Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 1585–1588, March 1985
Biochemistry

Base composition-independent hybridization in tetramethylammonium chloride: A method for oligonucleotide screening of highly complex gene libraries

(DNA melting)

WILLIAM I. WOOD, JANE GITSCHIER, LAURENCE A. LASKY, AND RICHARD M. LAWN

Departments of Molecular Biology and Vaccine Development, Genentech, Inc., 460 Pt. San Bruno Boulevard, South San Francisco, CA 94080

Communicated by Gary Felsenfeld, October 31, 1984

ABSTRACT **An oligonucleotide hybridization procedure has been developed that eliminates the preferential melting of A·T versus G·C base pairs, allowing the stringency of the hybridization to be controlled as a function of probe length only. This technique, which uses tetramethylammonium chloride, is especially helpful whenever a highly complex library is screened with a pool of oligonucleotide probes, which usually vary widely in base composition. The procedure can also be applied advantageously whenever an exact match to an oligonucleotide probe is desired, such as in screening for clones having as little as a single-base alteration generated by *in vitro* mutagenesis.**

Short oligonucleotide probes [range, 14–20 base pairs (bp)] are commonly used to screen libraries of cloned DNA for genes of interest (1–3). Typically, these probes are pools representing all possible codon choices for a short amino acid sequence. Although this method has been successful, there is considerable uncertainty in the hybridization conditions because the binding of the oligonucleotides depends on two factors: (*i*) the length of the hybrid formed and (*ii*) the G·C content of the probe. Empirically determined formulas allow for estimation of the oligonucleotide dissociation temperature (T_d) (4); however, these methods can be unsatisfactory when screening with pools of oligonucleotides. Although the length of the probes in the pool is constant, the individual probes differ considerably in G·C content, making suitably stringent and selective hybridization conditions difficult to find for all members of the pool. Thus, a large number of false positives can occur when screening highly complex libraries for genes of low abundance.

We describe here the use of tetramethylammonium chloride (Me_4NCl) in the hybridization of oligonucleotide probes to eliminate the dependence of T_d on the G·C content of the probe, reducing the problem to a simple dependence on length of the hybrid. Tetraalkylammonium salts were found some years ago to bind to A+T-rich polymers of DNA (5) and have been used to abolish the preferential melting of A·T versus G·C base pairs for fragments of DNA (6). Me_4NCl binds selectively to A·T base pairs, thus displacing the dissociation equilibrium and raising the melting temperature. At 3.0 M Me_4NCl, this displacement is sufficient to shift the melting temperature of A·T base pairs to that of G·C base pairs (6). For the melting of long DNA, this shift results in a remarkable sharpening of the melting profile. Natural DNAs that melt over a range of 5 to 10°C in the presence of Na^+ melt within 1°C in Me_4NCl (6–8). The data presented here show the utility of hybridizations with Me_4NCl when oligonucleotide probes are used. The T_d values of probes from 11 to >1000 bp have been determined so that hybridization conditions for probes of various lengths can be chosen easily. This method is applicable to a variety of circumstances in which an exact match to a probe is desired.

METHODS

DNA Synthesis, Binding, and Labeling. In this procedure, nitrocellulose filters are used routinely with DNA spotted directly or bound (9) from bacteriophage λ or M13 plaques or from plasmid-containing bacterial colonies. Filters removed from bacteriophage or bacterial colony plates are treated with 0.5 M NaOH/1 M NaCl, which is neutralized with 1 M Tris·HCl, pH 7.5/1.5 M NaCl. DNA spots are made by denaturing the sample with 0.3 M NaOH, neutralizing with 0.6 M Tris·HCl, pH 7.4/1.5 M NaCl, and immediately spotting on nitrocellulose filters previously soaked in 3.0 M NaCl/0.3 M Na citrate (20× NaCl/Cit) and dried. The DNA-containing filters are baked for 2 hr at 60–80°C in a vacuum oven. Oligonucleotide probe pools were synthesized from trimers by the triester method (10). Pools 9.3 and 9.4 were end-labeled with T4 polynucleotide kinase and [γ-^{32}P]ATP (9).

Hybridization and Me_4NCl Wash Procedure. In this procedure, an initial nonstringent hybridization with radiolabeled probe is followed by washing with 3.0 M Me_4NCl to control the stringency of the hybridization. The filters are prehybridized in 6× NaCl/Cit/50 mM sodium phosphate, pH 6.8/5× Denhardt's solution (1× Denhardt's solution is albumin/polyvinylpyrrolidone/Ficoll, each at 0.2 mg/ml) containing boiled sonicated salmon sperm DNA at 0.1 mg/ml for 4–16 hr at 37°C. The filters are hybridized overnight at 37°C in the same solution plus dextran sulfate at 100 mg/ml and the pool of end-labeled oligonucleotide probes (each at 180 pM; 560,000–1,680,000 cpm/pmol) (11). [For a 17-mer probe this is 1 μg/liter (100–300 cpm/pg). For a pool of 16 probes, the probe concentration is 16 μg/liter.] The filters are rinsed three times with 6× NaCl/Cit at 4°C and washed twice for 30 min with 6× NaCl/Cit at 4°C. The filters are then rinsed with the Me_4NCl wash solution at 37°C to remove the NaCl/Cit [the NaCl/Cit must be substantially removed because Na^+ will compete for Me_4N^+ binding (12)] and washed twice for 20 min with the Me_4NCl wash solution at the desired temperature. In screening experiments, we typically use a temperature 2–4°C below the T_d shown in Fig. 3. The wash temperature needs to be well controlled (±1°C); suitable heat exchange can be obtained only in a shaking or circulating water bath. The Me_4NCl wash solution is 3.0 M

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.

Abbreviations: bp, base pair(s); T_d, dissociation temperature; Me_4N^+, tetramethylammonium ion; Et_4N^+, tetraethylammonium ion; NaCl/Cit, 0.15 M NaCl/0.015 M Na citrate.

Me_4NCl/50 mM Tris·HCl, pH 8.0/2 mM EDTA containing $NaDodSO_4$ at 1 mg/ml. Me_4NCl is purchased from Aldrich and made up as a nominal 5 M stock solution. Since Me_4NCl is hydroscopic, the actual molar concentration (c) is determined from the refractive index (η) by the formula $c = (\eta - 1.331)/0.018$ (12). The concentration needs to be fairly precise since the T_d varies with the Me_4NCl concentration (6).

The prehybridization and hybridization described above are used for oligomers in the range 14–50 nucleotides. For shorter oligonucleotides, a room temperature hybridization is preferable. For oligonucleotides of $\geq$ 50 nucleotides, the prehybridization and hybridization are as above except that formamide at 400 ml/liter is included at 37°C. In the range 46–75 bp, we have compared hybridization with and without formamide; the two give equal results. Nitrocellulose filters become fragile in washes at 85°C and above. If frequent use is made of high-temperature washes, it is possible to use 2.4 M tetraethylammonium chloride (Et_4NCl) rather than 3.0 M Me_4NCl. Et_4NCl has the same desirable effect on DNA melting as Me_4NCl [at least with long DNA (6, 7)], but the entire melting profile is shifted 33°C lower. Although with Et_4NCl the melting temperature of short oligonucleotides is inconvenient (below room temperature), it could be useful for 50- to 200-bp probes.

RESULTS AND DISCUSSION

The results of hybridization and melting of four oligonucleotide probes in the commonly used 6× NaCl/Cit system (1, 2, 11, 13) are compared with those in 3.0 M Me_4NCl in Fig. 1. These probes are 17-mers, and their hybridization is determined by the stretch of contiguous match, the hybridizing length, rather than by the percentage homology because even a single internal mismatch in probes of this length lowers the T_d by 5–10°C (refs. 11 and 13; unpublished observations). A series of DNA spots was hybridized under nonstringent conditions and washed at temperature intervals of 2 or 3°C. As shown in Fig. 1*A*, the melting in 6× NaCl/Cit does not depend linearly on the hybridizing length. In particular, the probe with the 13-bp hybridizing region melts some 5°C higher than the 15-bp probe and at about the same temperature as the 16-bp probe. This is because the probe with the 13-bp hybridizing region is 69% G·C while the 15- and 16-bp probes are 40 and 44% G·C. These melting results in 6× NaCl/Cit contrast with those in 3.0 M Me_4NCl (Fig. 1*B*) in which the probes melt according to their hybridizing length. These results illustrate that the preferential melting of A·T base pairs is abolished in oligonucleotide hybridizations in 3.0 M Me_4NCl, as previously shown for long DNA (6). The T_d values derived from the data in Fig. 1 are plotted as a function of hybridizing length in Fig. 2.

To extend the utility of the method over a wide range of probe lengths, the T_d values in 3.0 M Me_4NCl were determined as a function of length for 18- to 1374-bp restriction fragments of pBR322. These fragments were end-labeled, hybridized to a series of pBR322 DNA dots, and washed at various temperatures with 3.0 M Me_4NCl. The T_d, length, and G·C content of these fragments are shown in Table 1. When plotted as a function of length, the T_d values lie on a smooth curve from 45 to 93°C in spite of the widely varying G·C content (31–66%) of the probes (Fig. 3). This further demonstrates the lack of dependence of the T_d on the G·C content. The T_d is essentially independent of length above 200 bp, and the limiting T_d of 93°C agrees with that reported for long DNA (6, 7). From either Fig. 3 or Table 1 one can determine the hybridization wash temperature so that the stringency can be based solely on the probe length.

The utility of hybridization in Me_4NCl is perhaps best illustrated by the following example based on our own experience: A pool of sixteen 17-mers representing all possible codons for the protein sequence Glu-Cys-Trp-Cys-Gln-

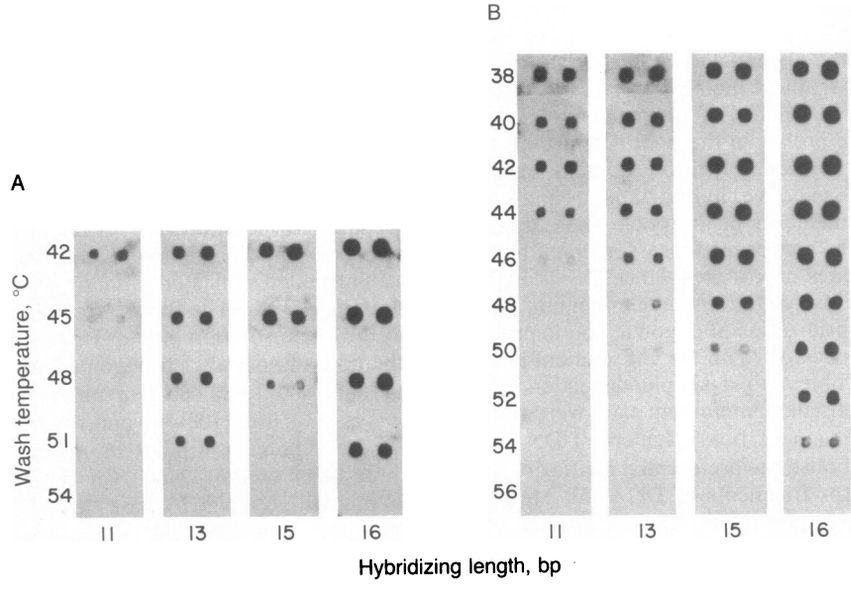

FIG. 1. Hybridization of oligonucleotides of various hybridizing lengths washed in 6× NaCl/Cit or 3.0 M Me_4NCl at increasing temperatures. Spots containing 125–250 ng of bacteriophage λ20a or λ21a DNA were made on nitrocellulose membranes, and these DNA spots were hybridized to ^{32}P-labeled probe pools 9.3 or 9.4. Pool 9.3 contains sixteen 17-mer probes of the sequence 5′ G-C-$\frac{C}{T}$-T-G-$\frac{A}{G}$-C-A-C-C-A-$\frac{A}{G}$-C-A-$\frac{C}{T}$-T-C. Pool 9.4 contains sixteen 17-mers with the sequence 5′ T-A-$\frac{C}{T}$-T-G-$\frac{C}{T}$-T-T-C-C-A-$\frac{A}{G}$-A-A-$\frac{C}{T}$-T-C. λ20a and λ21a are clones isolated from a bovine genomic library (14) by conventional hybridization with 6× NaCl/Cit washes to pools 9.3 and 9.4 (unpublished data). Pool 9.3 has a contiguous match of 11 bp to λ21a, A-*C-T-T-G-A-C-A-C-C-A-A*-A-C-*T*-C-A (the nucleotides in italic type are those of the λ21a sequence that match the probe), and a contiguous match of 13 bp to λ20a, *G-C-T-T-G-G-C-A-C-C-A-G-C*-T-*T*-G-C. Pool 9.4 has a 15-bp match to λ21a, *T*-G-*T-T-G-C-T-T-C-C-A-G-A-A-T-T-C*, and a 16-bp match to λ20a, A-*A-C-T-G-C-T-T-C-C-A-A-A-A-C-T-C*. Duplicate hybridized spots were washed at the indicated temperatures in 6× NaCl/Cit containing $NaDodSO_4$ at 1 g/liter twice for 10 min (*A*) or in 3.0 M Me_4NCl (*B*) as described in *Methods*. Spots are shown after autoradiography overnight at −70°C with DuPont Lightning Plus intensifying screens.

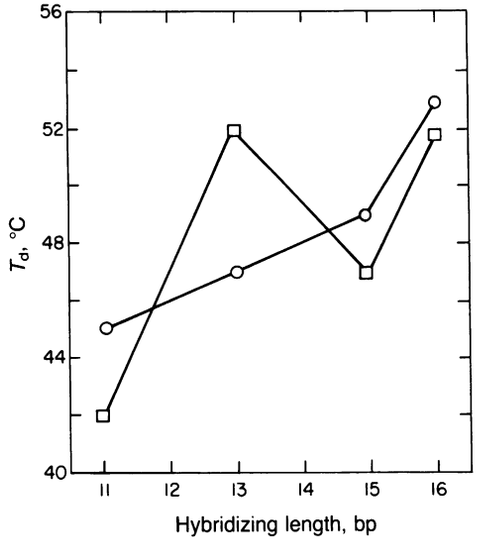

FIG. 2. T_d of oligonucleotide probes as a function of hybridizing length in 6× NaCl/Cit (□) or 3.0 M Me_4NCl (○). T_d values were determined visually from the data in Fig. 1 as the temperature at which half the intensity remained and plotted as a function of hybridizing length. T_d values are used rather than melting temperature (T_m) values, which are determined under conditions of thermodynamic equilibrium.

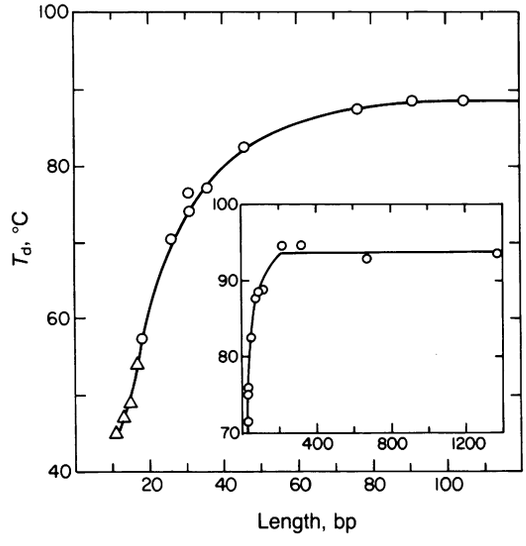

FIG. 3. T_d in 3.0 M Me_4NCl as a function of probe length. A series of pBR322 DNA spots was made on nitrocellulose membranes, and these were hybridized and washed with 3.0 M Me_4NCl. Washes of the duplicate spots were carried out at 3°C intervals. Hybridization probes of various lengths were made by cutting pBR322 DNA with *Sau*3AI, treating with alkaline phosphase, labeling with T4 polynucleotide kinase and [γ-^{32}P]ATP, and isolating the labeled fragments from a polyacrylamide gel (9). These double-strand probes were boiled before hybridization. From an autoradiogram of the hybridized spots (analogous to that shown in Fig. 1), a T_d was determined and plotted for each fragment (○). Also shown are melting data in Me_4NCl from Fig. 2 (△).

Ala was synthesized (probe 9.3; see Fig. 1). This pool consists of oligonucleotides ranging in G·C content from 47 to 71%. In 6× NaCl/Cit, a temperature that allows hybridization of the probe having the lowest G·C content will also allow hybridization of regions as short as 12 or 13 bp in the probes having the highest G·C contents. When screening a highly complex library, such as those derived from mammalian genomes, this can lead to a large number of false positives. The expected number of hybridizing sequences per haploid genome at random, N, is approximately

$$N = C(2)(n - h + 1)(p)/4^h,$$

where C = complexity of the genome in bp, n = probe length, h = contiguous hybridization match, and p = pool size (see *Appendix* for derivation). Thus, for a pool of sixteen 17-mers where matches as short as 13 are allowed ($C = 3.3 \times 10^9$, $n = 17$, $h = 13$, $p = 16$), the expected number of clones at random is 8000. These clones represent a large number of false positives. However, with Me_4NCl, a hybridization temperature can be selected in which only 17- and perhaps some 16-bp matches are allowed (52–54°C). In this case, the expected number of random clones is only 6 (or 49 for 16-bp matches), a number small enough to be searched through for the correct clone by DNA sequence analysis or by a precise melting in Me_4NCl as described below. As noted in the *Appendix*, the application of the formula for N to hybridization in 6× NaCl/Cit is not quite correct because all matches of length 13 (not just those of high G·C content) are counted. Thus, perhaps only 1000–4000 positives would be found with 6× NaCl/Cit hybridization. In any event, the number of false positives due to probes of high G·C content can be reduced significantly by using Me_4NCl.

Table 1. Probe G·C content and T_d in 3.0 M Me_4NCl

Hybridizing length, bp	G·C content, %	T_d in 3.0 Me_4NCl, °C
11	45	44–45
13	69	47
15	40	49–50
16	44	53–54
18	44	57–58
27	63	70–71
31	58	74 (76)
36	53	77
46	54	82–83
75/78	48/42	87–88
91	66	88–89
105	31	88–89
207	55	94–95
317	65	94–95
665	46	93
1374	55	93–94

Probes having hybridizing lengths of 11–16 bp are described in the legend to Fig. 1. Those 18–1374 bp long are *Sau*3AI fragments of pBR322.

In addition to screening highly complex libraries, the Me_4NCl procedure may be adapted advantageously in a variety of other situations. For example, it can be used to select the best candidates for DNA sequencing. If a number of candidate clones have been obtained, those that match the probe most closely can be found by plating out replicas of the clone or spotting the DNA, hybridizing to an oligonucleotide probe, and washing with 3.0 M Me_4NCl at increasing temperature in the manner of Fig. 1. We have used six duplicate spots of the DNA and washed at 2°C intervals to find the clones that match the probe exactly, thus eliminating unnecessary DNA sequencing. In this way, the match with the probe can be determined to ±1 bp in the 14- to 20-nucleotide range. This technique can be especially precise if standards of known hybridizing length are included at the same time. Another application of this method has been in screening of M13 clones that had been mutagenized *in vitro* (15). In this procedure, a 20-mer having one or more centrally located base changes is synthesized and used to generate M13 plaques that are screened for an exact match

with the same oligonucleotide as the probe. Since even a single-base-pair internal mismatch decreases the T_d of the hybrid 5 to 10°C (11, 13), the correct mutant can be found by screening in Me_4NCl, using the wash conditions shown in Fig. 3. Furthermore, for different oligonucleotides, empirical determination of a suitable wash temperature is unnecessary. Finally, the method can be used to select one particular gene out of a closely related gene family. For example, a clone for the γ subunit of mouse nerve growth factor was selected from a family of closely related kallikrein-like cDNA clones by hybridization in Me_4NCl (16). In this case, the probe was a pool of thirty-two 14-mers based on protein sequence data.

In summary, hybridization in Me_4NCl is useful whenever an exact match with an oligonucleotide probe is desired. This method is being used routinely for oligonucleotide hybridization to bacteriophage λ or M13 plaques, plasmid-containing bacterial colonies, and DNA spots. While we have not yet used the procedure for genomic blot hybridization with oligonucleotide probes, it should be useful in detecting single-base mismatches in hybridizations to genomic DNA (17). Here again, the hybridization conditions would not need to be determined empirically.

APPENDIX

Derivation of N, the Approximate Number of Sequence Matches Expected per Haploid Genome at Random. Let C = complexity of the genome in bp, n = probe length, h = contiguous hybridizing length, and p = pool size; then, considering the genome as random sequence, $N = C(2)f$, where f is the frequency of some sequence at random. The frequency times the number of bp, C, times 2 (the sequence of interest could be on either strand) is N, the number of sequence matches. The frequency of a particular sequence of length n, where every nucleotide matches, is $1/4^n$. If only h contiguous bp of a probe of length n match, then the frequency of sequence match is $1/4^h$ times the number of contiguous length h matches in the probe. For $n = 17$ and $h = 15$, the number of contiguous matches is $(n - h + 1) = 3$. Thus, for a single probe of length n, where contiguous matches of length h are allowed, the frequency of a sequence match is $(n - h + 1)/4^h$. If a pool of p such probes is used, then $f = p(n - h + 1)/4^h$ and $N = C(2)(n - h + 1)p/4^h$.

The expected number of clones at random, N, given by this formula is about the number we have found in screening a bovine genomic library by using probe pools 9.3 and 9.4 (Fig. 1). We screened 750,000 plaques of a library representing 3.5 haploid genomes with 3.0 M Me_4NCl washes at 50°C and found 80 and 130 positives per haploid genome for the two pools. For these conditions, we would expect 16- and 17-bp matches with perhaps some matches as short as 15 bp. For $C = 3.3 \times 10^9$, $n = 17$, $p = 16$, and $h = 16$ or 15, the formula gives $N = 49$–295 expected random positives, approximately the number found. Application of this formula, based on the contiguous match, is especially appropriate for hybridizations in Me_4NCl where the match length can be controlled. Use of the formula with hybridization in 6× NaCl/Cit is less exact because the number of hybridizing sequences in a pool is affected by the G·C content.

This formula can also be used to estimate the number of random matches to be found in cDNA or other libraries. For this purpose, C can be set to the total number of bp screened (assuming the clones are independent), and the number of hybridizing clones expected at random can be calculated. Clearly, this will overestimate the number of randomly hybridizing clones for most cDNA libraries because some clones are represented many times. However, it can still be used to obtain a rough approximation of what to expect for a particular sized library, oligonucleotide probe, and hybridization stringency.

Note Added in Proof. Recently we have successfully used the Me_4NCl procedure for genomic blot hybridization. The frequency in the population of a single base polymorphism in the factor VIII gene (18) was determined by the hybridization of duplicate lanes of human DNA to two 21-mer probes that differ only at a single central residue. Wash conditions based on Fig. 3 (58°C) revealed a single band with no lane-specific background and clearly showed that 7 of 11 individual DNAs have one sequence while the remaining 4 have the other.

1. Goeddel, D. V., Yelverton, E., Ullrich, A., Heyneker, H. L., Miozzari, G., Holmes, W., Seeburg, P. H., Dull, T., May, L., Stebbing, N., Crea, R., Maeda, S., McCandliss, R., Sloma, A., Tabor, J. M., Gross, M., Familletti, P. C. & Pestka, S. (1980) *Nature (London)* **287,** 411–416.
2. Suggs, S. V., Wallace, R. B., Hirose, T., Kawashima, E. H. & Itakura, K. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 6613–6617.
3. Szostak, J. W., Stiles, J. I., Tye, B.-K., Chiu, P., Sherman, F. & Wu, R. (1979) *Methods Enzymol.* **68,** 419–428.
4. Suggs, S. V., Hirose, T., Miyake, T., Kawashima, E. H., Johnson, M. J., Itakura, K. & Wallace, R. B. (1981) *ICN–UCLA Symp. Dev. Biol.* **23,** 683–693.
5. Shapiro, J. T., Stannard, B. S. & Felsenfeld, G. (1969) *Biochemistry* **8,** 3233–3241.
6. Melchior, W. B. & von Hippel, P. H. (1973) *Proc. Natl. Acad. Sci. USA* **70,** 298–302.
7. Chang, C.-T., Haim, T. C., Hutton, J. R. & Wetmur, J. G. (1974) *Biopolymers* **13,** 1847–1858.
8. Britten, R. J., Cetta, A. & Davidson, E. H. (1978) *Cell* **15,** 1175–1186.
9. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
10. Crea, R. & Horn, T. (1980) *Nucleic Acids Res.* **8,** 2331–2348.
11. Wallace, R. B., Johnson, M. J., Hirose, T., Miyake, T., Kawashima, E. H. & Itakura, K. (1981) *Nucleic Acids Res.* **9,** 879–894.
12. Orosz, J. M. & Wetmur, J. G. (1977) *Biopolymers* **16,** 1183–1199.
13. Wallace, R. B., Shaffer, J., Murphy, R. F., Bonner, J., Hirose, T. & Itakura, K. (1979) *Nucleic Acids Res.* **6,** 3543–3557.
14. Leung, D. W., Capon, D. J. & Goeddel, D. V. (1984) *Biotechnology* **2,** 458–464.
15. Seeburg, P. H., Colby, W. W., Capon, D. V., Goeddel, D. V. & Levinson, A. D. (1984) *Nature (London)* **312,** 71–75.
16. Ullrich, A., Gray, A., Wood, W. I., Hayflick, J. & Seeburg, P. H. (1984) *DNA* **3,** 387–392.
17. Conner, B. J., Reyes, A. A., Morin, C., Itakura, K., Teplitz, R. L. & Wallace, R. B. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 278–282.
18. Wood, W. I., Capon, D. J., Simonsen, C. C., Eaton, D. L., Gitschier, J., Keyt, B., Seeburg, P. H., Smith, D. H., Hollingshead, P., Wion, K. L., Delwart, E., Tuddenham, E. G. D., Vehar, G. A. & Lawn, R. M. (1984) *Nature (London)* **312,** 330–337.