16. O. Amster-Choder and A. Wright, unpublished data.
17. M. R. Diaz-Torres and A. Wright, unpublished data.
18. J. K. Bryan, *Anal. Biochem.* **78**, 523 (1977).
19. H. Towbin, T. Staehlin, J. Gordon, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4350 (1979).
20. R. J. Zagursky and M. L. Berman, *Gene* **27**, 183 (1984).
21. AG1688 [*ara*D139 Δ(*ara-leu*)7697 Δ*lac*X74 *gal*E *gal*K *rps*L *hsd*R (F'128 *laci*^{q} *lacZ*::*Tn5*)] is the same as strain AG115 described by D. J. Rudner, J. R. LeDeaux, K. Ireton, A. D. Grossman, *J. Bacteriol.* **173**, 1388 (1991).
22. pOAC100 was constructed by the deletion of the *zip::lacZ'* from pJH391 and the introduction of a translational stop codon. We constructed pOAC101 by replacing the *zip::lacZ'* in pJH391 with *bglG*.
23. The number of plaques in the plaque assay ranged from 0 for strains immune to the phage to 90 to 110 for strains sensitive to the phage.
24. H. R. Revel, *Virology* **31**, 688 (1967).
25. We thank M. Steinmetz for providing *sac Y* plasmids; J. Hu and R. T. Sauer for advice on the use of the λ repressor to study dimerization and for providing plasmids and bacterial strains; G. Tran Van Nhieu for advice on the native gel system; and A. L. Sonenshein and C. A. Kumamoto for critical reading of the manuscript.

16 March 1992; accepted 10 July 1992

MacroH2A, a Core Histone Containing a Large Nonhistone Region

John R. Pehrson*† and Victor A. Fried

A histone, macroH2A, nearly three times the size of conventional H2A histone, was found in rat liver nucleosomes. Its N-terminal third is 64 percent identical to a full-length mouse H2A. However, it also contains a large nonhistone region. This region has a segment that resembles a leucine zipper, a structure known to be involved in dimerization of some transcription factors. Nucleosomes containing macroH2A may have novel functions, possibly involving interactions with other nuclear proteins.

The nucleosome, as the basic structural unit of chromatin, must be adapted to diverse chromatin functions such as DNA packaging, transcription, replication, and mitotic chromosome segregation. Replication and transcription, for instance, require a disruption of nucleosomal structures to allow polymerases and accessory factors access to the DNA. The mechanisms used to alter nucleosomal structures are not well understood.

One way of varying nucleosome structure and function is to alter its protein composition. To identify proteins that may be involved in producing specialized nucleosome structures, we looked for proteins that remained bound to mononucleosomes during sedimentation in 0.5 M NaCl, conditions in which the core histones remain bound, but H1 and most nonhistone proteins are removed (*1*). Two proteins with a molecular mass of ~42 kD cosedimented with mononucleosomes under these conditions (Fig. 1A). These two proteins had similar electrophoretic mobilities in conventional SDS-polyacrylamide gels but were resolved in less cross-linked gels (Fig. 1B). Two proteins of identical electrophoretic properties cochromatographed with mononucleosomes and oligonucleosomes through molecular sieve columns (*2*); a preparation from oligonucleosomes is shown in Fig. 1B. The copurification of these two proteins with mononucleosomes during both sedimentation and gel filtration, which are oppositely affected by molecular shape, and also with oligonucleosomes, indicated that they were associated with the nucleosome.

Proteins other than the core histones and 42-kD proteins were also present in the mononucleosome region of the gradient (Fig. 1A). The proteins at the top of the gel probably sediment in this region because of their large size rather than because of their association with nucleosomes. However, some of the smaller proteins present in this region may be associated with nucleosomes.

To examine the association of the 42-kD proteins with the nucleosome more closely, we used a stepwise salt elution procedure developed by Simon and Felsenfeld (*3*) to resolve different core histones. The 42-kD proteins were present in the H2A and H2B fraction. Relative to H2A and H2B, the 42-kD proteins were diminished at the beginning of the elution of H2A and H2B and enriched at the end (Fig. 1C, lanes 3 to 5). Most importantly, the later fractions contained less H2A than H2B (Fig. 1C, lanes 5 and 6); scans of lane 6 indicate the deficit of H2A to be approximately 25% (*4*). In the early fractions, H2A and H2B are present in essentially equal amounts (Fig. 1C, lane 7), consistent with observations that they are eluted from the chromatin as a heterodimer. The gel scans indicated that the deficit of H2A in later fractions could be accounted for by the presence of the 42-kD proteins on a molar basis. The simplest explanation for the deficit of H2A, therefore, is that the 42-kD proteins form dimers with H2B. To confirm that this band contained the two 42-kD proteins that copurified with mononucleosomes and oligonucleosomes, we ran samples from this experiment in an SDS gel with low cross-linking (Fig. 1C, lanes 8 to 10).

The two 42-kD proteins were excised from SDS gels (Fig. 1B) and digested with trypsin or V8 protease for peptide mapping (*5*). These maps indicated that the two proteins have related primary structures. Of five peptides sequenced (*6*), three proved to be similar to regions of histone H2A, and the other two were unlike any protein in the SwissProt database. A polymerase chain reaction (PCR) primed with degenerate oligonucleotides based on the peptide sequences ARDNKK and KEFVEA obtained from the 42-kD proteins amplified a DNA fragment of ~550 bp from a rat lymphoma cDNA library (*7*). This fragment encoded sequences identical to the H2A-like peptides obtained from the 42-kD proteins (*8*).

The cloned PCR product was used to screen a rat liver cDNA library (*9*). Sequencing of a positive clone (*10*) revealed a single large open reading frame with an ATG context conforming to the consensus sequence of Kozak (*11*), indicating a probable site for initiation. This ATG is preceded by two in-frame stop codons. The open reading frame encodes a 39-kD protein consisting of 367 amino acids and contains sequences identical to all five peptides obtained from the 42-kD proteins (Fig. 2A). The NH_2-terminal third (amino acids 1 to 122) of the encoded protein is typical of full-length H2As, diverging mainly in regions known to differ between H2A subtypes within an organism and between species (*12*) (Fig. 2B).

The similarity of the NH_2-terminal region of this clone to H2A, together with the close association of the 42-kD proteins with nucleosomes and their apparent association with H2B, indicates that the 42-kD proteins are forms of H2A that occupy the place of conventional H2A in a subset of nucleosomes. We have accordingly named these proteins macroH2A (mH2A) and the subtype encoded by the cloned cDNA mH2A.1. The exact relation between mH2A.1 and the two mH2A bands seen in an SDS gel (Fig. 1B) is not known. On the basis of the relative staining intensity of gel bands, we estimate that there is about one mH2A for every 30 nucleosomes in rat liver.

Although the data presented above in-

J. R. Pehrson, Fox Chase Cancer Center, Institute for Cancer Research, Philadelphia, PA 19111.
V. A. Fried, Department of Cell Biology and Anatomy, Basic Sciences Building, New York Medical College, Valhalla, NY 10595.

*To whom correspondence should be addressed.
†Present address: Department of Animal Biology, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA 19104.

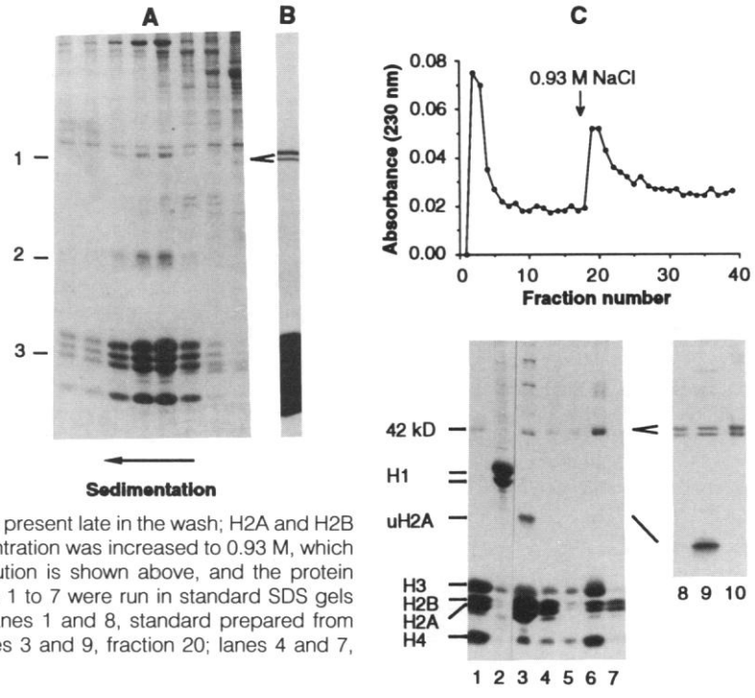

Fig. 1. (**A**) Detection of 42-kD proteins that cosediment with rat liver mononucleosomes in 0.5 M NaCl. Mononucleosomes were sedimented in a sucrose gradient (*24*). The gradient was fractionated and the protein composition of the fractions analyzed by electrophoresis in 15% polyacrylamide gels containing SDS (*25*). 1, 42-kD proteins; 2, ubiquitinated H2A (22.4 kD); 3, conventional H2A (14 kD). (**B**) Preparation of 42-kD proteins from oligonucleosomes. Oligonucleosomes were purified (*26*), extracted with 0.4 N H_2SO_4, and the extracted proteins resolved by electrophoresis in SDS-polyacrylamide gels (*25*) with low cross-linking (12% acrylamide, 1:125 bisacrylamide). (**C**) Salt elution of 42-kD proteins from rat liver chromatin. Chromatin fragments were bound to a hydroxylapatite column in 0.63 M NaCl, 0.1 M potassium phosphate (pH 6.7) and washed with the same buffer. At this salt concentration, H1 and most nonhistone proteins are removed from the chromatin (*3*). The 42-kD proteins were absent from these fractions except for traces present late in the wash; H2A and H2B were also present in these later fractions. The NaCl concentration was increased to 0.93 M, which removes H2A and H2B. The absorption profile of the elution is shown above, and the protein composition of individual fractions is shown below. Lanes 1 to 7 were run in standard SDS gels (*25*), and lanes 8 to 10 in gels with low cross-linking. Lanes 1 and 8, standard prepared from purified chromatin fragments (*26*); lane 2, fraction 2; lanes 3 and 9, fraction 20; lanes 4 and 7, fraction 25; lanes 5, 6, and 10, fraction 39.

A

```
SSRGGKKKST KTSRSAKAGV IFPVGRMLRY IKKGHPKYRI GVGAPVYMAA VLEYLTAEIL 60
ELAGNAARDN KKGRVTPRHI LLAVANDEEL NQLLKGVTIA SGGVLPNIHP ELLAKKRGSK 120
GKLEAIITPP PAKKAKSPSQ KKPVAKKTGG KKGARKSKKQ GEVSKAASAD STTEGAPTDG 180
FTVLSTKSLF LGQKLQVVQA DIASIDSDAV VHPTNTDFYI GGEVGSTLEK KGGKEFVEAV 240
LELRKKNGPL EVAGAAVSAG HGLPAKFVIH CNSPVWGADK CEELLEKTVK NCLALADDRK 300
LKSIAFPSIG SGRNGFPKQT AAQLILKAIS SYFVSTMSSS IKTVYFVLFD SESIGIYVQE 360
MAKLDAN                                                         367
```

B

```
                1                  10                  20                  30
mH2A.1 --SSRG---GKKKSTKTSRSAKAGVIFPVGRMLRYIKKG
H2A.2  --SGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKG
H2A.Z  AGGKAGKDSGKAKTKAVSRSERAGLQFPVGRIHRHLKSR
       40                  50                  60                  70                  80
HPKY-RLGVGAPVYMAAVLEYLTAEILELAGNAARDNKKGRVTPRHL
NYAE-RVGAGAPVYMAAVLEYLTAEILELAGNAARDNKKTRIIPRHL
TTSHGRVGATAAVYSAATLEYLTAEVLELAGNASKDLKVKRITPRHL
                90                  100                 110                 120
LLAVANDEELNQLLKGVTIASGGVLPNIHPELLAKKRGS----KGK
QLAIRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK
QLAIRGDEELDSLI-KATIAGGGVIPHIHKSLIGKKGQQKTV
```

Fig. 2. (**A**) Amino acid sequence of mH2A.1. The DNA sequence was sent to GenBank, accession number M99065. Sequences identified in peptides prepared from mH2A are underlined (*27*). The basic region following the H2A region is shown in bold, as are the residues in the d position of the putative leucine zipper. (**B**) Comparison of the H2A region of mH2A.1 (residues 1 to 122) to mouse H2A.2 (*28*) and H2A.Z (*29*). Boxes indicate regions of identity and (–) indicates a gap. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

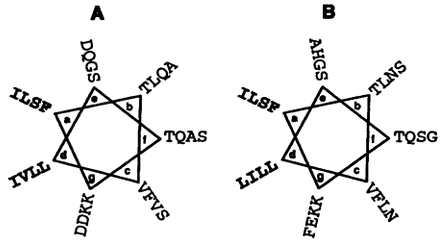

Fig. 3. Display of the leucine zipper–like region on a helical wheel with a heptad repeat. (**A**) mH2A.1, amino acids 181 to 208. (**B**) PCR product, the region of similar location to residues 181 to 208 of mH2A.1.

dicate that the H2A region of mH2A has retained general H2A function, including that of association with the nucleosome core, the divergence of this region from other mammalian H2As suggests that it may have specialized properties. A similar situation is seen with H2A.Z, an H2A subtype that is only 59% identical to conventional H2A. It has been reported that H2A.Z is enriched in transcriptionally active chromatin (*13*), and thus its divergence from conventional H2A appears to be functionally significant. Although H2A.Z and the H2A region of mH2A.1 are only 50% identical, both have retained similar regions of homology to conventional H2A (Fig. 2B).

The H2A region of mH2A.1 is followed by a highly basic region (residues 132 to 159). This region is similar to a number of proteins, most of which are known to bind nucleic acids. The greatest similarity is to the COOH-terminal region of histone H1 [57% identity to sea urchin H1γ (*14*) over 30 residues], which is thought to interact with the DNA between neighboring nucleosomes (*15*). Thus, it seems likely that this region of mH2A.1 interacts with DNA.

The region from amino acids 181 to 208 of mH2A.1 has several features of a leucine zipper (Fig. 3A), a helical structure involved in the dimerization of some transcription factors (*16*). The most important positions for dimerization by leucine zippers are the ones marked "a" and "d," which usually contain nonpolar residues and are the sites of interaction between the helices (*17*). Although many leucine zippers have only leucines in position d, there are several examples in which one or two of these

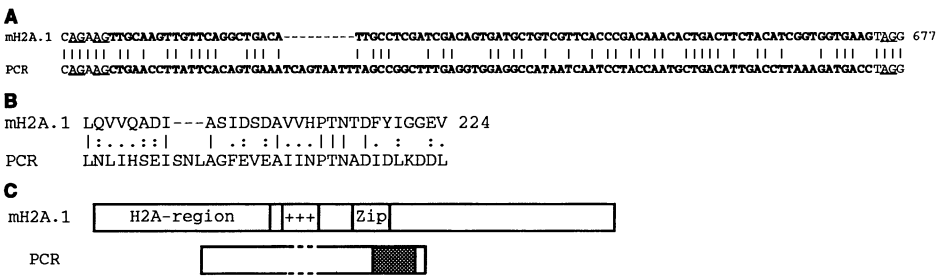

Fig. 4. Comparison of nucleotide and encoded amino acid sequence of the PCR product and mH2A.1. (**A**) Nucleotide sequences of the region of nonidentity. Sequence of PCR product sent to GenBank, accession number M99066. The putative splice donor sequences, AG, are underlined. (**B**) Amino acid sequences of the region of nonidentity. (|) indicates identities, (:) indicates conservative substitutions, and (.) indicates less conservative substitutions. The gap is indicated by (---). (**C**) Relation between the PCR product and mH2A.1. The region of nonidentity with mH2A.1 is shown as a filled box. The dashed lines indicate the region of the PCR product that has not been sequenced. (+++) indicates the basic region from residues 132 to 159, and Zip indicates the leucine zipper–like region (residues 181 to 208).

residues are not leucines (*18*), as in the case of mH2A.1. The positions flanking this region are marked "e" and "g" and typically contain hydrophilic residues (*17*). Proline residues, which would break the helical structure, are not found in leucine zippers or in this region of mH2A.1.

Comparison of the cDNA clone with the partial sequences of the PCR product revealed a region of nonidentity. Although the 185-nucleotide sequence at the 5′ end of the PCR product was identical to that of the mH2A.1 clone, the 3′ end had a region of low homology flanked by regions of complete identity (Fig. 4A), which suggests that mH2A mRNA may be subject to alternate splicing. Consistent with this possibility is the presence of the consensus sequence (AG) for the end of the splice donor (*19*) near the borders of the region of nonidentity at positions consistent with splicing (Fig. 4A).

The amino acid sequence encoded by the region of nonidentity is only 24% identical to the similar region in mH2A.1 and is three residues longer (Fig. 4B). Most of the substitutions, however, are conservative. The region of nonidentity begins in the middle of the leucine zipper–like structure at amino acid 195 of mH2A.1 (Fig. 4C), and when displayed on a helical wheel, the amino acid sequence encoded by the PCR product is also consistent with the formation of a leucine zipper (Fig. 3B); in this case, three of the four residues at position d are leucines. If this region of mH2A is involved in protein-protein interactions, then these alternate structures could have different specificities of interaction.

MacroH2A is apparently the result of a fusion of histone and nonhistone structures in a single gene product. The only other known linkage of histone and nonhistone structures is the posttranslational ubiquitination of histones H2A and H2B (*20*), the function of which is unclear. Another product of the fusion of histone and nonhistone sequences may be CENP-A, an unusual form of histone H3 localized to centromeric heterochromatin (*21*). However, because it is not substantially larger than conventional H3, it is probably a highly diverged subtype of H3.

The unique combination of features seen in mH2A.1 suggests that it has functions different from other known histones. One possibility is that it helps to produce sequence-specific nucleosome positioning. Nucleosome positioning appears to be important in the regulation of some genes (*22*) and can influence the activity of an autonomously replicating sequence (*23*). MacroH2A could produce specific nucleosome positioning by binding DNA with sequence specificity or by interacting with a protein that binds specific sites in the chromatin. The latter possibility is supported by the presence of a region that resembles a leucine zipper. MacroH2A-directed nucleosome positioning could serve to regulate chromatin function or contribute to the formation of specific chromosomal structures.

REFERENCES AND NOTES

1. P. J. G. Butler and J. O. Thomas, *J. Mol. Biol.* **140**, 505 (1980).
2. Nucleosomes were chromatographed through BioGel A-1.5m (Bio-Rad) in 0.5 M NaCl
3. R. H. Simon and G. Felsenfeld, *Nucleic Acids Res.* **6**, 689 (1979).
4. Data not shown.
5. Tryptic and V8 protease peptides were prepared [C. Eckerskorn and F. Lottspeich, *Chromatographia* **28**, 92 (1989); D. W. Cleveland, S. Fischer, M. W. Kirschner, U. K. Laemmli, *J. Biol. Chem.* **252**, 1102 (1977)].
6. Peptides were sequenced as described by Y. Yarden *et al.* [*Nature* **323**, 226 (1986)].
7. A PCR reaction of 40 cycles of 1 min at 95°C and 5 min at 58°C was used.
8. The PCR product was cloned into pBluescript SK$^-$ (Stratagene) and sequenced [F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977)] with deaza G (U.S. Biochemicals Corporation).
9. The H2A sequences were removed from the PCR product by digestion with Bam HI, and the subfragment was used to screen a rat liver cDNA library. We switched to a rat liver library to obtain a clone derived from normal tissue.
10. Both strands were sequenced with the use of subclones generated by Bal 31 and exonuclease III digestion [J. Sambrook, E. F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989).
11. M. Kozak, *Nucleic Acids Res.* **15**, 8125 (1987).
12. D. Wells and D. Brown, *ibid.* **19**, 2173 (1991).
13. F. Gabrielli, R. Hancock, A. J. Faber, *Eur. J. Biochem.* **120**, 363 (1981); C. D. Allis, Y. S. Ziegler, M. A. Gorovsky, J. B. Olmsted, *Cell* **31**, 131 (1982).
14. J. A. Knowles, Z.-C. Lai, G. J. Childs, *Mol. Cell. Biol.* **7**, 478 (1987).
15. D. J. Clark, S. S. Hill, S. R. Martin, J. O Thomas, *EMBO J.* **7**, 69 (1988).
16. W. H. Landschulz, P. F. Johnson, S. L. McKnight, *Science* **240**, 1759 (1988); T. Kouzarides and E. Ziff, *Nature* **336**, 646 (1988); P. Sassone-Corsi, L. J. Ransone, W. W. Lamph, I. M. Verma, *ibid.*, p. 692; R. Turner and R. Tjian, *Science* **243**, 1689 (1989); R. Gentz, F. J. Rauscher III, C. Abate, T. Curran, *ibid.*, p. 1695; M. Schuermann *et al.*, *Cell* **56**, 507 (1989).
17. E. K. O'Shea *et al.*, *Science* **254**, 539 (1991).
18. F. Katagiri, E. Lam, N.-H. Chua, *Nature* **340**, 727 (1989); C. R. Vinson, P. B. Sigler, S. L. McKnight, *Science* **246**, 911 (1989); I. V. Subramanian, D. L. Davis, S. N. Seal, J. E. Burch, *Mol. Cell. Biol.* **11**, 4863 (1991).
19. S. M. Mount, *Nucleic Acids Res.* **10**, 459 (1982).
20. I. L. Goldknopf and H. Bush, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 864 (1977).
21. D. K. Palmer, K. O'Day, H. L. Trong, H. Charbonneau, R. L. Margolis, *ibid.* **88**, 3734 (1991); B. Kingwell and J. B. Rattner, *Chromosoma* **95**, 403 (1987).
22. A. Almer, H. Rudolph, A. Hinnen, W. Horz, *EMBO J.* **5**, 2689 (1986); H. Richard-Foy and G. L. Hager, *ibid.* **6**, 2321 (1987); S. Y. Roth, A. Dean, R. T. Simpson, *Mol. Cell. Biol.* **10**, 2247 (1990).
23. R. T. Simpson, *Nature* **343**, 387 (1990).
24. Rat liver nuclei were isolated and digested with micrococcal nuclease (*30*). The digested nuclei were pelleted, and the supernatant, containing mononucleosomes [K. S. Bloom and J. N. Anderson, *Cell* **15**, 141 (1978)], was layered onto a 5 to 20% sucrose gradient containing 0.5 M NaCl, 10 mM triethanolamine, and 1 mM EDTA (pH 7.5), which was spun at 25,000 rpm in a Beckman SW-28 rotor for 18 hours at 4°C.
25. U. K. Laemmli, *Nature* **227**, 680 (1970).
26. Chromatin fragments were prepared and precipitated (*30*). They were dissolved in 0.5 M NaCl, 10 mM triethanolamine, and 1 mM EDTA (pH 7.5) and chromatographed through BioGel A-1.5m (Bio-Rad) equilibrated and run at 4°C with the same buffer.
27. All underlined residues were unambiguously identified except those corresponding to positions 73, 74, 232, 331, and 335 in mH2A.1.
28. M. M. Hurt, N. Chodchoy, W. F. Marzluff, *Nucleic Acids Res.* **17**, 8876 (1989).
29. C. L. Hatch and W. M. Bonner, *ibid.* **16**, 1113 (1988).
30. J. R. Pehrson, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9149 (1989)

31. We thank L. H Cohen for facilities, support, suggestions, and critical reading of the manuscript; W. Li and A. Cywinski for technical assistance; A. Bellacosa, J. Burch, J. Pugh, M. Radic, J. Hittle, D. Davis, and C. Patroitis for technical suggestions; C Patroitis for supplying the cDNA libraries, and A. Zweidler, R. Perry, M. Atchison, N. Avadhani, I. Akerblom, and J. Burch for comments on the manuscript. Supported by grants GM 24019, CA 06927, and RR 05539 from the National Institutes of Health and by an appropriation from the Commonwealth of Pennsylvania

24 March 1992, accepted 6 July 1992