Isolation of a cDNA clone encoding pancreatic polypeptide

(recombinant DNA/icosapeptide/hybridization/endocrine neoplasm)

Toshiyuki Takeuchi and Tadataka Yamada

Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI 48109

ABSTRACT   We have isolated a cDNA clone encoding pancreatic polypeptide from a cDNA library constructed with RNA from an endocrine neoplasm of the human pancreas. The cDNA was inserted into plasmid pBR322 and the plasmid was cloned in *Escherichia coli*. Oligonucleotides d(G-C-$\frac{A}{G}$-T-A-$\frac{T}{C}$-T-G-N-G-C-C-A-T) specific for the amino acid sequence Met-17 18 19 20 21 Ala-Gln-Tyr-Ala of pancreatic polypeptide were used as hybridization probes. The pancreatic polypeptide cDNA isolated was 465 base pairs long and encoded a peptide of 95 amino acids in the coding region. The 36-amino acid sequence of pancreatic polypeptide was flanked by a 29-amino acid putative leader sequence at the amino terminus and a connecting tripeptide (Gly-Lys-Arg) followed by a 27-amino acid peptide at the carboxyl terminus. The first 20 of the amino acids in the carboxyl-terminal heptacosapeptide were identical to the structure of human pancreatic icosapeptide with the single exception of an isoleucine substitution for valine in the 18th position. This alteration results from an A→G substitution in the nucleotide sequence of the cDNA and may represent a genetic variation or a point mutation in the pancreatic polypeptide cDNA.

Human pancreatic polypeptide is a 36-amino acid hormone that may be important in the regulation of pancreatic secretion and gall bladder relaxation (1). As with other biologically active polypeptides, pancreatic polypeptide is thought to result from post-translational processing of a larger precursor form. In biosynthetic studies with dog pancreas, Schwartz and Tager (2) have found evidence that pancreatic polypeptide is cosynthesized with an icosapeptide that may represent a carboxyl-terminal extension of pancreatic polypeptide in the structure of its precursor. A homologous, but not identical, icosapeptide has been purified from human pancreas (3). Recently, the structure of the pancreatic polypeptide precursor has been deduced from the nucleotide sequence of a cDNA clone obtained from human pancreatic tissue (4). The predicted structure confirms the presence of the icosapeptide in the carboxyl-terminal extension of pancreatic polypeptide. In the present paper, we report the independent construction from an endocrine tumor of a recombinant cDNA clone encoding pancreatic polypeptide. Our cDNA clone contained a variation in a single nucleotide, possibly representing a point mutation, which would result in an alteration in the amino acid sequence of the icosapeptide.

MATERIALS AND METHODS

Enzymes and Reagents. Enzymes and reagents were purchased from the following suppliers: restriction enzymes from New England Biolabs, Bethesda Research Laboratories, or International Biotechnologies (New Haven, CT);

DNA polymerase I (Klenow fragment), terminal deoxynucleotidyltransferase, and T4 polynucleotide kinase from P-L Biochemicals; avian myeloblastosis virus reverse transcriptase from Life Sciences (St. Petersburg, FL); oligo-$(dT)_{12-18}$, oligo$(dG)_{12-18}$, and oligo(dT)-cellulose (type 77 F) from P-L Biochemicals; dG-tailed pBR322 from New England Nuclear; $[\alpha\text{-}^{32}P]dATP$, $[\alpha\text{-}^{32}P]dCTP$, $[\gamma\text{-}^{32}P]ATP$, $[^3H]dCTP$, and $[^3H]dGTP$ from Amersham and New England Nuclear; and cellulose powder (CF-11) from Whatman. The pancreatic polypeptide-specific oligonucleotide probes were custom synthesized by Applied Biosystems (Foster City, CA).

Poly(A)$^+$ RNA Isolation and cDNA Library Synthesis. Total RNA was extracted from a 0.4-g fragment of a human endocrine tumor containing pancreatic polypeptide [57 nmol/g (wet weight)] by the guanidine thiocyanate method of Chirgwin *et al*. (5). Poly(A)$^+$ RNA was isolated by using oligo(dT)-cellulose. With 5 μg of poly(A)$^+$ RNA as a template, first-strand cDNA was synthesized in a 100-μl reaction mixture containing oligo$(dT)_{12-18}$ as a primer, $[\alpha\text{-}^{32}P]dATP$ and unlabeled dNTP. The reaction was catalyzed with reverse transcriptase and incubated at 42°C for 60 min. The resultant first-strand cDNA was tailed with $[^3H]dCTP$ (average length, 20 nucleotides) using terminal deoxynucleotidyltransferase. The dC-tailed cDNA was then reprimed with oligo$(dG)_{12-18}$, for second-strand cDNA synthesis, by incubation with $[\alpha\text{-}^{32}P]dCTP$, unlabeled dNTP, and DNA polymerase I (Klenow fragment) at 37°C for 30 min. The double-stranded cDNA was tailed with $[^3H]dCTP$ with terminal deoxynucleotidyltransferase then hybridized with dG-tailed pBR322 that had been cut with *Pst* I. Newly constructed plasmid was used to transform *Escherichia coli* MM294 strain and transformants were selected on the basis of tetracycline resistance. One thousand five hundred transformants were obtained from 16 ng of double-stranded cDNA; thus, transformation efficiency was ≈2.8 × 10⁴ colonies per pmol of DNA.

cDNA Library Screening. The cDNA library was screened with synthetic 14-base oligonucleotides d(G-C-$\frac{A}{G}$-T-A-$\frac{T}{C}$-T-G-N-G-C-C-A-T)   corresponding   to   a   pancreatic 17 18 polypeptide-specific   amino   acid   sequence   (Met-Ala-19 20 21 Gln-Tyr-Ala). The oligonucleotides were labeled with ^{32}P at Athe 5′ end with T4 polynucleotide kinase in the presence of $[\gamma\text{-}^{32}P]ATP$, and ^{32}P-labeled oligonucleotides were separated from unreacted $[\gamma\text{-}^{32}P]ATP$ by CF-11 column chromatography.   The   specific   radioactivity   of   the ^{32}P-labeled oligonucleotides was ≈3 × 10⁶ cpm/pmol. With this probe, 1500 transformants were screened by the colony hybridization method of Grunstein and Hogness (6). Nitrocellulose filters were prehybridized for 6 hr at 37°C in a solution containing 3× NaCl/citrate (1× NaCl/citrate is 0.15 M NaCl/0.015 M trisodium citrate), 0.05 M Hepes (pH 7.0), yeast RNA at 200 mg/ml, denatured salmon sperm DNA at 50 mg/ml, and 0.02% bovine serum albumin/Ficoll/polyvinylpyrrolidone. Then the nitrocellulose papers were hybridized at 37°C for 24 hr with the ^{32}P-labeled oligo-

5'   pBR-(G$_{23}$)-GGATCCGGCCTGAGGCAGGTGCTCGCTTGGTCTAGTGCCCATTTACTCTGGACTCCGG
                    *Bam*HI   *Hinf*I
-29   -20   -10
MET ALA ALA ALA ARG LEU CYS LEU SER LEU LEU LEU LEU SER THR CYS VAL ALA LEU LEU
ATG GCT GCC GCA CGC CTC TGC CTC TCC CTG CTG CTC CTG TCC ACC TGC GTG GCT CTG TTA
START --LEADER SEQUENCE--
                                                -1 1   10
LEU GLN PRO LEU LEU GLY ALA GLN GLY ALA PRO LEU GLU PRO VAL TYR PRO GLY ASP ASN
CTA CAG CCA CTG CTG GGT GCC CAG GGA GCC CCA CTG GAG CCA GTG TAC CCA GGG GAC AAT
----

                                                20   30
ALA THR PRO GLU GLN MET ALA GLN TYR ALA ALA ASP LEU ARG ARG TYR ILE ASN MET LEU
GCC ACA CCA GAG CAG ATG GCC CAG TAT GCA GCT GAT CTC CGT AGA TAC ATC AAC ATG CTG
                        OLIGONUCLEOTIDE

                36   40   50
THR ARG PRO ARG TYR |GLY|LYS|ARG| HIS LYS GLU ASP THR LEU ALA PHE SER GLU TRP GLY
ACC AGG CCT AGG TAT GGG AAA AGA CAC AAA GAG GAC ACG CTG GCC TTC TCG GAG TGG GGG
                                            --------------------------ICOSAPEPTIDE--------------------

                (VAL)   60   66
SER PRO HIS ALA ALA |ILE| PRO ARG GLU LEU SER PRO LEU ASP LEU STOP
TCC CCG CAT GCT GCT ATC CCC AGG GAG CTC AGC CCG CTG GAC TTA TAA TGCCACCTTCTGTCTC
Ava II ----------------------------------

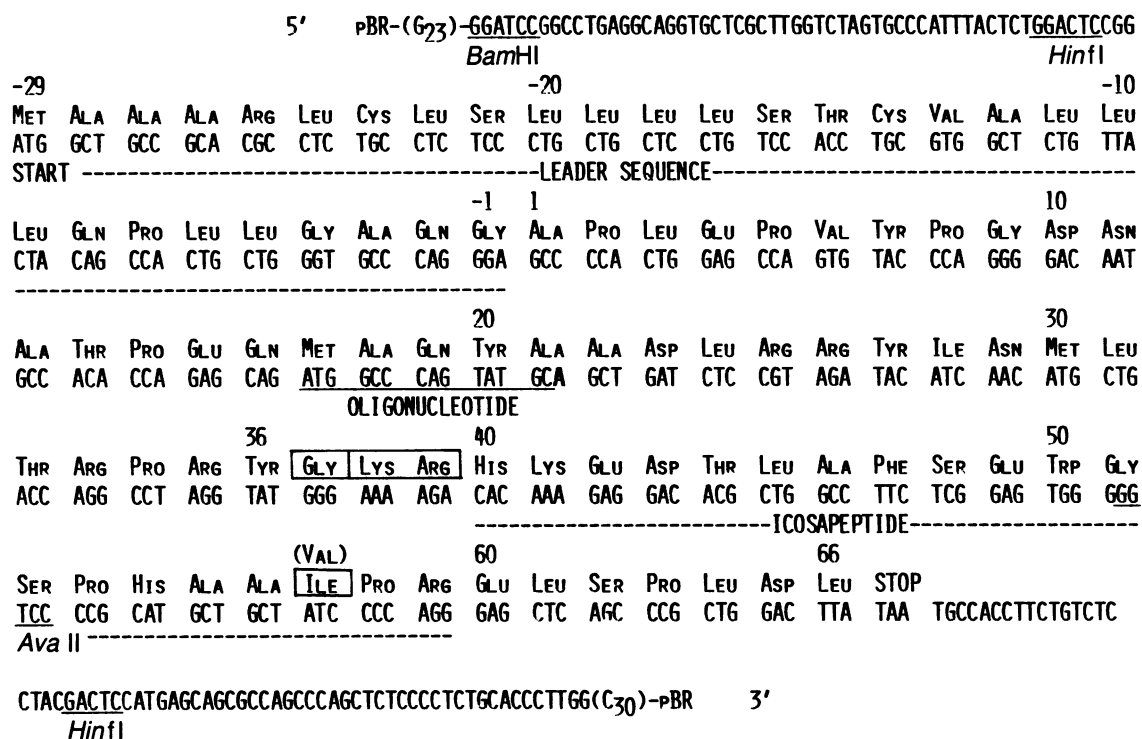CTACGACTCCATGAGCAGCGCCAGCCCAGCTCTCCCCTCTGCACCCTTGG(C$_{30}$)-pBR   3'
     *Hinf*I

FIG. 1. Nucleotide sequence of the pancreatic polypeptide cDNA clone. Restriction endonuclease sites for *Bam*HI, *Hinf*I, and *Ava* II, which were used for DNA sequencing, and the oligonucleotide sequence corresponding to the screening probe are indicated by the solid lines. The deduced amino acid sequence is numbered from the amino-terminal alanine of the pancreatic polypeptide portion of the sequence. The proposed leader sequence and the icosapeptide portion are indicated by dashed lines. The connecting sequence (Gly-Lys-Arg) and the variant amino acid are boxed.

nucleotide probes at 4×10^6 cpm/ml in the above solution. The filters were washed at 37°C in five changes of $3\times$ NaCl/citrate/0.5% sodium dodecyl sulfate. The filters were autoradiographed with Kodak XAR-5 x-ray film with an intensifying screen.

Plasmid DNA Isolation and Restriction Site Analysis. Recombinant plasmid was isolated from chloramphenicol-amplified cultures by using the mild alkaline condition method (7). The supercoiled plasmids were isolated by CsCl density gradient centrifugation or 0.7% agarose gel electrophoresis. The plasmid DNAs were digested with various restriction enzymes according to the instructions by the suppliers, and the resultant products were separated by agarose gel (0.7–1.5%) or polyacrylamide gel (5–8%) electrophoresis.

DNA Sequence Analysis. DNA fragments were prepared from the recombinant plasmid by digestion with *Bam*HI, *Ava* II, or *Hinf*I and radiolabeled with [α-^{32}P]dNTP by using DNA polymerase I (Klenow fragment) or with [γ-^{32}P]ATP by using T4 polynucleotide kinase. ^{32}P-labeled DNA fragments were separated by 0.7% agarose gel electrophoresis or 5% polyacrylamide gel electrophoresis and electroeluted from the gels. DNA fragments labeled at the both ends were digested with appropriate restriction enzymes and single fragments labeled at a single end were obtained by electrophoresis and electroelution. The nucleotide sequence was determined using the Maxam and Gilbert chemical reaction method (8).

Oligonucleotide Primed cDNA Synthesis. cDNA was synthesized from poly(A)$^+$ RNA (3–6 μg) primed with ^{32}P-labeled oligonucleotide (20–50 pmol) in a 20-μl reaction mixture containing 50 mM Tris·HCl (pH 8.3), 8 mM MgCl$_2$, 50 mM KCl, and 1 mM dNTP. Alternatively, the same reaction was primed with nonradioactive oligonucleotide in the presence of [α-^{32}P]dCTP and nonradioactive dTTP, dATP, and dGTP. The reaction mixture was incubated at

42°C for 90 min with reverse transcriptase. The [^{32}P]cDNA·mRNA complex was extracted with phenol/chloroform/isoamyl alcohol, 25:24:1 (vol/vol) and precipitated with 2 vol of ethanol. After digestion of the mRNA in 0.1 M NaCl/1 mM EDTA, the [^{32}P]cDNA products were separated on an 8% polyacrylamide/7 M urea slab gel and visualized by autoradiography using Kodak XAR-5 x-ray film. Several of the [^{32}P]cDNA bands were electroeluted from the gel for DNA sequencing or for use as hybridization probes.

RESULTS

On colony hybridization with the pancreatic polypeptide-specific oligonucleotide probes, two transformants gave a positive signal, although only one of the clones could be harvested. By polyacrylamide gel electrophoresis, the cDNA insert in the plasmid appeared to be 480 base pairs in size, and it could be digested with restriction enzymes *Bam*HI, *Pvu* II, *Hinf*I, *Ava* II, and *Hpa* II. The nucleotide sequence of the cloned cDNA and the corresponding deduced amino acid sequence of the pancreatic polypeptide precursor are depicted in Fig. 1. According to this structure, the 36 amino acids comprising pancreatic polypeptide are preceded by a 29-amino acid putative leader sequence. There is a specific amino acid complex consisting of a glycine residue followed by two basic amino acids that follows the tyrosine that serves as the carboxyl-terminal amino acid of pancreatic polypeptide (position 36). The cDNA encoded a second peptide consisting of 27 amino acids, of which the first 20 are identical to the icosapeptide purified by Schwartz *et al.* (3) with the single exception that an A→G substitution in the nucleotides encoding the 57th amino acid (position 18 in the icosapeptide) results in the replacement of a valine by an isoleucine residue. This alteration was confirmed as a T→C substitution in the nucleotide sequence of the second cDNA strand (Fig. 2). The translated region of the cDNA consists of 285 nucle-
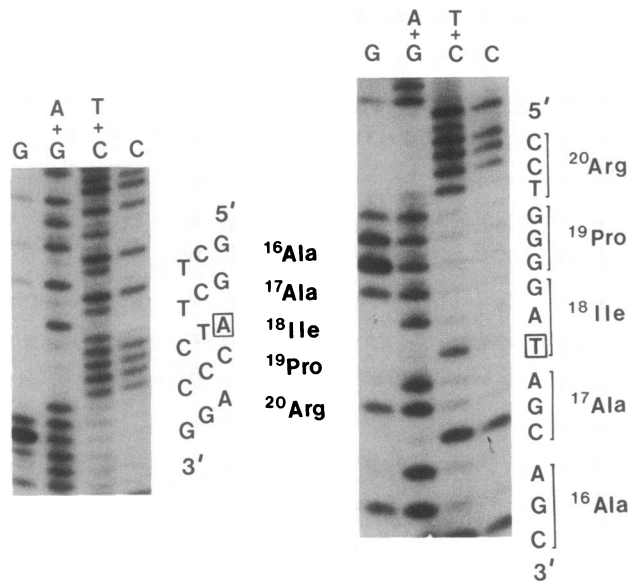
FIG. 2. Demonstration of the variant site in the nucleotide sequence of pancreatic polypeptide cDNA by polyacrylamide gel electrophoresis. The nucleotide sequence of both complementary strands of cDNA are depicted and the variant nucleotide is boxed. The amino acids encoded by the cDNA are depicted to the right of each autoradiogram and are numbered from the first amino acid of the icosapeptide.

otides encoding 95 amino acids with a molecular weight of 10,400. The cDNA insert also contains 56-nucleotide 5′- and 69-nucleotide 3′-untranslated segments. The additional 5′ dG(23) and 3′ dC(30) tails combine to account for a total insert size of 465 nucleotides.

The cDNA synthesized by using the pancreatic polypeptide-specific oligonucleotide as a primer could be fractionated into eight major and numerous minor bands by polyacrylamide gel electrophoresis (Fig. 3). Two of the major bands, nos. 5 and 2, were of 105 and 210 nucleotides, respectively. When the eight major labeled cDNAs were electroeluted from the polyacrylamide gels and used as detection probes, only cDNA 2 hybridized with the pancreatic polypeptide cDNA clone. Nevertheless, the nucleotide sequences of both cDNA 2 and cDNA 5 were determined. Although the nucleotide sequence of cDNA 2 exhibited a perfect match with the pancreatic polypeptide cDNA clone, that of cDNA 5 failed to exhibit any homology except in the specific region of the oligonucleotide primer.

DISCUSSION

In addition to its localization in healthy endocrine and exocrine pancreatic tissue, the frequent presence of pancreatic polypeptide in endocrine neoplasms has made it an occasionally useful tumor marker. The availability of one such tumor containing large amounts of pancreatic polypeptide has permitted us to isolate from a newly constructed cDNA library a recombinant clone encoding pancreatic polypeptide. As with the pancreatic polypeptide cDNA structure previously reported (4), the deduced amino acid sequence for the pancreatic polypeptide-encoding region is identical to the sequence reported for peptide purified by traditional biochemical techniques. Pancreatic polypeptide itself appears to be at the amino terminus of a 66-amino acid precursor that is preceded by a putative 29-amino acid hydrophobic leader sequence. Much in the manner observed for cDNA sequences of other carboxyl-terminally amidated peptides (9–12), a specific sequence of amino acids, glycine followed by two basic amino acid residues, appears to mark
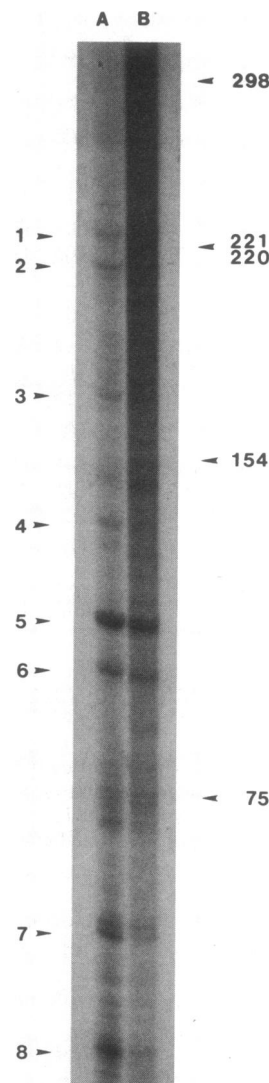

FIG. 3. Electrophoresis of cDNAs synthesized with pancreatic polypeptide-specific oligonucleotides as primers. Lane A: the oligonucleotides were labeled with [γ-^{32}P]-ATP by using T4 polynucleotide kinase, and oligonucleotide primer extension was carried out with unlabeled dNTP and reverse transcriptase. Lane B: the cDNAs were synthesized using nonradioactive oligonucleotide primers and extended with the use of [α-^{32}P]dCTP and reverse transcriptase. Numbers on the left refer to the specific cDNA bands used for hybridization with the pancreatic polypeptide cDNA clone. Numbers on the right refer to the nucleotide position as determined from HinfI-cut pBR322 fragments.

the cleavage site in the precursor for the formation of the amide moiety. Bradbury *et al.* (13) has postulated that the extended glycine moiety is converted to a carboxyl-terminal α-amide group by a process of dehydrogenation followed by hydrolysis of the resulting imino linkage.

The remaining peptide encoded by the pancreatic polypeptide cDNA clone is 27 amino acids long, the first 20 of which correspond to the pancreatic icosapeptide described by Schwartz *et al.* (3). Since the carboxyl-terminal amino acid of the icosapeptide is arginine, it is possible that the icosapeptide is cleaved from the heptacosapeptide by a tryptic reaction as part of a post-translational processing mechanism or as an artifact of the extraction and isolation methods used in purification. The unique aspect of the cDNA clone that we isolated as compared with the pancreatic polypeptide cDNA clone isolated by Schwartz and co-workers (4) is that a single A→G substitution in the translated region results in the substitution of an isoleucine moiety for valine in the deduced amino acid sequence of the icosapeptide. Since both isoleucine and valine are branched-chain hydrophobic amino acids, the amino acid substitution is a relatively conservative one. We cannot confirm from the present studies whether this observed alteration in nucleotide sequence represents a genetic variation or a true mutation, nor can we determine whether the alteration is present in the genetic structure of the patient's non-neoplastic tissue.

Potential difficulties in using peptide-specific oligonucleotides as cDNA primers or as hybridization probes are

Medical Sciences: Takeuchi and Yamada

Proc. Natl. Acad. Sci. USA 82 (1985) 1539

the possibilities that hybridization may occur despite base-pair mismatches or that nonspecific hybridization may occur with incidental regions of homology. Noyes *et al.* (14), using a dodecamer specific for hog gastrin, primed the synthesis of two cDNAs, only one of which encoded gastrin. Similarly, Chan *et al.* (15) primed the synthesis of two cDNAs using an insulin-specific decamer, but only one encoded insulin. Using conditions of relative stringency, we obtained at least eight cDNAs with our pancreatic polypeptide-specific oligonucleotides, but only one of them hybridized with the pancreatic polypeptide cDNA clone. We were able to identify that particular cDNA as being completely homologous with the pancreatic polypeptide cDNA, while at least one of the other cDNAs was homologous only in the region of the oligonucleotide primer.

In conclusion, we have isolated a cDNA clone encoding pancreatic polypeptide from a cDNA library constructed with RNA from an endocrine neoplasm. The nucleotide sequence of the clone exhibits an A→G substitution that would result in an alteration in the deduced amino acid sequence of the icosapeptide region of the pancreatic polypeptide precursor. The physiological importance of this potential alteration awaits studies to determine the biological activity of the icosapeptide.

1. Schwartz, T. W. (1983) *Gastroenterology* **85**, 1411–1425.
2. Schwartz, T. W. & Tager, H. S. (1981) *Nature (London)* **294**, 589–591.
3. Schwartz, T. W., Hansen, H. F., Hakanson, R., Sundler, F. & Tager, H. S. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 708–712.
4. Boel, E., Schwartz, T. W., Norris, K. E. & Fiil, N. P. (1984) *EMBO J.* **3**, 909–912.
5. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299.
6. Grunstein, M. & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 3961–3965.
7. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513–1523.
8. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
9. Amara, S. G., David, D. M., Rosenfeld, M. G., Roos, B. A. & Evans, R. M. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 4444–4448.
10. Land, H., Schutz, G., Schmale, H. & Richter, D. (1982) *Nature (London)* **295**, 299–303.
11. Yoo, O. J., Powell, C. T. & Agarwal, K. L. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 1049–1053.
12. Deschenes, R. J., Lorenz, L. J., Hauer, R. S., Roos, B. A., Colbier, K. J. & Dixon, J. E. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 726–730.
13. Bradbury, A. F., Finnie, M. D. A. & Smyth, D. G. (1982) *Nature (London)* **298**, 686–688.
14. Noyes, B., Mevarech, M., Stein, R. & Agarwal, K. L. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 1770–1774.
15. Chan, S. J. S., Noyes, B. E., Agarwal, K. L. & Steiner, D. F. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 5036–5040.