J. Enzyme Inhibition, 1989, Vol. 3, pp. 49–56
Reprints available directly from the publisher
Photocopying permitted by license only

cDNA CLONING OF HUMAN CALPASTATIN: SEQUENCE HOMOLOGY AMONG HUMAN, PIG, AND RABBIT CALPASTATINS

KIYOZO ASADA,* YOSHIZUMI ISHINO, MASAMITSU SHIMADA, TOMOKO SHIMOJO, MASAHIRO ENDO, FUSAO KIMIZUKA, IKUNOSHIN KATO, MASATOSHI MAKI,[†] MASAKAZU HATANAKA,[†] and TAKASHI MURACHI[§]

Pharmaceutical Research Laboratories, Takara Shuzo Co., Ltd., Kyoto 600-91, Japan, and [†]Institute for Virus Research and §Department of Clinical Science and Laboratory Medicine, Faculty of Medicine, Kyoto University, Kyoto 606, Japan

(Received 11 October, 1988)

cDNA of human calpastatin, an inhibitor protein specific for calpain (EC 3.4.22.17; Ca^{2+}-dependent cysteine proteinase) was isolated by screening of a library prepared from human liver mRNA with pig calpastatin cDNA fragment as a probe. The primary structure of human calpastatin was deduced from the nucleotide sequence of the cDNA and compared with that of pig and rabbit calpastatins already reported. Human calpastatin consisted of 673 amino acid residues and had 78% and 77% indentity to pig or rabbit calpastatins. respectively. Human calpastatin had a domain structure with four internally repetitive sequences and one N-terminal non-homologous sequence like the other calpastatins. Human calpastatin had two deletions, 22 and 13 residues long in domain L and domain 1, respectively, compared to pig or rabbit calpastatins.

KEY WORDS: Proteinase inhibitor, library screening, nucleotide sequence, predicted amino acid sequence.

INTRODUCTION

Calpastatin is an endogenous protein inhibitor acting specifically on calpain (EC 3.4.22.17; Ca^{2+}-dependent cysteine proteinase).[1] Both calpain and calpastatin are widely distributed in mammalian and avian cells.[2,3] The physiological roles of calpain have not been clarified, but this proteinase–proteinase inhibitor system may be important in various cellular functions, coupled with calcium ion mobilization.[4–9]

The primary structure of pig calpastatin has been identified by nucleotide sequencing of cloned cDNAs and by Edman degradation of the inhibitor purified from pig heart. The inhibitor is composed of 713 amino acid residues, and contains four internally repetitive sequences (domains 1–4) and one non-homologous sequence on the amino-terminal side (domain L), each separated by about 140 amino acid residues.[10,11] This structural feature is similar to that of rabbit calpastatin.[12]

Each repetitive region, domains 1–4, inhibited both calpain I (low-Ca^{2+}-requiring form) and calpain II (high-Ca^{2+}-requiring form) and each structural domain was a functional unit of calpastatin.[13,14] The structure-function relationship for pig calpas-

*Correspondence.

tatin has been studied by functional analysis of various deletion mutants of domain 3.[15] The central region, composed of 52 amino acid residues (spanning 441 to 492 bp by our coordinates) seems to be an important region for the function of the inhibitor. We have proposed that there is a central consensus sequence of 12 amino residues in each domain, which is essential for the formation of the tertiary structure. A similar kind of analysis also has been reported for rabbit calpastatin[16] and a peptide of 34 residues long in the central region of domain 1 inhibits both kinds of calpains.[17]

To find if human calpastatin has structural features similar to those for pig and rabbit calpastatins,[11,12] we cloned human calpastatin cDNA and compared its primary structure deduced from the cDNA nucleotide sequence with those of different mammalian sources.

MATERIALS

The cDNA library and poly(A+) mRNA were purchased from Clontech Inc. (Palo Alto, CA, USA). Reverse transcriptase, DNA polymerase I, RNase H, T4 DNA ligase, methylase, and restriction enzymes were the products of Takara Shuzo Co., Ltd. (Kyoto, Japan); oligonucleotides were prepared at the same company. Nylon membrane was purchased from Amersham (UK). λgt 10 DNA and packaging components were obtained from Stratagene (San Diego, CA, USA).

METHODS

Construction of cDNA Library

Double-stranded cDNA was synthesized from 1–8 μg of human liver poly(A+) RNA as described by Gubler and Hoffman[18] except that 0.3–1.2 μg of random hexanucleotide was used as the primer. cDNA was then treated with *Eco*RI methylase, ligated with *Eco*RI linker, and digested with *Eco*RI. Linker-attached cDNA was gel-filtered with Sepharose CL-4B and ligated with *Eco*RI-digested phage vector λgt 10; the ligation mixture was packaged *in vitro* as described by Huynh *et al.*[19]

Screening of cDNA Library

About 2×10^4 phages were spread on a square plate (10×10 cm) and transferred to a nylon membrane. Hybridization was performed overnight at 65°C in a solution containing 6 × SSC (1 × SSC was 0.15 M NaCl/0.15 M sodium citrate, pH 7.0), 1% SDS, 100 μg/ml heat-denatured salmon sperm DNA, 5 × Denhardt's (1 × Denhardt's was 0.02% bovine serum albumin/0.02% polyvinylpyrrolidone/0.02% Ficoll), and ^{32}P-labelled probe DNA at 10^6 cpm/ml. Filters were washed four times for 15 min each time in 2 × SSC and 0.1% SDS at room temperature. Excess liquid was drained off the filters, which were autoradiographed at −70°C with an intensifying screen.

Nucleotide Sequencing

Nucleotides were sequenced as described by Messing.[20] Restriction fragments of cDNA were subcloned into M13 mp 18/19 and single-stranded DNA was prepared. Chain termination reactions were carried out with a kit prepared by Takara Shuzo.

Homology Search

A program DNASIS developed in Hitachi Software Engineering Co., Ltd. (Yokohama, Japan) was employed for dot matrix analysis and a best fit alignment of two similar sequences.

RESULTS

cDNA Cloning of Human Calpastatin

A commercial cDNA library constructed from human liver mRNA with a phage vector λ-gt 11 was screened for human calpastatin cDNA clones. The 935-bp *Pst*I fragment of pCSFL713, spanning the region of domain 1 to domain 2 in pig calpastatin (Takano *et al.*),[11] was used as a probe. In all, 51 candidate clones were obtained from 10^5 recombinant phages, and eight clones arbitrarily selected were found to have an insertion fragment of the same size, 1.8 kb. One of these clones, designated λcs 19, was used to sequence the nucleotides of the insertion fragment.

Another clone was obtained that covered the 5′-terminal region of the cDNA that λcs 19 lacked. In this screening, the desired clones were isolated from the cDNA library constructed with random hexanucleotide primers with the ^{32}P-labelled *Eco*RI–*Acc*I probe fragment or λcs 19 (shown in Figure 1). Four candidate clones were isolated from the recombinant phages and one of these clones, designated λcs 143, covered the 5′-terminal region of the cDNA shown in Figure 1.

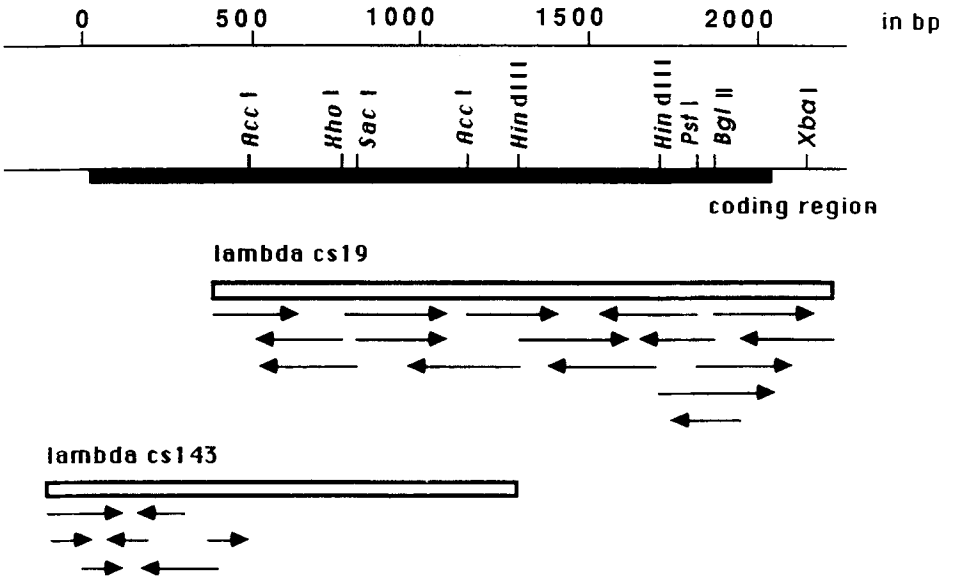
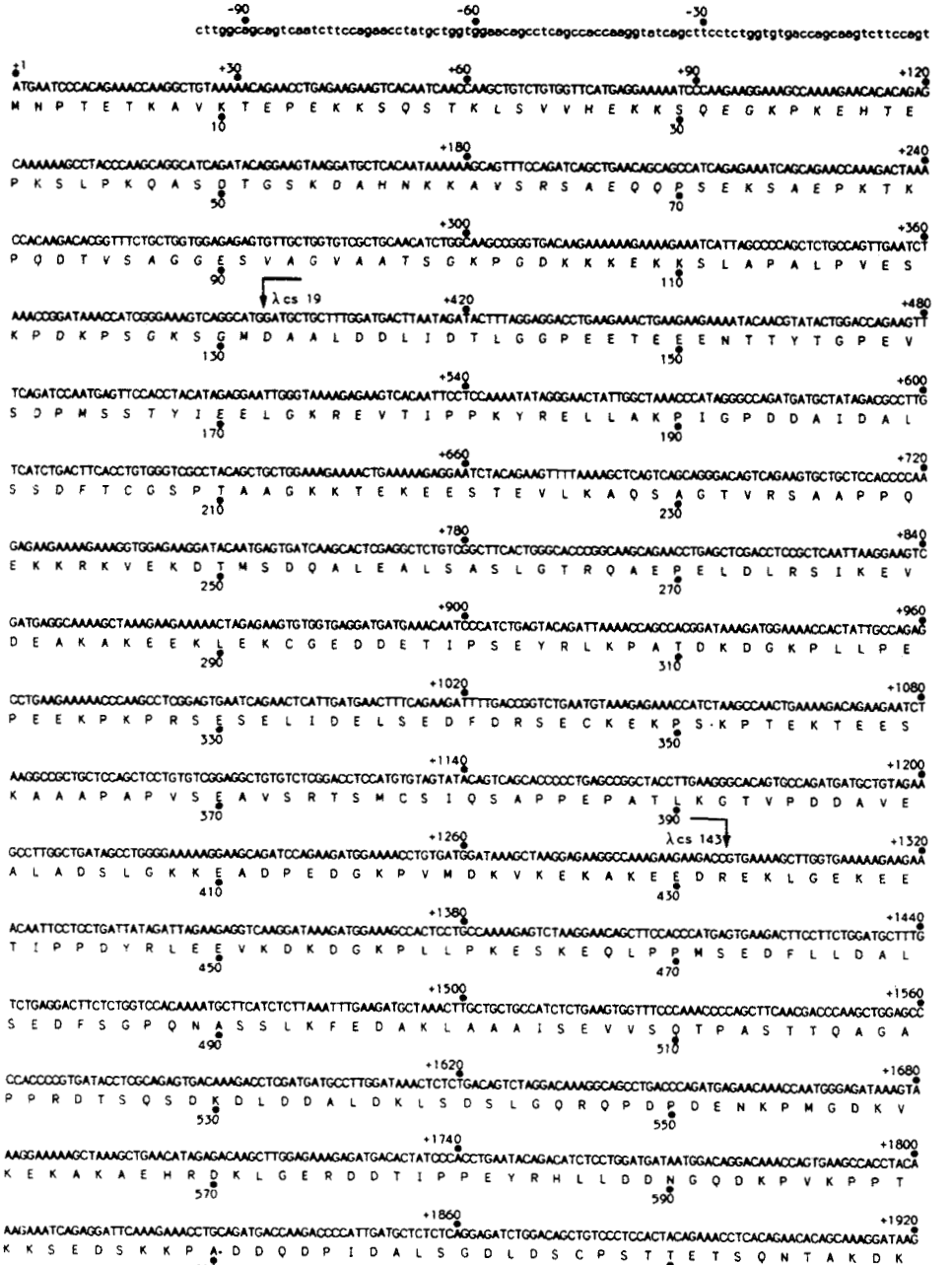


FIGURE 1 Restriction map of cDNA clones of human calpastatin and sequencing strategy. The human calpastatin coding region is indicated by the solid box. Coordinates are given as the nucleotide base number (bp) starting from the 5′-terminal end. Negative numbers are used for the 5′ untranslated regions. Open boxes under the restriction map show the regions two clones (λcs 19 and λcs 143) cover. Arrows below each clone show the direction of sequencing and the region sequenced.



Nucleotide Sequences of Cloned cDNAs and Primary Structure Deduced for Human Calpastatin

The human calpastatin cDNA was sequenced with the strategy shown in Figure 1. No difference was observed between λcs 19 and λcs 143 in the region from 394 to 464 bp in these sequence coordinates. Within the 2314-bp sequence obtained (Figure 2), 31 open reading frames were identified and the longest one, 2019 bp long, was predicted to code for human calpastatin because of its high homology to the nucleotide sequences of pig and rabbit calpastatin. This prediction was confirmed by the following experiment. The longest frame was subcloned in frame into an expression vector and crude extract of *E. coli* cells that produced the expression plasmid was tested for inhibition. The crude extract caused inhibition (Maki, M., *et al.* manuscript in preparation). If human calpastatin is assumed to start from the methionine, which is the same relative position here as in rabbit calpastatin,[12] then human calpastatin consists of 673 amino acids residues with the molecular weight of 72605.1. The clones obtained covered the 5′ and 3′ untranslated regions, which were 96 bp long and 199 bp long, respectively.

Comparison of the Primary Structure of Human Calpastatin with Those of Pig and Rabbit Calpastatins

Pig and rabbit calpastatins have a domain structure with four internally repetitive sequences (domain 1, 2, 3, and 4) and one non-homologous sequence on the N-terminal region.[11,12] To find whether human calpastatin also has this structural feature, internal homology was searched. Human calpastatin also had four internally repetitive sequences, each about 140 amino acid residues long (Figure 3a). Human calpastatin was then compared with pig (Figure 3b) and rabbit calpastatins (Figure 3c). The amino acid sequence of human calpastatin was highly homologous with that of pig and rabbit calpastatin, as expected from the nucleotide sequence analysis, but human calpastatin had two deletions in its internal regions. The amino acid sequence of human calpastatin was aligned with the sequences of pig and rabbit calpastatins (Figure 4), allowing deletion, gaps, and insertions to give maximum homology. Human calpastatin had 78% identity (526 identical residues in the 673 residues) and 77% identity (519 identical residues in the 673 residues) to pig and rabbit calpastatins, respectively. One deletion that was 22 amino acid residues long existed in domain L and another that was 13 residues long, in domain 1, according to the definition of the domains′ boundary in pig calpastatin.[11] This result is summarized in Table I.

DISCUSSION

In the cDNA cloning of human calpastatin, we isolated two clones in addition to λcs 19 and 143. One of these, designated λcs 131, had a high heterogeneity compared to λcs 143. λcs 131 had 10 substitutions of single bases in the region from + 87 to 464

FIGURE 2 Nucleotide sequence of cloned cDNA and amino acid sequence deduced for human calpastatin. Nucleotides are numbered as in Figure 1. The amino acid sequence deduced shown beneath is numbered from the N-terminal end of human calpastatin. The coordinates are given above the sequence in base pairs and below in amino acid residues. Dots show exact locations. The 5′ boundary of λcs 19 and the 3′ boundary of λcs 143 are indicated by arrows.

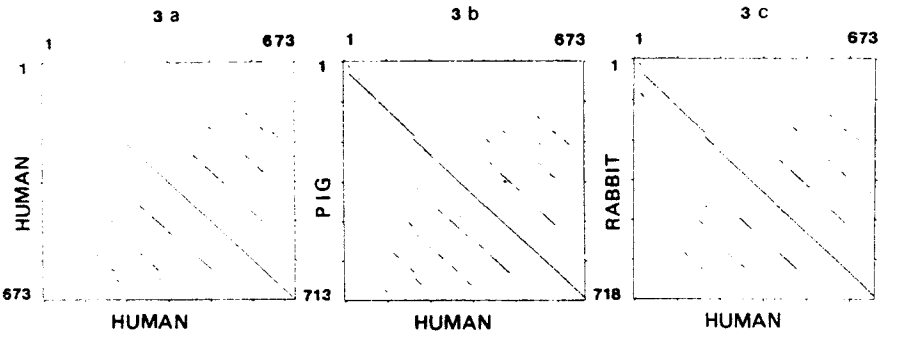


FIGURE 3 Comparison matrix between human and other calpastatins. Parameters were set so that dots were pointed when 8 residues or more out of 18 matched. a, human to human; b, pig to human; and c, rabbit to human.

by our coordinates that we examined. This suggests that two or more copies of the calpastatin gene may exist in the human genome, if they didn't arise from an error during cDNA cloning or polymorphism of the calpastatin gene in the allele or the individual.

Here, the primary structure of human calpastatin was deduced from the nucleotide sequence of its cDNA and compared with calpastatin from two other mammals. Human calpastatin had around 77% identity to both calpastatins, and retained the consensus sequences proposed by Maki[15] and Emori.[16] These findings suggest that the proposed consensus sequence is essential for the inhibition of calpains.

Human calpastatin had two deletions in domains L and 1, compared to pig and rabbit calpastatins. These deletions were unlikely to be artifacts of the cloning steps, because the deletion in domain 1 was also observed in human calpastatin cDNA derived from the heart as well as the liver (unpublished data). The 34-mer peptide of rabbit calpastatin spanning the residues asp-184 to Gly-217 is the shortest calpastatin fragment so far reported that inhibits both kinds of calpains.[17] The 34-mer peptide of

TABLE I
Summary of the comparison of sequence alignments between human and other calpastatins

	Comparison to pig calpastatin		Comparison to rabbit calpastatin	
	Position	Number of residues	Position	Number of residues
Deletions	9–10	22	9–10	22
	189–190	13	189–190	13
Gaps	93–94	1	93–94	1
	222–223	1	219–220	3
	237–238	1	390–391	1
			644–645	1
Insertions	390–391	2	16	1
	494	1		
Number of identical residues		526		519
Homology		78.2%		77.1%

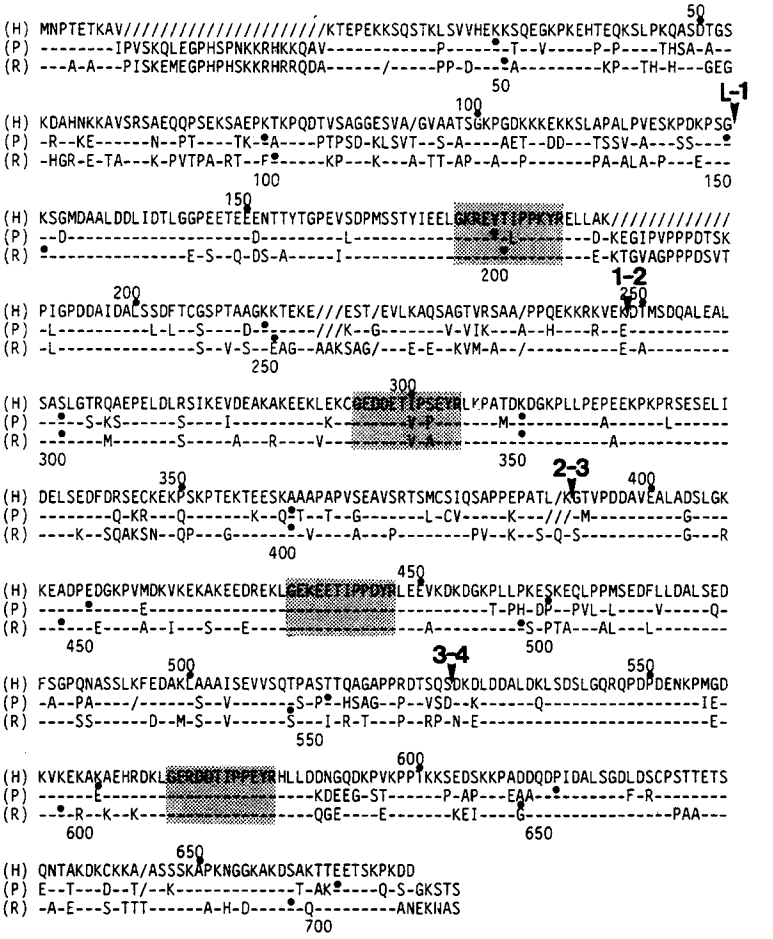


FIGURE 4 Comparison of amino acid sequences of the three calpastatins. Amino acid sequences of human (H), pig (P), and rabbit (R) calpastatins, in this order, are aligned from top to bottom. The coordinates of the amino acid sequence are shown every 50 residues (dotted). The numbers above the sequence refer to human calpastatin and those below refer to the pig and rabbit calpastatins. Gaps (indicated by slashes) were introduced to achieve maximum homology. Residues identical with human calpastatin are marked with dashes. The boundaries of each domain defined for pig calpastatin[11] are indicated by vertical arrows above the sequence. The reported "central consensus sequence" in each domain[15] is shaded.

rabbit calpastatin and domain 1 of human calpastatin have 26 residues in common (residues 184 to 209 in rabbit calpastatin and residues 162 to 187 in human calpastatin). By analogy to isolated domain 1 proteins of pig[15] and rabbit[16] calpastatins, it is reasonable to assume that domain 1, alone, of human calpastatin can also exert inhibition of calpains. This would, in turn, imply that the common peptide 26 residues long can still retain inhibitory potency against calpains. The synthesis and test for inhibitory activity of this and related peptides are now in progress.

Acknowledgements

We thank Miss Uchimura for the preparation of the oligonucleotides and Miss Ohara for technical assistance.

References

1. Parkes, C. (1986) in *Proteinase Inhibitors* (Barrett, A. J., and Salvesen, G., eds.) pp. 571–587, Elsevier, Amsterdam.
2. Murachi, T., Tanaka, K., Hatanaka, M., and Murakami, T. (1981) *Adv. Enzyme Regul.*, **19,** 407–424.
3. Murachi, T. (1983) *Trends Biochem. Sci.*, **8,** 167–169.
4. Murachi, T. (1983) in *Calcium and Cell Functions* (Cheung, Y.W., ed.) vol. 4, pp. 377–410, Academic Press, New York.
5. Murachi, T. (1984) *Biochem. Soc. Symp.*, **49,** 149–167.
6. Pontremoli, S., and Melloni, E. (1986) *Annu. Rev. Biochem.*, **55,** 455–481.
7. Murray, A.W., Fournier, A., and Hardy, S.J. (1987) *Trends Biochem. Sci.*, **12,** 53–54.
8. Suzuki, K. (1987) *Trends Biochem. Sci.*, **12,** 103–105.
9. Mellgren, R.L. (1987) *FASEB J.*, **1,** 110–115.
10. Takano, E., Maki, M., Hatanaka, M., Mori, H., Zenita, K., Sakihama, T., Kannagi, R., Marti, T., Titani, K., and Murachi, T. (1986) *FEBS Lett.*, **208,** 199–202.
11. Takano, E., Maki, M., Mori, H., Hatanaka, M., Marti, T., Titani, K., Kannagi, R., Ooi, T., and Murachi, T. (1988) *Biochemistry*, **27,** 1964–1972.
12. Emori, Y., Kawasaki, H., Imajou, S., Imahori, K., and Suzuki, K. (1987) *Proc. Natl. Acad. Sci. USA*, **84,** 3590–3594.
13. Maki, M., Takano, E., Mori, H., Kannagi, R., Murachi, T., and Hatanaka, M. (1987) *Biochem. Biophys. Res. Commun.*, **143,** 300–308.
14. Maki, M., Takano, E., Mori, H., Sato, A., Murachi, T., and Hatanaka, M. (1987) *FEBS Lett.*, **223,** 174–180.
15. Maki, M., Takano, E., Osawa, T., Ooi, T., Murachi, T., and Hatanaka, M. (1988) *J. Biol. Chem.*, **263,** 10254–10261.
16. Emori, Y., Kawasaki, H., Imajoh, S., Minami, Y., and Suzuki, K. (1988) *J. Biol. Chem.*, **263,** 2364–2370.
17. Emori, Y. (1988) *Seikagaku (Tokyo)*, (mini-review in Japanese) **60**, 266–272.
18. Gubler, U., and Hoffman, B.J. (1983) *Gene*, **25,** 263–269.
19. Hyunh, T.V., Young, R.A., and Davis, R.W. (1985) in *DNA Cloning Techniques, A Practical Approach* (Glover, D., ed.) vol. 1, pp. 49–78 Oxford, IRL Press.
20. Messing, J. (1983) *Methods Enzymol.*, **101,** 20–78.