



inNovations

NEWSLETTER OF NOVAGEN, INC. • ADVANCED PRODUCTS AND PROTOCOLS FOR MOLECULAR BIOLOGY RESEARCH

NUMBER 3
JUNE 1995

FEATURE ARTICLE

IN THIS ISSUE

A Novel Strategy for Directional Cloning of Random Primed cDNA

Barbara Morris, Beth Hammer and Robert Mierendorf — Novagen, Inc.

The first synthetic step in the creation of a cDNA library is reverse transcription of the mRNA to produce a cDNA copy. Like other DNA polymerases, RNA-dependent DNA polymerases are not capable of initiating cDNA synthesis *de novo*; *i.e.*, the enzymes require a short piece of DNA or RNA as a primer on which to add nucleotides. Most current priming strategies (oligo(dT)-containing primers or random hexamers) do not permit directional cloning while maintaining unbiased sequence representation. Novagen's newly introduced Directional RH Random Primer cDNA Library Construction System combines the directional cloning advantages of oligo(dT) with the sequence independent priming of random hexamers.

Directional Cloning with Extended Primers

Oligo(dT)-containing primers have been most frequently used for cDNA synthesis (1). Since oligo(dT) anneals to the 3' poly(A) tail carried by most eukaryotic mRNA species, the resulting cDNA molecules share a common (dA:dT) tract on this end. In recent years, oligo(dT) primers have been engineered to include specific restriction sites (2). This relatively short sequence does not affect the annealing of the primer to the mRNA, so the cDNA is copied with the intended restriction site at the original 3' end of the coding sequence. After second strand synthesis, a blunt ended linker molecule containing a second restriction site (or a partially double stranded adaptor containing a protruding end compatible with a

second restriction site) is ligated to both ends of the cDNA. The site encoded by the linker is now on both ends of the cDNA, but the 3' end of the cDNA has the site introduced by the primer internal to the site added by the linker. Following the ligation step, the product is digested with both restriction enzymes (or just the enzyme which cuts the primer if a partially double stranded adaptor was used in the ligation). The result of the digestion is a population of cDNA molecules which all have one restriction site on the 5' end and a different restriction site on the 3' end with respect to their coding sequences.

Sequence-specific primers are used less commonly because they require some knowledge of the target gene to be cloned. Generally, sequence-specific primers can be designed to incorporate a restriction site at the 3' end of the cDNA coding sequence in a directional strategy, comparable to the oligo(dT) extension strategy described above. Specific sequence-primed libraries have a very low complexity, but they greatly increase the probability of cloning the target sequence.

The Directional Linker Approach

An alternative approach to directional cloning using "plain," unextended oligo(dT) originally described by Meissner *et al.* (3) and later modified by Novagen, takes advantage of the sequence recognized by *Hind* III. Following oligo(dT)-primed cDNA synthesis, an *EcoR* I/*Hind* III "directional" linker is ligated to the cDNA and the prod-

continued on page 2

ARTICLES

A Novel Strategy for Directional Cloning of Random Primed cDNA

1

λ SCREEN™: A High Efficiency cDNA Cloning Vector for Protein Expression, Purification and Subtractive Probe Synthesis

5

Frequencies of Restriction Sites in Murine cDNA

6

pET TRX Fusion System for Increased Solubility of Proteins Expressed in *E. coli*

7

Novagen on the Internet

10

Q&A

Commonly Asked Questions About cDNA Libraries

12

NEW PRODUCTS

13

ANNOUNCEMENTS

16

continued from page 1

ucts digested with *EcoR* I and *Hind* III. The sequence of the directional linker 5' d (GCTTGAATTCAAGC) includes an internal *EcoR* I site flanked on each side by 4 of the 6 bases defining a *Hind* III site. Regardless of the sequence to which this directional linker ligates, the internal *EcoR* I site will always be present. However, *Hind* III can only cut the linker if it ligates next to a d(AA):d(TT) dinucleoside base pair. In an oligo(dT)-primed strategy, a *Hind* III site will always be generated at the 3' end of the cDNA coding sequence after ligation to this directional linker. For cDNAs having d(TT) at the 5' end relative to their coding sequence (statistically, one in 16 molecules), linker addition will yield *Hind* III sites at both ends. However, because the 5' ends of cDNA are heterogeneous due to the lack of processivity of reverse transcriptases, cDNA products from every gene will be represented in the library.

Oligo(dT) Priming and the 3' Bias

Whereas oligo(dT)-primed synthesis theoretically should give rise to full-length cDNA, in practice the yield of such molecules is less than 50% (4). Obtaining a full-length cDNA depends on the quality of the reverse transcriptase and on the length and secondary structure of the mRNA to be copied. Although high RNase H activity in enzyme preparations can cause loss of sequence representation (5), it is also believed that reverse transcriptases tend to become disengaged from long mRNA templates or those with extensive secondary structure before traversing their entire length (1). Statistically, the 3' ends of mRNAs are more likely to be copied than sequences toward the 5' end in oligo(dT)-primed libraries because the reverse transcription reaction always commences from the point at which the primer anneals. The resulting cDNA population is therefore biased toward the 3' ends of coding sequences. In practical terms the effect is particularly noticeable with long mRNAs and results in few or no complete cDNAs for certain genes in the library.

It should also be noted that oligo(dT) can also efficiently prime from internal dA stretches within a given mRNA. In these cases, the resulting cDNA may be missing sequences that lie to the 3' side of the dA-rich priming site in the mRNA. Although internal priming usually has little effect on the overall 3' bias of oligo(dT)-primed libraries,

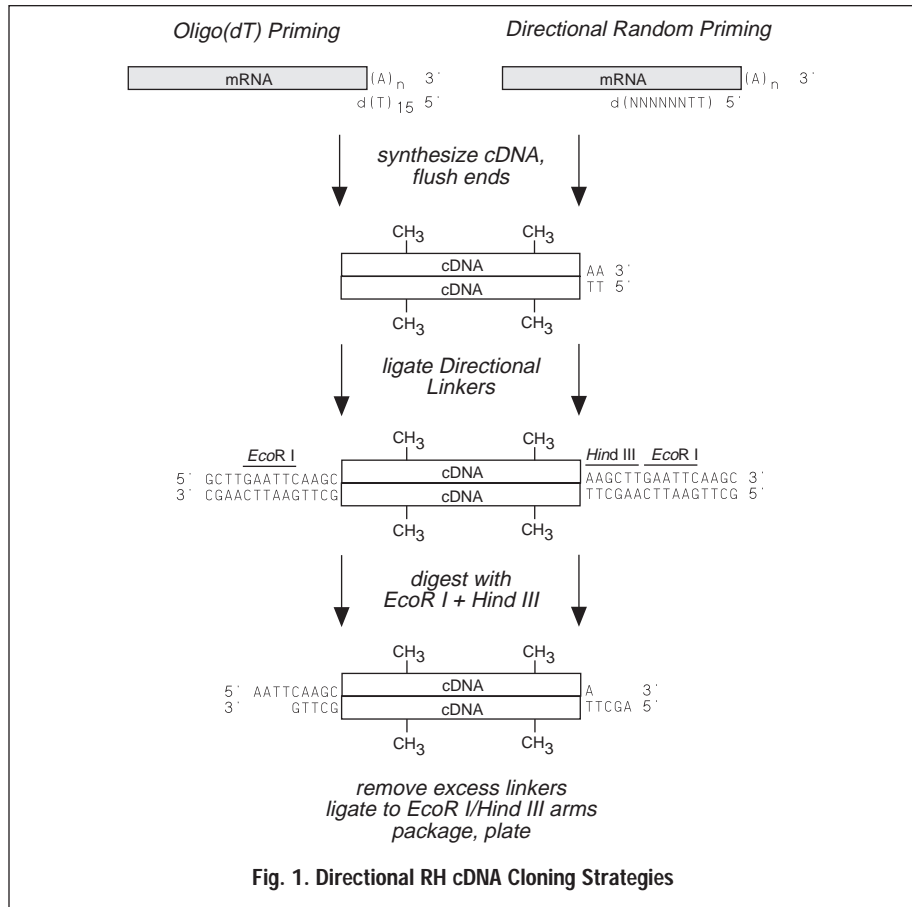


Fig. 1. Directional RH cDNA Cloning Strategies

it can be significant when analyzing individual cDNA species.

Some mRNAs have either a poly(A) tail which is too short to anneal to the oligo(dT) primer or no poly(A) tail at all (6-9). The inability of some mRNAs to be primed with oligo(dT) makes it essential to construct random primed libraries in these cases.

Directional Random Priming

Random primed libraries are not biased toward 3' ends because the primer mix of all possible hexamers promotes initiation of cDNA synthesis at any point on the mRNA (10-12). However, the lack of a common sequence on the 3' end of the cDNA coding sequences synthesized with conventional random hexamers makes them unsuited for directional cloning. The use of random hexamers extended by six or more bases comprising a restriction site, analogous to the oligo(dT) directional primer strategy, has been unsuccessful for the efficient construction of large libraries. Since in this case the extension sequence is as long or longer than the random domain, annealing of the primer may be energetically unfavorable.

To address this problem, Novagen has developed a directional random primer that consists of a random hexamer sequence preceded by two defined bases at the 5' end.* Having only two bases at this position allows proper priming at random sites while sufficiently defining the 5' end of the first strand to allow directional cloning of the cDNA into two different restriction sites. In the configuration described here, the two defined bases are d(TT). This primer mixture allows the identical strategy as that described above for directional cloning with oligo(dT) primer and the *EcoR* I/*Hind* III directional linker (see Fig. 1). Since the two 5' terminal bases are d(TT) defined by the primer, digestion of the cDNA with both *EcoR* I and *Hind* III (after ligation to the directional linker) will produce molecules with an *EcoR* I site at the 5' end and a *Hind* III site at the 3' end relative to the coding sequence.

Model for Testing the Strategy

To test our directional random priming strategy, we used a 3256nt synthetic "mRNA" prepared from a pSCREEN™-1

* patent pending

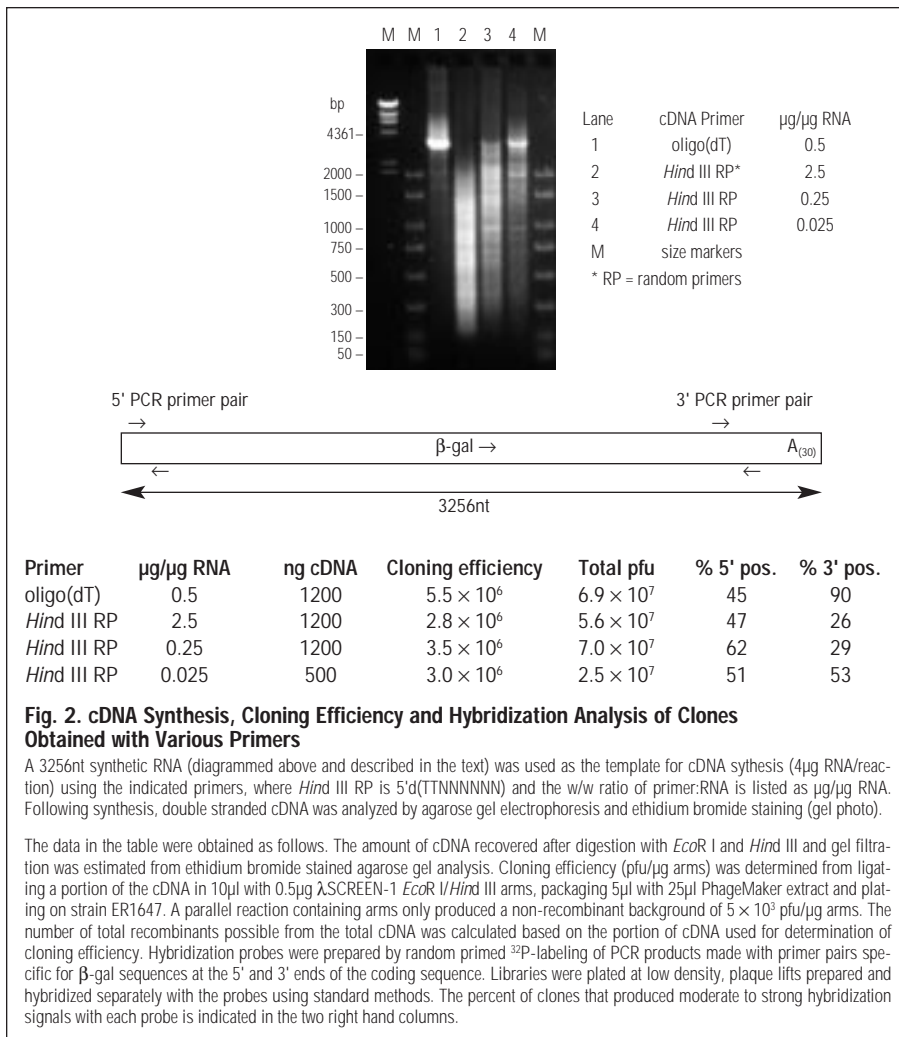


Fig. 2. cDNA Synthesis, Cloning Efficiency and Hybridization Analysis of Clones Obtained with Various Primers

A 3256nt synthetic RNA (diagrammed above and described in the text) was used as the template for cDNA synthesis (4µg RNA/reaction) using the indicated primers, where *Hind* III RP is 5'd(TTNNNNNN) and the w/w ratio of primer:RNA is listed as µg/µg RNA. Following synthesis, double stranded cDNA was analyzed by agarose gel electrophoresis and ethidium bromide staining (gel photo).

The data in the table were obtained as follows. The amount of cDNA recovered after digestion with *Eco*R I and *Hind* III and gel filtration was estimated from ethidium bromide stained agarose gel analysis. Cloning efficiency (pfu/µg arms) was determined from ligating a portion of the cDNA in 10µl with 0.5µg λSCREEN-1 *Eco*R I/*Hind* III arms, packaging 5µl with 25µl PhageMaker extract and plating on strain ER1647. A parallel reaction containing arms only produced a non-recombinant background of 5×10^3 pfu/µg arms. The number of total recombinants possible from the total cDNA was calculated based on the portion of cDNA used for determination of cloning efficiency. Hybridization probes were prepared by random primed ³²P-labeling of PCR products made with primer pairs specific for β-gal sequences at the 5' and 3' ends of the coding sequence. Libraries were plated at low density, plaque lifts prepared and hybridized separately with the probes using standard methods. The percent of clones that produced moderate to strong hybridization signals with each probe is indicated in the two right hand columns.

recombinant with SP6 RNA polymerase. The construct produces an RNA containing a 3.1kb coding sequence for *E. coli* β-galactosidase and a stretch of 30 consecutive A residues at the 3' end. This RNA was used to compare the directional oligo(dT) and random primer approaches described above for cDNA cloning efficiency, sequence representation, and directionality.

Cloning Efficiency

Fig. 2 summarizes experiments in which the test RNA was used to construct libraries with four different priming conditions, including oligo(dT) and three random primer:RNA ratios. Unlike oligo(dT) priming strategies, random priming allows control over the average cDNA size by adjusting primer:template ratios. At high ratios, priming events become more frequent and smaller cDNAs are produced than at low ratios. Smaller inserts may be desirable when screening libraries for functional

domains or other applications. The gel shown in Fig. 2 illustrates the effect of adjusting random primer:template ratios on cDNA size ranges.

The table in Fig. 2 shows that under all priming conditions, high cloning efficiencies were obtained with λSCREEN-1 *Eco*R I/*Hind* III arms and PhageMaker® packaging extracts. These efficiencies corresponded to 2.5 - 7.0×10^7 primary recombinants.

Sequence Representation

Sequence representation was determined by plaque hybridization of probes corresponding to 165nt and 160nt of 5' and 3' ends of the β-gal coding sequence, respectively. The table in Fig. 2 shows that, as expected, the highest number of clones that hybridized with a 3' probe (90%) was obtained with oligo(dT). Since the non-recombinant background was much less than 1% as determined by control ligations, close to

10% of the clones appeared to represent molecules lacking the 3' end. While we did not analyze these clones further, it is possible that at least some of them could represent cDNA molecules initiated at internal oligo(A) sequences. Two stretches of six consecutive dA residues occur in the β-gal coding sequence, which could theoretically serve as priming sites for oligo(dT). In contrast, roughly half of the clones were positive with the 5' probe, indicating that the oligo(dT) library had a significant 3' bias.

The hybridization analysis also shows that all of the random primer libraries contained a higher representation of 5' coding sequences, ranging from 1:1 to 2:1 vs. 0.5:1 (5' ends:3' ends) seen for oligo(dT). The highest 5' representations were obtained with medium and high primer:RNA ratios, which is consistent with the model that more cDNA chains would be initiated at internal sites and be elongated toward the 5' end of the RNA template under these conditions. Therefore, with high random primer:RNA ratios it appears possible to obtain a significant 5' bias in libraries, which may be useful for a variety of applications.

Directionality

To determine the orientation of cDNA inserts, we amplified plasmid subclones positive for hybridization with the 3' β-gal probe with a combination of insert and vector-specific primers. Fig. 3 shows the priming strategy and examples of PCR

continued on page 4

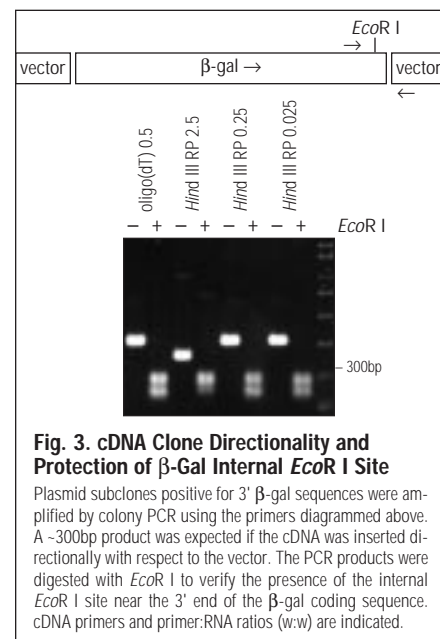


Fig. 3. cDNA Clone Directionality and Protection of β-Gal Internal *Eco*R I Site

Plasmid subclones positive for 3' β-gal sequences were amplified by colony PCR using the primers diagrammed above. A ~300bp product was expected if the cDNA was inserted directionally with respect to the vector. The PCR products were digested with *Eco*R I to verify the presence of the internal *Eco*R I site near the 3' end of the β-gal coding sequence. cDNA primers and primer:RNA ratios (w:w) are indicated.

results. All clones tested from each of the four libraries produced the expected ~300bp species using this primer pair, indicating that all cDNAs were inserted in the correct orientation.

Protection of Internal *EcoR* I and *Hind* III Sites

Since the cloning strategies required digestion with *EcoR* I and *Hind* III to cleave the directional linker, a means of protecting adventitious restriction sites within cDNA was necessary. A widely used method for protection of *EcoR* I sites is treatment of cDNA with *EcoR* I methylase prior to linker addition (13). It is also possible to protect against *Hind* III digestion using *Alu* I methylase, since the *Alu* I recognition sequence d(AGCT) is contained within that of *Hind* III (3). To minimize the steps required for cloning, we chose as an alternative the incorporation of 5-methyl dCTP in place of unmodified dCTP for both first and second strand synthesis. We have previously shown that randomly chosen clones from mouse and *Drosophila* embryo cDNA libraries synthesized by this strategy carry both *EcoR* I and *Hind* III sites (14), which suggests that at least some of these sites are protected using the cloning strategy. However, we were able to obtain more quantitative data as described below.

The β -gal coding sequence contains one *EcoR* I site 54bp upstream from the C-terminus (see Fig. 3). This site is present within the PCR product made with β -gal and T7 terminator primers. PCR products from all clones tested from oligo(dT)-primed and *Hind* III directional random primed libraries were digested with *EcoR* I, indicating that this site was efficiently protected during cDNA synthesis and initial cloning.

Since the 3256nt β -gal test RNA lacks an internal *Hind* III site, we chose another test RNA and method that allowed analysis of *Hind* III digestion following double stranded cDNA synthesis. This 1.3kb synthetic RNA contains a *Hind* III recognition sequence approximately 300nt from its 3' end, so digestion was easily assessed by agarose gel electrophoresis. In this case we were analyzing cDNA directly, so internal sites should resist digestion. In the above *EcoR* I analysis we were testing for the presence of sites within cDNA clones that had been propagated in a non-methylating host

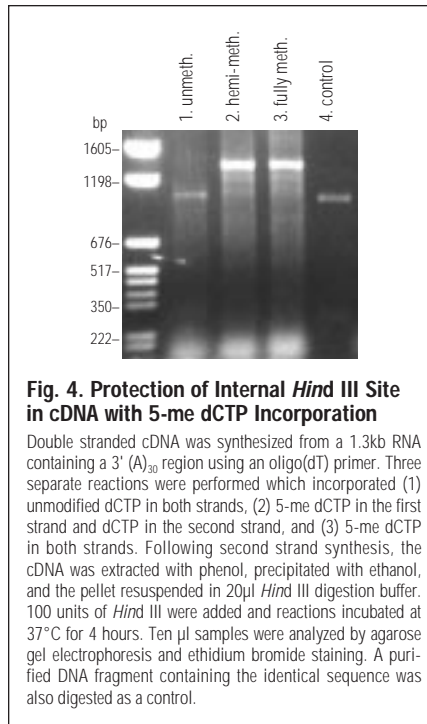


Fig. 4. Protection of Internal *Hind* III Site in cDNA with 5-me dCTP Incorporation

Double stranded cDNA was synthesized from a 1.3kb RNA containing a 3' (A)₂₀ region using an oligo(dT) primer. Three separate reactions were performed which incorporated (1) unmodified dCTP in both strands, (2) 5-me dCTP in the first strand and dCTP in the second strand, and (3) 5-me dCTP in both strands. Following second strand synthesis, the cDNA was extracted with phenol, precipitated with ethanol, and the pellet resuspended in 20 μ l *Hind* III digestion buffer. 100 units of *Hind* III were added and reactions incubated at 37°C for 4 hours. Ten μ l samples were analyzed by agarose gel electrophoresis and ethidium bromide staining. A purified DNA fragment containing the identical sequence was also digested as a control.

and thus no longer retained their methylation patterns.

Fig. 4 shows the results of *Hind* III digestion of cDNA synthesized in the presence of 5-methyl dCTP. Whereas the unmethylated cDNA control was completely digested, hemi-methylated and fully methylated cDNA (prepared using 5-me dCTP incorporation into the first strand only or both strands, respectively) were completely resistant to digestion with excess *Hind* III. These data verify our earlier analysis of *Hind* III sites in individual clones isolated from a variety of libraries constructed by the method described here.

Summary

We have developed a simple method for directional cloning of random primed cDNA. As described, the method is also compatible with oligo(dT) priming for cloning into *EcoR* I and *Hind* III sites of suitable vectors (e.g., λ SCREEN-1; see accompanying article). We have demonstrated that this method produces high cloning efficiencies, with inserts directionally cloned and protected from digestion at internal *EcoR* I and *Hind* III sites.

It should also be noted that the directional random priming strategy described here for *EcoR* I and *Hind* III can be applied to many other restriction enzyme combinations. By changing the sequence of the

directional linker and of the two 5' bases on the random primer, virtually any enzyme combination can be used for cloning. Several examples are presented in Table 1.

Table 1. Design of Several Directional Random Primer/Directional Linkers for Cloning into Specific Restriction Sites

Directional Random Primer	Directional Linker	Enzymes (5', 3')
TTNNNNNN	GCTTGAATTC AAGC	<i>EcoR</i> I, <i>Hind</i> III
AGNNNNNN	CGAGGAATTCCTCG	<i>EcoR</i> I, <i>Xho</i> I
GCNNNNNN	GGCCGCGAATTCGCGGCC	<i>EcoR</i> I, <i>Not</i> I
GCNNNNNN	GGCCGCGTCGACGCGGCC	<i>Sal</i> I, <i>Not</i> I

Kits for rapid construction of cDNA libraries using the directional random primer and oligo(dT) approaches described here are available from Novagen (see p. 13).

References

1. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) "Molecular Cloning, A Laboratory Manual," Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
2. Palazzolo, M.J. and Meyerowitz, E. M. (1987) *Gene* **52**, 197-206.
3. Meissner, P.S., Sisk, W.P. and Berman, M.L. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4171-4175.
4. Kimmel, A.R. and Berger, S.L. (1987) *Meth. Enzymol.* **152**, 307-316.
5. Krug, M.S. and Berger, S.L. (1987) *Meth. Enzymol.* **152**, 316-325.
6. Greenberg, J.R. (1976) *Biochemistry* **15**, 3516-3522.
7. Adesnik, M. and Darnell, J.E. (1972) *J. Mol. Biol.* **67**, 397-406.
8. Houdebine, L.M. (1976) *FEBS Lett.* **66**, 110-113.
9. Rutledge, R.G., Seligy, V.L., Cote, M.J., Dimock, K., Lewin, L.L. and Tenniswood, M.P. (1988) *Gene* **68**, 151-158.
10. Haymerle, H., Herz, J., Bressan, G.M., Frank, R. and Stanley, K.K. (1986) *Nucleic Acids Res.* **14**, 8615-8624.
11. Koike, S., Sakai, M. and Muramatsu, M. (1987) *Nucleic Acids Res.* **15**, 2499-2513.
12. Bonaldo, P., Russo, V., Buccioti, F., Bressan, G.M. and Colombatti, A. (1989) *J. Biol. Chem.* **264**, 5575-5580.
13. Huynh, T.V., Young, R.A. and Davis, R.W. (1985) in "DNA Cloning" (D.M. Glover, ed.), Vol. 1, p. 49. IRL Press, Washington, D.C.
14. Novagen 1991-1995 catalogs.