Nucleic Acids Research, Vol. 19, No. 22 6123–6127

Isolation of a large number of novel mammalian genes by a differential cDNA library screening strategy

Christer Höög
Department of Molecular Genetics, Karolinska Institutet, Box 60 400, S-104 01 Stockholm, Sweden

Received September 13, 1991; Revised and Accepted October 8, 1991

EMBL accession nos X58486, X61801–X61899 (incl.)

ABSTRACT

As part of the ongoing human and mouse genome projects, the aim of this study was to isolate novel, previously uncharacterized, genes from mouse testis. Two approaches were compared for their effectiveness in isolating novel genes: random, *vs* differential, complementary DNA (cDNA) cloning methods. In the differential approach, only the cDNA clones containing rare sequences (as determined by preliminary clone hybridization) are further analyzed; in the random approach, cDNA clones are isolated at random from the cDNA library. More than two hundred cDNA clones altogether were analyzed, using a PCR-mediated amplification and sequencing strategy. A comparison of these sequences to nucleic acid and protein sequence databases, revealed that 84% of the isolated rare cDNA clones represented new, previously uncharacterized mouse genes. In contrast, less than 63% of the cDNA clones isolated at random from cDNA libraries, contained novel genes. Thus, the probability of isolating new, previously uncharacterized, mammalian genes from cDNA libraries can be markedly improved by focusing efforts on clones containing rare sequences.

INTRODUCTION

The human genome project is an international effort to determine the location and structure of all genes active in a number of organisms (1, 2). The human and the mouse genome each contain an estimated 50,000 to 100,000 genes (3), of which only a few thousand have been sequenced (4). Theoretically, it should be possible to isolate a majority of the remaining uncharacterized genes from a set of cDNA libraries representing different tissues and developmental stages (5). In one version of this strategy, a small stretch of each cDNA clone is partially sequenced to create a unique expressed sequence tag (6). These expressed sequence tags (ESTs) can be used to map the expressed part of the genome, to verify the significance of predicted open reading frames in genomic sequencing projects, and to aid in the isolation of candidate disease genes.

The progress of cDNA sequencing projects depends on the rate with which new, previously uncharacterized, cDNA clones can be isolated. A typical mammalian cell contains between 10,000 and 30,000 unique mRNA sequences (7). These mRNA sequences can be divided into a few frequency classes, where the relative size and complexity of eac ı class varies according to the tissue (8–10). A pilot study by \dams *et al.* (6), where cDNA clones were selected at random from a set of human brain cDNA libraries, revealed that only 46% of the analyzed cDNA clones contained novel human sequences. Instead, a large fraction of the isolated cDNA clones represented a small group of highly abundant genes, or contained only polyadenylated sequences (6).

The identification and subsequent removal of abundantly expressed, as well as noninformative, cDNA clones prior to sequence analysis, would maximize the representation of rare cDNAs selected from a cDNA library. In this study abundant and rare cDNA clones have been divided into two groups, using a differential hybridization strategy. The likelihood of identifying new, previously uncharacterized genes, within both groups of cDNAs were investigated. It was found that 84% of the rare cDNA clones represented novel, not previously described, mouse genes, whereas only 12% of the cDNA clones from the abundant group, were found to be novel.

MATERIALS AND METHODS

Construction of cDNA library

Total RNA was purified from testis of prepubertal mice, 6–12 days of age, or from a variety of somatic tissues, as previously described (11). Poly $(A)^+$ RNA was purified from total RNA using oligo (dT) chromatography (10). To synthesize cDNA appropriate for directional cloning, 5 μg of testis poly $(A)^+$ RNA was primed with a oligo $(dT)_{12-16}$ primer having a Xho1 site in the 3′ end, using a commercial cDNA synthesis kit from Stratagene. After ligation of Eco R1 linkers onto the cDNA, it was digested with Eco R1 and Xho 1, and finally ligated into an Eco R1, Xho1-cut λ uni-ZAP vector from Stratagene. Ligated DNA was packaged *in vitro*, using a commercial extract from Stratagene. The cDNA library contained 1×10^6 primary plaques, and was amplified to a titer of 2.6×10^{10} pfu/ml.

Isolation of cDNA clones

1000 plaques were screened at a density of 200 plaques per filter (Hybond C, extra strong, Amersham), after infection of Xl1 Blue cells (Stratagene). The cDNAs to be used as probes were synthesized from a mixture of poly $(A)^+$ RNAs prepared from liver, kidney and heart, or made from poly $(A)^+$ RNA prepared from prepubertal testis (6–12 days of age), using an oligo

$(dT)_{12-16}$ primer and AMV reverse transcriptase (Promega Biotec). The cDNA was labeled using the random priming technique (12). Typical labeling reactions, using a commercial labeling kit from Promega Biotec and alpha-^{32}P-dATP (3000 Ci/mmol, New England Nuclear), yielded cDNA probes with a specific activity of about 1×10^9 cpm/μg. Duplicate filters were hybridized in 50% formamide, 6×SET, 0.5% SDS and 10 mg/ml of denatured herring sperm DNA at 42°C, using 2×10^6 cpm/ml of probe (11). The filters were washed at 65°C for 30 min, in 0.2×SSC, 0.5% SDS, (11). The autoradiographic film was exposed for four days before it was developed. Selected cDNA clones were picked and suspended in SM (11).

DNA amplification

Complementary sequences to the M13 primer and the M13 reverse primer are located on the opposite sides of the cDNA cloning box in the λ uni-ZAP vector (Stratagene). The polymerase chain reaction (13) was performed using a Perkin-Elmer 'GeneAmp' kit, according to a slightly modified procedure from Applied Biosystems. Briefly, 4 μl of the phage suspension was combined with the following reagents in a final volume of 40 μl: 2 pmol each of the M13 primer and the M13 reverse primer, 20 μM of each deoxynucleotide, PCR buffer (10 mM Tris HCl, pH 8.3, 50 mM KCl, 0.001% gelatin and 2 mM $MgCl_2$), and 1.5 U of AmpliTaq polymerase. Samples were overlaid with mineral oil and placed in a preheated (95° C) microtiter-plate thermal cycler (Techne). After a 3 min pre-incubation at 95°C, 27 cycles of PCR were performed, with the following parameters: 95°C (1 min), 50°C (1 min), and 72°C (2 min). 15 μl samples from each reaction were analyzed on a 1% agarose gel.

Table 1. Comparison of sequence data to the EMBL nucleic acid database.

Rare group	no.	%
No database match	90/120	75
Database match - mouse	19/120	16
Nuclear genes	9	
Mitochondrial genes	-	
Ribosomal genes	1	
Repeated sequences	9	
Database match - other species	11/120	9
Rat	5	
Human	6	
Abundant group	**no.**	**%**
No database match	5/51	10
Database match - mouse	45/51	88
Nuclear genes	11	
Mitochondrial genes	34	
Ribosomal genes	-	
Repeated sequences	-	
Database match - other species	1/51	2
Human	1	

Partial 5' end sequences from 171 cDNA clones were compared to the EMBL database, using the program FASTA (17). 120 of these cDNA clones were taken from the rare group and 51 cDNA clones were taken from the abundant group. The mouse ESTs for which no matches were found were submitted to the EMBL database (accession numbers X61801 to X61891).

DNA sequencing

Dideoxy-terminator sequencing reactions were performed with fluorescent dye-labeled primers and cycle sequencing kits purchased from Applied Biosystems. A modified cycle sequencing protocoll for PCR templates was used (Applied Biosystems). Briefly, the A and the C reaction tubes each contained 1.5 μl of the PCR-amplified DNA, 0.8 pmol of fluorescent sequencing primer, 2 μl of deoxy/dideoxy nucleotide mixes, cycle sequencing buffer, and 1 U of AmpliTaq polymerase, in a final volume of 13 μl. The G and the T reactions were performed at twice that volume. The deoxy/dideoxy nucleotide mixes, as well as the sequencing buffer, were taken from a commercial kit (Applied Biosystems). All samples were overlaid with mineral oil and placed in a preheated (95°C) microtiter-plate thermal cycler (Techne). 15 cycles of PCR were performed, where each cycle consisted of 95°C (30 sec), 55°C (30 sec), and 72°C (1 min). A second set of 15 cycles was subsequently performed, consisting of 95°C (30 sec) and 72°C (1 min). The A, C, G and the T reactions were all pooled, precipitated with ethanol and resuspended in deionized formamide : 50 mM EDTA (pH 8.0) 5:1 (v/v). The samples were denatured, then processed on a 373A automated DNA sequencer (Applied Biosystems); the primary sequence data was edited using the program package SeqEd on a Macintosh IIcx (Applied Biosystems). Edited sequence data was analyzed using software from the Genetic Computer Group (14), on a VAX computer.

RESULTS

Isolation of cDNA clones

mRNA was isolated from testis of 6−12 day old prepubertal mice, converted into cDNA using reverse transcriptase and oligo $(dT)_{12-16}$ primers, and cloned directionally into the λ uni-ZAP vector (Stratagene). About 1000 uni-ZAP cDNA clones were distributed onto 5 plates. To mark the abundantly expressed cDNA clones (10), duplicate filters were hybridized with a mixture of labeled cDNA prepared from liver, kidney and heart,

Table 2. EST similarities in the EMBL nucleic acid database.

Clone name (TSG)	Heterologous match (species)	Similarity (length)	Number	Redundancy
20	La autoantigen (H)	65 % (293 bp)	HSLAANT	1
41	Lipoamide dehydrogenase (H)	86 % (198 bp)	HSLPDH	1
46	Collagen (III) α-1 (R)	92 % (209 bp)	RRCOL3A1	3
48	Ret transforming gene (H)	88 % (120 bp)	HSTYKRET	1
70	Cytochrome oxidase VIIa (H)	79 % (426 bp)	HSCOX7AL	1
147	Poly A binding protein (H)	92 % (279 bp)	HSPOLYAB	1
150	MHC protein (H)	90 % (231 bp)	HSMHB123	1
151	Ribosomal protein L18a (R)	94 % (217 bp)	RRRPL18A	2
215	Ribosomal protein L23 (H)	89 % (290 bp)	HSL23MR	1

14 ESTs were found to have similarities to a set of heterologous genes from humans (H) and rats (R). The similarity between the heterologous matches found are shown, with the length of the sequence overlap within parentheses. TSG 70 was sequenced from both ends of the cDNA insert. TSG 215 belongs to the abundant group, whereas the other clones originate from the rare group. The column labeled number, marks the accession number of the matched sequences taken from the EMBL database. The above ESTs were submitted to the EMBL database and were given the accession numbers; TSG20, X61893; TSG41, X61894; TSG46, X61892; TSG48, X61897; TSG70, X58486; TSG147, X61895; TSG150, X61898; TSG215, X61899.

or with labeled cDNA prepared from prepubertal testis (6–12 days of age). An identical set of cDNA clones, representing about 30% of the total number of plaques, reacted with both probes. The cDNA clones labeled in these hybridization experiments represented abundant mRNAs with a frequency in the total mRNA population exceeding 0.1% (10). The cDNA clones that were negative in the hybridization assay represented a diverse set of low abundance mRNAs, and can be divided into several abundance classes (7). Here, for reasons of simplicity, this set of mRNAs is collectively referred to as the rare group. 61 of the cDNA clones from the abundant group, and 140 from the rare group, were chosen for further analysis.

To be able to rapidly analyze a large set of cDNA clones, a linked PCR amplification and linear PCR sequencing strategy was developed (see Materials and Methods). Briefly, individual λ plaques were picked and suspended in SM buffer. A small aliquot of the phage suspension was mixed with two primers, complementary to DNA sequences on the opposite sides of the cDNA insert, and amplified with AmpliTaq DNA polymerase (15) for 27 cycles. The length and the amount of each amplified cDNA insert were analyzed using agarose gel electrophoresis. In case no amplification product was visible, a second amplification was performed, this time for 33 cycles. The average size of the 171 cDNA clones that contained cDNA inserts, was 1150 bp; 30% (51) of these clones belonged to the abundant group, whereas 70% (120) of the clones came from the rare group.

Sequencing of PCR-amplified cDNA inserts

The cDNA clones were sequenced using an automated 373A DNA sequenator and a modified cycle-sequencing protocoll. When sequencing primers located 5′ of the poly $(A)^+$ tail in the amplified cDNA were used, however, the sequence following the long homonucleotide stretch in the cDNA insert was not readable. Sequencing primers located on the opposite side of the poly $(A)^+$ stretch in the cDNA insert, in contrast, routinely generated 250–350 bp of high quality sequence data. The most likely explanation for the difference in results is that during the amplification process, an unequal number of deoxyadenosines and deoxythymidines were introduced in the poly (A/T) tail of the amplification products. When Taq polymerase then traversed this region during the subsequent sequencing step, the length heterogeneity of the poly (A/T) tail in the population of amplified fragments, resulted in an uninterpretable DNA sequence. A similar effect caused by a combination of sequence repetition and DNA amplification has been observed for other repetitive sequences (16). As a control, the same cDNA clones were also purified by an alkaline-lysis DNA preparation method, omitting the PCR amplification step (10); with template DNA prepared in this manner, high quality DNA sequence data were generated from both ends of tested cDNA inserts.

All PCR-amplified DNA sequences were routinely sequenced from their 5′ end, i.e. from the side opposite to the poly (A/T) tail, thereby avoiding the problems with the 3′ end described above. A total of 47 024 bp were sequenced from the 5′ end of 171 cDNA clones, yielding an average of 275 bp of DNA sequence from each clone. These partial cDNA sequences were denoted expressed sequence tags, (ESTs), in accordance with Adams *et al.* (6).

Comparison of DNA sequences to databases

The ESTs from the 120 rare cDNAs were first examined for similarities in the EMBL nucleic acid database (release 27), using the program FASTA (17). 84% of the cDNA clones from this group revealed no significant homologies to mouse sequences included in the EMBL database (Table 1). A small subset of these novel mouse ESTs were, however, found to be similar to a set of genes previously characterized in humans, including the human gene for La autoantigen, lipoamide dehydrogenase, the ret transforming protein, the poly A binding protein, and the MHC

Table 3. EST matches to mouse genes in the EMBL database.

Clone name (TSG)	Mouse gene identity	Number	Redundancy
11	t complex polypeptide 1	MMTCPAA	1
34	Interferon γ receptor	MMIFNGA	1
45	Ribosomal protein S6	MMRPS6	1
76	Histone H3.3 gene	MMH33REP	1
87	SPARC/osteonectin	MMSPARCR	5
140	MuLV-related RNA	MMMLVA2A	1
172	Cytochrome oxidase II	MITOMM	4
173	URF 4	MITOMM	2
174	Elongation factor 1 α	MMEF1A	10
175	URF A6L	MITOMM	8
182	Cytochrome B	MITOMM	11
185	URF 2	MITOMM	1
195	Cytochrome oxidase I	MITOMM	3
197	Heat shock protein 86	MMHSP86A	1
201	URF 1	MITOMM	1
202	Cytochrome oxidase III	MITOMM	4

54 ESTs were found to be identical to a set of mouse genes included in the EMBL DNA database. The homology for these matches all exceeded 98%. The contents of the columns are as described in Table 2. The EMBL number MITOMM refers to the mouse mitochondrial genome. TSG 11, 34, 45, 76, 86 and 140 were all isolated from the rare group, whereas the rest of the clones came from the abundant group.

Table 4. Accuracy of linked PCR amplification and automated PCR sequencing.

No. of clones compared	Average length	Aligned bases	Mismatches	Insertion/ deletions	Accuracy
54	289 bp	15607 bp	73	10	99.5 %

All 54 ESTs listed in Table 3 were aligned with their homologous mouse genes taken from the EMBL database, using the program GAP from the Genetic Computer Group (14).

Table 5. A comparison of random and differential cDNA clone isolation strategies.

	Random Adams et al.*	Random (rare + abundant)	Differential (rare)
No database match	38 % (49 %)	56 %	75 %
Database match-humans	33 % (41 %)	-	-
Database match-mouse	-	37 %	16 %
Database match-other species	8 % (10 %)	7 %	9 %

Column 1 (random): 606 randomly isolated cDNA clones were analyzed by Adams *et al.* (6), and compared to the Gen Bank nucleic acid database and the Protein Information Resource database. The percentages of ESTs classified into different categories are shown. 22% of the analyzed cDNAs in this study either had no insert or contained only polyadenylate sequences. Exclusion of these noninformative clones from the study by Adams *et al.* (6), results in the percentages shown within parentheses. Column 2 (random): The 171 cDNA clones analyzed here, were classified into different categories after comparison to the EMBL database. Column 3 (differential): The same as column 2, but only including the 120 rare cDNA clones.

protein (Table 2). A second set of ESTs within this group, for which no mouse gene homologues were found in the EMBL DNA database, had similarities to a group of rat genes, including the genes for ribosomal protein L35a, L18a, and collagen (III) $\alpha-1$.

16% of the mouse ESTs from the rare group matched a set of mouse sequences already included the EMBL database. Half of these matches represented nuclear genes, including the t complex polypeptide 1 gene, the interferon γ receptor gene, histone H3.3, and the SPARC/osteonectin gene (Table 3). The second half of matches were similar to various repetitive sequences included in the database, e.g. Alu-like sequences and B2 repeats.

A very different pattern emerged when the abundant group of cDNA sequences were compared to the EMBL nucleic acids database. Here, only 12% of the ESTs were found to represent novel mouse DNA sequences (Table 1). Instead, the majority of the abundant clones (88%), were identical to genes already included in the EMBL database. Most of these frequently expressed ESTs (44/45), were either of mouse mitochondrial origin or represented one nuclear gene, the mouse elongation factor 1 α gene (Table 3). One significant heterologous match, to the human ribosomal protein L 23 gene, was also found (Table 2).

A total of 95 ESTs from the rare and the abundant group showed no significant homology to sequences included in the EMBL database (Table 1). To analyze the internal redundancy within this set of novel sequences, all 95 ESTs were assembled into a small database, using the program Dataset (14) and compared to each other (17). No sequence overlap was, however, found. To allow more sensitive comparisons to be made, the three forward frames for all 171 ESTs were translated. Open reading frames longer than 50 amino acids were found in 80 of the ESTs. These amino acid sequences were compared to the Swissprot protein sequence database (release 18), using the program FASTA (17). Excluding the matches summarized in Tables 2 and 3, however, no additional significant sequence similarities were found.

The accuracy of DNA sequencing with the linked PCR amplification-sequencing approach described here was also evaluated. All 54 cDNA sequences that matched either mitochondrial or nuclear mouse genes (Table 3) were compared to their homologous mouse genes taken from the EMBL database. A total of 15,607 bp was aligned using the program Gap from the Genetic Computer Group (14). The accuracy of PCR-mediated DNA sequencing was found to be 99.5% (Table 4). The majority of the recorded sequencing errors were mismatches, whereas very few insertions or deletions were recorded.

DISCUSSION

To facilitate the understanding of human biology and disease, the mapping and the characterization of the human genome is an important undertaking (18, 19). Of equal importance is the parallel analysis of the genomes from a number of model organisms, as functional predictions can be tested *in vivo* in these systems. The mouse, for a long time, has been the mammal of choice for comparative genetic analysis (20, 21). A cDNA sequencing strategy focusing on a molecular description of a majority of all human and mouse genes would be an important step towards a more complete understanding of various aspects of mammalian biology. Recently, Adams *et al.*, (6) partially sequenced more than 600 randomly isolated cDNAs from a set of human brain cDNA libraries. Of these clones, only 46% represented novel human sequences (Table 5). If the cDNA clones that contained no insert, or consisted entirely of polyadenylate sequences, were excluded, 59% of the remaining ESTs were found to represent novel human sequences (6).

In this study two different cDNA cloning and isolation strategies have been analyzed, with the aim of improving the probability of finding novel genes, and reducing the redundancy involved in a large scale cDNA sequencing project. The directional cDNA cloning strategy used here, together with oligo $(dT)_{12-16}$ priming of mRNA, combines two desirable goals in a cDNA sequencing project. The sequences localized next to the poly $(A)^+$ tail in the 3′ end of the cDNA will serve as an internal control of the redundancy within a large set of ESTs. The sequences from the 5′ end of the cloned cDNAs, on the other hand, are more likely to contain information from the coding region of the gene, and will allow more sensitive database comparisons to be made. To minimize the time spent on preparation of DNA templates for sequencing, a linked PCR amplification-sequencing strategy was tested. The amplification products were analyzed using agarose gel electrophoresis, in order to exclude those cDNAs that had very short inserts, lacked inserts completely or failed to produce a PCR product. In this way, 30 of the 201 candidate cDNAs could be eliminated, optimizing the sequencing effort.

A total of 171 randomly selected cDNA clones have been partially sequenced in this study. A comparison of these 171 ESTs to nucleic acid and protein databases, should give a result similar to Adams *et al.* (6). Indeed, it was found that 63% of the randomly isolated cDNAs described here, *vs* 59% of the randomly isolated cDNAs analyzed by Adams *et al.* (6), represented novel sequences (Table 5).

To improve the probablity of isolating novel cDNAs, a differential cDNA screening strategy was evaluated, as an alternative to selection of cDNA clones at random. The cDNA hybridization experiment performed here revealed that approximately 30% of the cDNA clones in a prepubertal testicular cDNA library represented abundant cDNA clones. Sequence analysis and comparison of 51 abundant ESTs to the EMBL database showed that only 12% of the mouse sequences had not been previously incorporated into the database (Table 1). In contrast, a similar comparison to the EMBL database of 120 cDNA clones belonging to the rare group, showed that 84% of these clones contained novel mouse sequences.

To isolate very rare, tissue-specific and developmentally regulated mRNAs, it will most likely be necessary to develop technically more demanding cDNA library preprocessing techniques, e.g., normalization (22) and subtraction (23). The differential hybridization strategy used here, however, removes abundantly expressed cDNA clones, and therefore markedly improves the probability of finding new, previously not characterized genes. This strategy is convenient to use, and is applicable to all currently available cDNA libraries.

ACKNOWLEDGEMENTS

I thank Drs Bertil Daneholt, Lars Wieslander and Thomas Perlman for helpful suggestions, Eva Grunder-Brundell, Maria Starborg, Elisabeth Long for technical support, and Ann E. Cutting as well as Dr. Kristen Mayo for providing valuable comments on the manuscript. I also thank Ulf Skoglund, and in particular Hans Mehlin, for invaluable computer support. This work was mainly supported by funds provided by Axel and

Margaret Ax:son Johnsons Stiftelse. Additional support was provided by the Swedish National Research Council, the Karolinska Institutet, Magnus Bergvalls Stiftelse, Socialstyrelsen fonder and T.A.A. Amundsons stiftelse.

REFERENCES

1. Watson J.D. (1990) *Science* **248**, 44–48.
2. Human Genome News (1991) **2** (no. 6), 1.
3. Bishop, J.O., Morton, J.G., Rosebach, M. and Richardson, M. (1974) *Nature* **250**, 199–204.
4. Schmidtke, J. and Cooper, D.N. (1990) *Nucleic Acids Res.* **18**, 2413–2414.
5. Brenner, S. (1990) *Ciba Found. Symp.* **149**, 6–8.
6. Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., Kerlavage, A.R., McCombie, W.R. and Venter, J.C. (1991) *Science* **252**, 1651–1656.
7. Davidson, E.H. (1976) Gene activity in early development, Academic Press, Inc., New York.
8. Hastie, N.D. and Bishop, J.O. (1976) *Cell* **9**, 761–774.
9. Davidson, E.H. and Britten, R.J. (1979) *Science* **204**, 1052–1059.
10. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A laboratory manual. Cold Spring Harbor University Press, Cold Spring Harbor.
11. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A laboratory manual, 2 ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
12. Feinberg, A.P. and Vogelstein, B. (1983) *Anal. Biochemistry* **132**, 6–13.
13. Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Ehrlich, H.A. and Arnheim, N. (1985) *Science* **230**, 1350–1354.
14. Devereux, J. Haeberli, P. and Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
15. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuichi, R., Horn, G.T., Mullis, K.B. and Ehrlich, H.A. (1988) *Science* **239**, 487–491.
16. Hagelberg, E., Gray, I.C. and Jeffreys, A.J. (1991) *Nature* **352**, 427–429.
17. Pearson, W.J. (1990) *Methods Enzymol.* **183**, 63–98.
18. Cantor, C.R. (1990) *Science* 248, 49–51.
19. Stephens, J.C., Cananaugh, M.L., Gradie, M.I., Mador, M.L. and Kidd, K.K. (1990) *Science* **250**, 237–244.
20. Lyon, M.F. and Searle, A.G. (1990) Genetic Variants and Strains of the Laboratory Mouse, 2nd edn. Oxford University Press, Oxford.
21. O'Brien, S.J. (1991) *Current Opinion in Genetic and Development* **1**, 105–111.
22. Patanjali, S.R., Parimoo, S. and Weissman, S.M. (1991) *Proc. Natl. Acad. Sci.* USA **88**, 1943–1947.
23. Fargnoli, J., Holbrook, N.J. and Fornace, Jr.,A.J. (1990) *Anal. Biochem.* **187**, 364–373.