal transfer of a plasmid to introduce $\gamma\delta$ insertions into a target sequence (12). The target DNA fragment has been subcloned into a minimal plasmid in which nearly all the plasmid sequences are selectable. Under these conditions recovered transpositions almost always contain a $\gamma\delta$ transposon inserted into the target sequences. The sites of insertion can be readily mapped by polymerase chain reaction (PCR) (13). Finally, orientation-specific sequencing primers allow sequence analysis in both directions from the insertion point.

METHODS

Bacterial Mating. Two *Escherichia coli* strains are grown overnight under appropriate antibiotic selection. The donor strain, DPWC (*supE42* Δ*recA*[*Sst II–Eco*RI] *srl*::Tn*10*-[Tet^s], F$^+$), contains the target plasmid, which confers resistance to ampicillin. The recipient, strain JGM, is strain MC1061 (14) that carries Tn5seq1 (9) and is F$^-$ and kanamycin resistant. One-tenth milliliter of each overnight culture and 2 ml of LB medium are combined in a sterile 15-ml tube and incubated on a rotary wheel (30 rpm) for 3 hr or longer at 37°C. One-tenth milliliter of a 100-fold dilution of the mating mixture is plated on an LB-agar plate containing both ampicillin at 100 μg/ml and kanamycin at 50 μg/ml. This plate is incubated overnight at 37°C. Generally, this procedure yields 50–500 Amp^{r}Kan^r colonies. Plating nondiluted and 10-fold dilutions of the mating mixture results in bacterial lawns that consist primarily of bacteria resistant only to kanamycin.

Construction of the Miniplasmid Vector. The miniplasmid vector, used as a transposon target, was constructed by the PCR (14) followed by standard recombinant DNA techniques (15). (*i*) Two oligonucleotides were synthesized that would PCR-amplify the β-lactamase gene and the replication origin of the pUC plasmids. One oligonucleotide, N-AMP (5′-ATGAGACAATAACCCTGA-3′), hybridizes just upstream of the β-lactamase gene (near position 4210 of the pBR322 map) (16). The second oligonucleotide, ORI F (5′-GCCCGGGCGTTGCTGGCGTTTTTCC), is located around pBR322 position 2520 and contains a *Sma* I site. A PCR was performed by using this oligonucleotide pair as primers and pBluescriptKS2 (Stratagene) as template. The PCR product was phosphorylated with T4 polynucleotide kinase, ligated, and introduced into *E. coli* to generate plasmid pOAS. The polylinker from pBluescriptKS2 was then introduced into pOAS to generate pMOB (Fig. 2). To accomplish this construction, the polylinker was first PCR-amplified by using the reverse and universal sequencing oligonucleotides that are commercially available (New England Biolabs). Plasmid pOAS was linearized with *Sma* I and ligated to the polylinker.

DNA Sequence Analysis. Templates for DNA sequence analysis were prepared by using rapid-boil plasmid preparations (17). The sequencing reactions were done with the United States Biochemical T7 Sequenase version 2.0 kit according to the enclosed protocols. The sequencing primers are oligonucleotides complementary to transposon sequences adjacent to the inverted repeat ends. GD1 (5′-CAACGAATTATCTCCTT-3′) will sequence outward from the $\gamma\delta$ end

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: PCR, polymerase chain reaction.

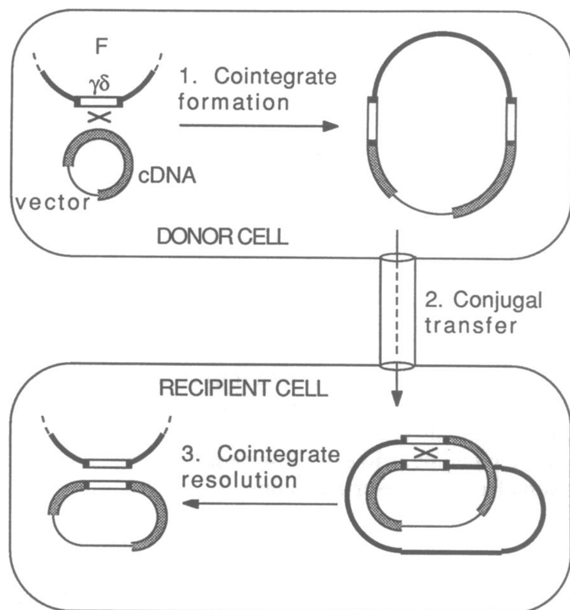

FIG. 1. Transposon mobilization into target DNA sequences. This diagrams the probable mechanism of $\gamma\delta$-transposon insertion into the target sequences by bacterial mating. According to this model (5), a cointegrate is formed in the donor cell between the F factor and the target plasmid. In this cointegrate, both plasmids are flanked by $\gamma\delta$ transposons. After transfer, the cointegrate is resolved, leaving a transposon copy in each plasmid.

closest to the *Sac* I site in the transposon (12). GD2 (5'-TCAATAAGTTATACCAT-3') will sequence outward from the opposite end.

PCR Conditions. All PCRs were done in 40-μl total volume with a light mineral oil overlay. For the construction steps, a three-step protocol was used that included the following: 94°C for 45 sec, 55°C for 45 sec, and 72°C for 2 min. This protocol was then repeated for 10 cycles. For the transposon mapping experiments a two-step PCR protocol was used: 92°C for 20 sec and 72°C for 2 min. One-half microliter of a

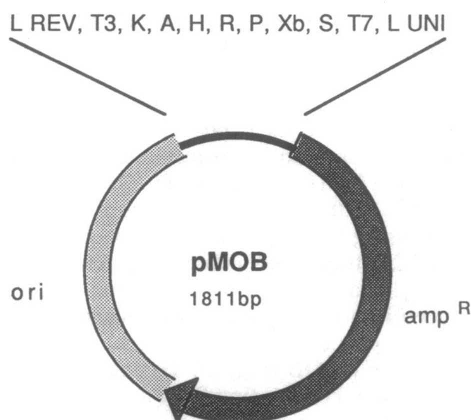

FIG. 2. Map of pMOB, the miniplasmid used for transposon-targeted DNA sequence analysis. This plasmid contains (*i*) a pUC plasmid origin of replication (pUC ori), (*ii*) β-lactamase gene (Amp^r), (*iii*) a multiple cloning site (MCS), (*iv*) bacteriophage RNA polymerase promoters (T3 and T7), (*v*) primer sites for PCR mapping of the transposon insertion sites (LREV, LUNI). The polylinker was derived from a PCR from the plasmid pBluescriptKS2; many restriction sites occur in this polylinker. We have tested for and used only the *Apa* I (A), *Eco*RI (R), *Kpn* I (K), *Pst* I (P), *Hin*dIII (H), *Sac* I (S) and *Xba* I (Xb) sites (these sites are all unique and found only in the MCS region). Because this plasmid was constructed from DNA fragments amplified by PCRs, the DNA sequence may vary somewhat from sequence of the parent plasmids.

bacterial overnight culture was added to the 40-μl reaction mixture and subjected to the PCR for 40 cycles. For each template two PCRs were done. Both PCRs used an oligonucleotide (GDIR; 5'-TTTCGTTCCATTGGCCCTCAAACCCC-3') complementary to the inverted repeat of the $\gamma\delta$ transposon. The second oligonucleotide primer used in one PCR (LREV; 5'-AACAGCTATGACCATGATTACGCCAAG-3') was complementary to a sequence just upstream of the T7 promoter (see Fig. 2). The second oligonucleotide primer used in the other PCR mapping (LUNI; 5'-GTAAAACGACGGCCAGTGAGCGCG-3') is complementary to a region immediately adjacent to the T7 RNA promoter. The PCR buffer was purchased from Cetus.

RESULTS AND DISCUSSION

Efficiency of $\gamma\delta$ Transposition into Target Plasmids. Transposition of $\gamma\delta$ from an F factor to a plasmid is believed to produce a cointegrate that can be conjugally transferred. After conjugation the cointegrate is resolved in the recipient cell (Fig. 1).

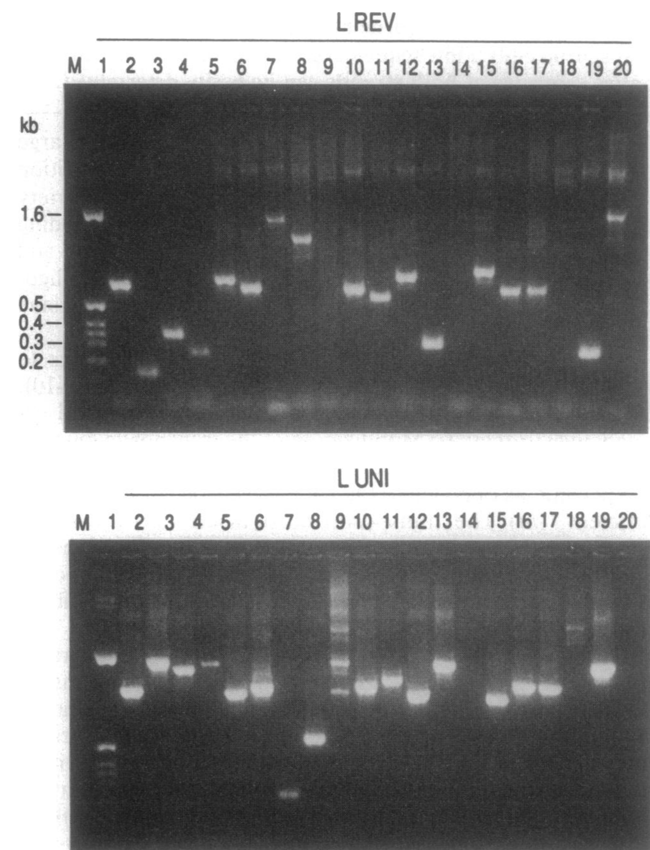

FIG. 3. PCR mapping of transposon-insertion sites. $\gamma\delta$ transposons were introduced by bacterial mating into 1.7-kb cDNA (1H2) that was subcloned into pMOB. Bacterial overnight cultures from each of 20 individual transposition events provided the templates in two separate PCRs. In the first set of PCRs LREV and GDIR were the primers; PCR products were then analyzed by agarose gel electrophoresis (*Upper*, labeled LREV). The second set of PCRs were identical, except that LUNI and GDIR oligonucleotides were the primers (*Lower*); results of these PCRs are labeled LUNI. Lanes in each of these gels are matched so that the upper gel shows the PCR products generated by using the first set of primers, and the bottom gels show the analogous products with the second set of primers. Exact sequences of the oligonucleotides and PCR conditions are indicated in text; interpretation of this data is presented in Fig. 4. As mentioned in text, some PCRs gave ambiguous results: for example, multiple PCR products can be seen in lane 9 (*Lower*), whereas no apparent products can be seen in either reaction in lane 14. Size markers (M) are *Hin*fI digests of pBR322.

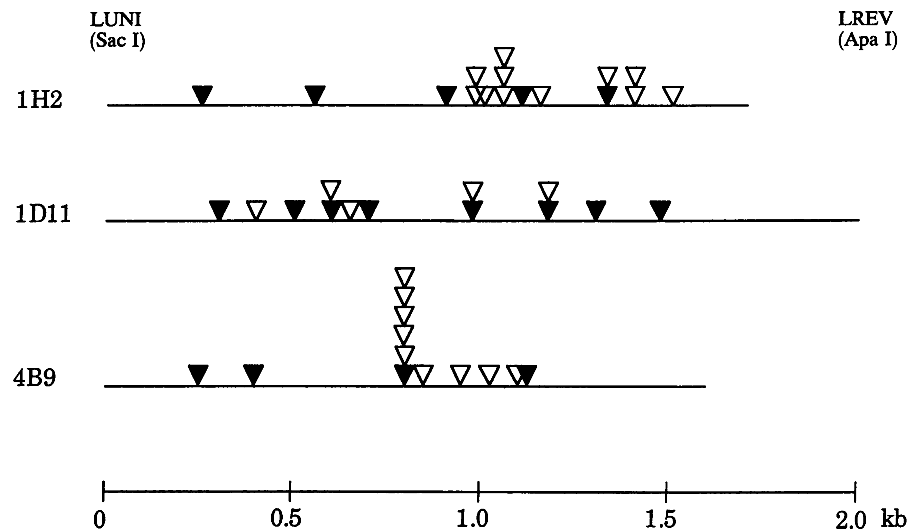

FIG. 4. PCR mapping analysis allows the generation of maps of the $\gamma\delta$-transposon-insertion points. Insertions into three different cDNAs (1H2, 1.7 kb; 1D11, 2.0 kb; and 4B9, 1.6 kb) are shown. These plasmids are all longer versions of cDNA clones originally isolated by Palazzolo *et al.* (19). Each triangle represents the insertion of a single $\gamma\delta$ transposon into a given cDNA. PCR mappings that gave ambiguous results were excluded from the map. The filled triangles represent clones selected for DNA sequence analysis. A complete double-stranded analysis was possible from just these selected clones.

Although the mobilized plasmid is expected to contain a single copy of $\gamma\delta$, some experiments (18) indicated that an alternative method of plasmid transfer is possible in certain *E. coli* strains. Examination of transferred plasmids in these strains showed that only 30% of the plasmids contained transposons after transfer. To determine the fraction of mobilized plasmids that contain $\gamma\delta$, *E. coli* DPWC (a donor strain; see *Methods*) was transformed with a 3-kilobase (kb) plasmid vector conferring ampicillin resistance, pBluescriptKS2 (Stratagene), into which a 1.5-kb *Drosophila* cDNA fragment (4B9; ref. 19) had been subcloned. These cells were then mated to JGM (a kanamycin-resistant recipient strain, see *Methods*). Recipient cells that received the plasmid were selected on plates containing ampicillin and kanamycin. Restriction enzyme analysis of 20 mobilized plasmids indicated that, in each instance, the plasmid contained a $\gamma\delta$ insertion. Furthermore, restriction fragment length polymorphisms in the restriction enzyme digests suggested that the transposons had inserted at different sites.

Construction of a Miniplasmid Sequencing Vector. An important criterion for the successful application of transposition to DNA sequence analysis is that the transposon be forced into the target DNA sequences and not the plasmid. Restriction analysis of the $\gamma\delta$ insertions described above suggested that most insertion events occurred in the vector and not in the insert. We used dideoxynucleotide chain-termination DNA sequencing to identify the transposon-insertion sites in more detail. Of eight clones chosen at random and sequenced, all transposition sites occurred at different locations in the vector.

These experiments suggest that a plasmid, in which insertions into the vector sequences can be selected against, is useful, as then only the transposon insertions in the target can be recovered. Ideally, the vector should contain only an origin of replication, a drug-resistance gene, and a multiple cloning site [the construction of such a miniplasmid, pMOB (Fig. 2), has been described].

To test where transposons insert in this 1.8-kb construct, three different *Drosophila* cDNA molecules (1H2, 1D11, 4B9) (19) were subcloned into this plasmid, and each subclone was separately used as target for $\gamma\delta$ transposition. Twenty clones from each of the mating mixtures were selected and analyzed by restriction mapping. All 60 clones contained a $\gamma\delta$ transposon, and most insertion sites were in or near the target cDNA fragments.

Analysis of Transposon-Insertion Sites by PCR. The ability to rapidly and simply map the sites of insertion is important for minimizing the labor required to sequence a given target DNA fragment. PCR promised to allow such an identification. For this purpose, we synthesized three oligodeoxynucleotides. One (GDIR) matches the inverted repeat found at each end of $\gamma\delta$. The other two oligonucleotides (LREV and LUNI) flank the cloning site of the miniplasmid (see Fig. 2). Two separate PCRs can be used to determine the point of insertion of a given transposon. In one reaction LREV and GDIR are used as primers, whereas LUNI and GDIR are used in the second reaction. In both cases the same plasmid containing a $\gamma\delta$ transposon is the template. Size of the LUNI–GDIR PCR product allows determination of the distance from the transposon-insertion site to the LUNI site at one end of the target fragment, whereas the size of the LREV–GDIR PCR product allows a similar determination of the distance from transposition site to the opposite end of the subclone. Furthermore, the two PCR products should add up to approximately the same size as the fragment subcloned into the miniplasmid.

Such an analysis was performed on the 60 plasmids isolated in the transposon experiments described above. Bacterial overnight cultures provided the templates in two separate PCRs. These reactions were subsequently analyzed by agarose gel electrophoresis. Analysis of 20 transposition events into one clone is shown in Fig. 3. The size of the fragment in each lane delimits the distance of the transposition site from the fixed points in the plasmid.

These experiments allowed us to map the $\gamma\delta$ insertion sites for most of the 60 plasmids containing transposons (Fig. 4). Three conclusions can be drawn from these results. (*i*) Forty-two of the 60 transposition events occurred within the cDNA inserts and could be rapidly localized. (*ii*) The insertion sites were sufficiently dispersed within the target to be useful for DNA sequencing. (*iii*) Eighteen of the PCRs gave apparently anomalous results, including multiple PCR products or the apparent absence of PCR products. The transposon insertions that gave such results could not be placed on the transposition maps by this technique and were not further characterized.

DNA Sequence Analysis By Using Transposon Sequences as Priming Sites. Several plasmids were then chosen for DNA sequence analysis of the different cDNA clones. The plasmids that were sequenced were selected because the transposons were spaced ≈300–400 bp apart (Fig. 4). Because end-specific sequences are found immediately adjacent to the inverted repeats, oligonucleotides complementary to these regions can be used to sequence outward from each end of the transposon (see *Methods* for primer sequences). Thus, it is straightforward to simultaneously obtain complete sequence information from both strands of the target sequence from a relatively small number of plasmid–transposon templates. All three plasmids presented in this paper were completely sequenced by a small number of transposon-containing templates (Fig. 4). In addition, we have sequenced six other cDNA clones with inserts from 1.2 to 2 kb. For each clone, a screen of 20–30 $\gamma\delta$ insertions was sufficient to obtain a subset of transposon insertions that were spaced every 300–400 bp along the cDNA insert.

Some Limitations to this Sequencing Strategy. One major limitation to this strategy is the inability of the PCR to map transpositional events that are relatively distant from the fixed points of the plasmid (LREV and LUNI). Specifically, PCRs typically yield anomalous products on templates in which transposons have inserted >3 kb away from the fixed plasmid point. One potential response to this limitation is the use of strand-switching PCR to map the position of unknown transposon insertions relative to known transposon-insertion points. In other words, two plasmids that contain the same initial insert but have transposons in different locations can be mapped relative to each other in a single PCR. This reaction contains both plasmids as templates but uses only the inverted-repeat oligonucleotide (GDIR) as primer. The PCR product should be the DNA sequence between the two transpositional events, and its size will map the position of the unknown site relative to the known one.

To test this strategy we used various combinations of the plasmids containing the 1H2 cDNA and different transpositional events. PCRs containing different pairwise combinations of 1H2 plasmids as templates and only GDIR as primer resulted in PCR products of the sizes predicted by the results presented in Figs. 3 and 4.

We thank K. VijayRaghavan, Susan Celniker, and Christopher Martin for many helpful discussions. We also thank Douglas Berg for λTnSeq5 and Randy Reed for the sequence of GD1 and GD2. M.J.P. is a Lucille P. Markey Scholar of the Lucille P. Markey Charitable Trust. B.A.H. is supported, in part, by a U.S. Public Health Service Predoctoral National Research Service Award (T32 GM07616). This work was also supported by U.S. Public Health Service Program Project Grant GM40499 (E.M.M.) and by U.S. Department of Energy Grant FG03-89ER60891 (M.I.S.).

1. Barnes, W. M., Bevan, M. & Sons, P. H. (1980) *Methods Enzymol.* **65,** 98–122.
2. Messing, J. (1983) *Methods Enzymol.* **101,** 20–78.
3. Henikoff, S. (1984) *Gene* **28,** 351–359.
4. Strauss, E. C., Kobori, J. A., Siu, G. & Hood, L. E. (1986) *Anal. Biochem.* **154,** 353–360.
5. Guyer, M. (1983) *Methods Enzymol.* **101,** 362–369.
6. Ahmed, A. (1985) *Gene* **39,** 305–310.
7. Adachi, T., Mizuchi, M., Robinson, E. A., Appella, E., O'Dea, M. H., Gellert, M. & Mizuchi, K. (1987) *Nucleic Acids Res.* **15,** 771–784.
8. Liu, L., Whalen, W., Das, A. & Berg, C. M. (1987) *Nucleic Acids Res.* **15,** 9461–9469.
9. Nag, D. K., Huang, H. V. & Berg, D. E. (1987) *Gene* **64,** 135–145.
10. Barrett, B. K. & Berget, P. B. (1989) *DNA* **8,** 287–295.
11. Grindley, N. F. (1988) *The Recombination of Genetic Material,* ed. Low, K. B. (Academic, New York), pp. 283–360.
12. Guyer, M. (1978) *J. Mol. Biol.* **126,** 347–365.
13. Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. & Arnheim, N. (1985) *Science* **230,** 1350–1354.
14. Casadaban, M. J. & Cohen, S. N. (1980) *J. Mol. Biol.* **138,** 179–207.
15. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
16. Sutcliffe, J. G. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **43,** 77–90.
17. Holmes, D. S. & Quigley, M. (1981) *Anal. Biochem.* **114,** 193–197.
18. Goto, N., Shoji, A., Horiuchi, S. & Nakaya, R. (1984) *J. Bacteriol.* **159,** 590–596.
19. Palazzolo, M. J., Hyde, D. R., VijayRaghavan, K., Mecklenburg, K., Benzer, S. & Meyerowitz, E. M. (1989) *Neuron* **3,** 527–539.
20. Strausbaugh, L. D., Bourke, M. T., Sommer, M. T., Coon, M. E. & Berg, C. M. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 6213–6217.