Pachyonychia Congenita Project

15 March 2005

Use of Articles in the Pachyonychia Congenita Bibliography

The articles in the PC Bibliography may be restricted by copyright laws. These have been made available to you by PC Project for the exclusive use in teaching, scholarship or research regarding Pachyonychia Congenita.

To the best of our understanding, in supplying this material to you we have followed the guidelines of Sec 107 regarding fair use of copyright materials. That section reads as follows:

> Sec. 107. - Limitations on exclusive rights: Fair use
> Notwithstanding the provisions of sections 106 and 106A, the fair use of a copyrighted work, including such use by reproduction in copies or phonorecords or by any other means specified by that section, for purposes such as criticism, comment, news reporting, **teaching (including multiple copies for classroom use), scholarship, or research**, is not an infringement of copyright. In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall include - (1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4) the effect of the use upon the potential market for or value of the copyrighted work. The fact that a work is unpublished shall not itself bar a finding of fair use if such finding is made upon consideration of all the above factors.

We hope that making available the relevant information on Pachyonychia Congenita will be a means of furthering research to find effective therapies and a cure for PC.

2386 East Heritage Way, Suite B, Salt Lake City, Utah 84109 USA
Phone +1-877-628-7300 • Email—Info@PC-Project.org
www.PC-Project.org

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 260, No. 10, Issue of May 25, pp. 5867-5870, 1985
Printed in U.S.A.

Communication

Organization of a Type I Keratin Gene

EVIDENCE FOR EVOLUTION OF INTERMEDIATE FILAMENTS FROM A COMMON ANCESTRAL GENE*

(Received for publication, January 8, 1985)

Thomas M. Krieg‡§, Millie P. Schafer¶, Christina K. Cheng‡, David Filpula‖, Patrick Flaherty‡, Peter M. Steinert, and Dennis R. Roop‡**

*From the ‡Laboratory of Cellular Carcinogenesis and Tumor Promotion and **Dermatology Branch, National Cancer Institute, the ¶Laboratory of Molecular Hematology, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20205, and ‖Genex Corporation, Gaithersburg, Maryland 20877*

The genomic structure of the mouse 59-kDa keratin gene, a Type I intermediate filament (IF) gene is presented. A comparison of the organization of this gene with that of the human 67-kDa keratin, a Type II IF gene, and hamster vimentin, a Type III IF gene, suggests a common evolutionary origin for Type I, II, and III IF genes. Most introns in these three types of IF genes occur at similar positions within the region encoding sequences predicted to form coiled-coils, but do not delineate structural subdomains. Interestingly though, most of the introns interrupt at or near the beginning of the characteristic 7-residue (heptad) repeat of sequences which form the coiled-coil. These data suggest that the three types of IF genes arose from a common ancestor which may have been assembled from smaller units containing multiple heptad repeats. Subsequent duplication events may then have formed the three known α-helical types and each of their various members.

Intermediate filaments (IF[1]) are major cytoskeletal components of most eucaryotic cells (1, 2). They have been classified into at least 5 distinct subclasses, each of which contains from 1 to 20 subunits encoded by separate genes (1, 3). These genes are differentially expressed in different tissues and during different stages of differentiation (reviewed in Refs. 1-6). An analysis of amino acid sequence information (7-14) has shown that all IF subunits possess a common secondary structure, consisting of a conserved central α-helical domain that can form four coiled-coil tracts and non-α-helical end domains of variable size and sequence (8, 11). The sequences that are predicted to form coiled-coils have been divided into three distinct types (2, 3) as follows: the acidic keratins form Type I, the neutral-basic keratins form Type II; and desmin, vimentin, and a glial fibrillary acidic protein form Type III. To explore the evolutionary relationship of these conserved and variable domains of IF, we have compared the genomic structure of the mouse 59-kDa keratin, a Type I IF gene, with that of a Type II IF gene, the human 67-kDa keratin, and a Type III IF gene, hamster vimentin.

A cDNA probe for the mouse 59-kDa keratin (15) was used to screen (16) a total genomic library containing fragments of a partial *Mbo*I digestion of Balb/c mouse embryo DNA cloned in λ Charon 28 (17). Nine phage clones were found to hybridize to the cDNA probe and one, λ59K-4, contained information from the 5'- and 3'-end of the molecule as judged by hybridization to probes constructed from different regions of the cDNA. Identical hybridization patterns were obtained when this phage DNA and total mouse DNA were subjected to restriction enzyme and DNA blot analysis (data not shown). This indicates that the 59-kDa keratin gene is present as a single copy in the mammalian genome as was recently shown for the chicken (18) and the hamster vimentin genes (12).

Initially we were able to determine that there were at least 8 exons within the 59-kDa keratin gene and to estimate their size by S1 nuclease mapping (data not shown). By R-loop analysis (Fig. 1), the genomic clone λ59k-4 was shown to encode the complete sequence of the gene which was ~4.5 kb in length and contained 7 introns. With the exception of the first intron, which was found to be ~615 bp by R-loop analysis, all of the exons and introns were subjected to direct sequence analysis (Fig. 2). The putative cap site within the first exon was determined from a primer extension experiment (9). Characteristic sequences for a TATA box and CAT box present in many other eucaryotic genes (20) are found 26 and 79 bp upstream from the cap site. The last exon contains two polyadenylation signals (AAT AAA) (21) which are separated by 4 bp.

The sequence of the genomic clone was similar to our previously published sequence for the cDNA clone (9), although several differences were noted. Accordingly, we resequenced the cDNA clone using chain-termination procedures (22, 23) and believe that most of the differences are due to sequencing errors or variations with the chemical degradation methods of Maxam and Gilbert (24). Many of the differences were single base changes which did not affect the original deduced amino acid sequence. However, in three other cases, we have corrected the amino acid sequence between residues 178 and 187 from VVREARQLKP to WYEKHGNSSQ, between residues 263 and 268 from QSVLEL to KSDMEM, and between residues 394 and 399 from VESLLR to EGRYCV. None of these changes result in an alteration in the predicted secondary structure of the protein. After modifying the cDNA sequence in this way, only one difference remained, which may represent a transcriptional error that arose during the construction of the cDNA clone.

It was of obvious interest to determine if the introns present within this keratin gene occur at locations delineating functional domains as has been observed for several other eucaryotic structural genes (25, 26). We also wanted to compare the structure of this Type I IF gene with that of other IF gene types. This was possible since the structure of a Type III IF

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Recipient of a Heisenberg Grant of the Deutsche Forschungsgemeinschaft.

[1] The abbreviations used are: IF, intermediate filament; kb, kilobases; bp, base pairs.

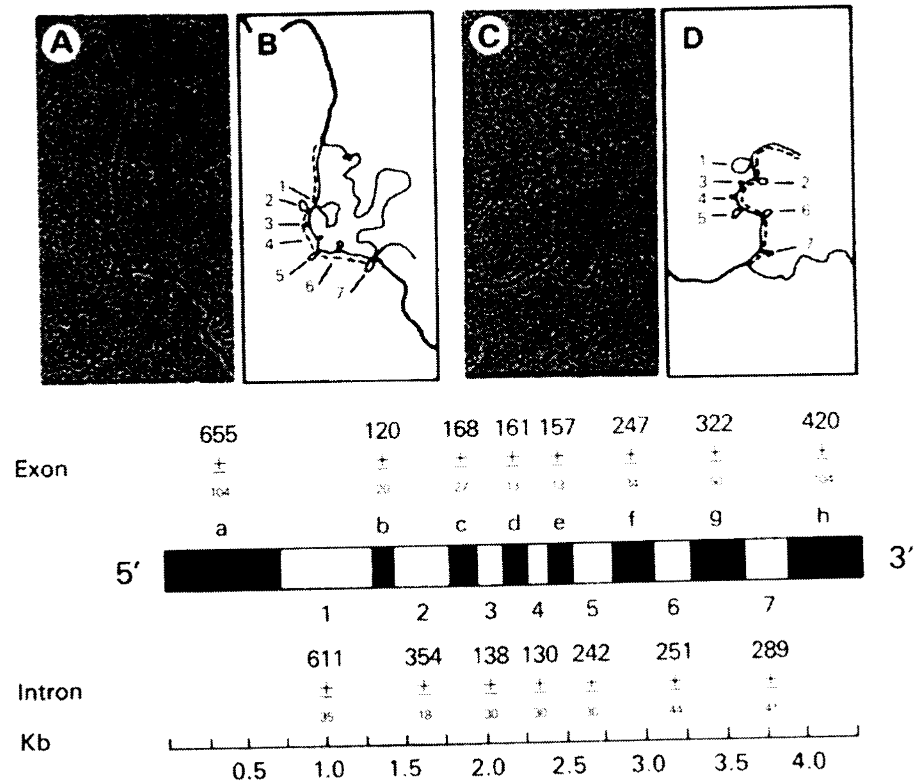

FIG. 1. **The exon-intron structure of the 59-kDa mouse keratin gene.** Previously published conditions for R-looping were utilized (19). Mouse epidermal poly(A) RNA was hybridized to the mouse λ59K-4 clone. *A* and *B*, shown are the electron microscopic visualizations of the introns and exons. The introns are labeled *1–7*. The exons are illustrated in the interpretative tracing (-----). Intron 3 and 4 are barely visible. *C* and *D*, a clear representation of intron 3 and 4 is depicted. One arm of Charon 28 is missing. The bottom illustration shows the exon (▬) and intron (▭) order and size within the λ59K-4 clone. The sizes of all exons and introns were determined using exon d and intron 5 as standards; the standards were sized by direct sequence analysis.

gene, hamster vimentin, has been described (12) and the structure of a Type II IF gene, the human 67-kDa keratin, has been recently completed.[2] Six of 7 introns within the 59-kDa keratin gene are located within the region predicted to form coiled-coil tracts (Fig. 3). Identical locations for several introns were found in the genes coding all three IF types. Some of these introns are found at or near positions defining functional domains of the protein. One is located near the end of the coiled-coil tract 1B, one near the beginning of the 2B tract, and the third near the end of the 2B tract, which separates this coiled-coil from the non-α-helical carboxyl-terminal end domain. The other introns located within the coiled-coil tracts do not appear to delineate functional domains or subdomains. However, by close inspection it is evident that many of these introns splice in a nonrandom manner at the beginning of the characteristic heptad repeat within the coiled-coil tracts (indicated by *arrows* with *solid circles*, Fig. 3), *e.g.* 4 of 6 in the Type I IF gene, 3 of 6 in the Type II IF gene, and 3 of 6 in the Type III IF gene.

This striking occurrence of the common locations of several introns in the α-helical domains of the three types of IF genes suggests that these domains arose from a common ancestral gene. Moreover, it is tempting to speculate that the putative progenitor IF gene was assembled from exons containing multiple heptad repeats. A similar formation of a progenitor gene from a multiple exon unit has already been proposed for the gene family coding for different collagen types (27). Two early duplications of the progenitor IF gene would have been required to form the three types of IF genes, and more recent duplications would have generated other members within each IF type. The presence of unique introns in the Type I and Type II IF genes suggests that the progenitor gene may have had additional introns within this domain that were subsequently lost during the evolution of the different IF genes. If IF genes did evolve by this mechanism, one would predict that genes of a given Type might be linked. Indeed, we have recently found by restriction fragment length polymorphism analysis that two mouse Type II IF genes, the 60- and 67-kDa keratin genes, are located on chromosome 15 and that they are linked.[3]

The other introns present within the IF genes are located in the carboxyl-terminal end domain. One of these introns occurs at a similar location in both the Type II and Type III genes; however, this intron does not appear to demarcate a functional domain since the amino acids encoded by the interrupted sequences are quite different. The other introns are located at unique positions within the Type I and Type III genes and occur near the boundary between coding and 3′ noncoding sequences; the last exon of the Type III gene encodes 13 amino acids and the last exon of the Type I gene only encodes a single amino acid. The end domains of different IF subunits in general show wide divergence in size and amino acid sequence (2, 11, 28). However, a pattern is emerging that co-expressed pairs of Type I and Type II keratins have similar subdomains; *e.g.* terminally differentiating mouse epidermal cells express the 59- (Type I) and 67-kDa (Type II) keratins which share similar glycine-rich end domains (3, 9, 10). We have suggested that these end domain sequences are important in defining the functions of the IF in which the subunits are present (9, 11). Consistent with the variability of the end domains, we suspect that future data may show wide variations in the locations of introns in the end domains of different IF genes.

Although IF genes contain numerous introns within the coiled-coil domain, this is not a general feature of other genes encoding coiled-coil regions such as fibrinogen (29) and myosin (30). The conservation of some intron positions if IF

[2] Johnson, L. D., Idler, W. W., Zhou, X.-M., Roop, D. R., and Steinert, P. M. (1985) *Proc. Natl. Acad. Sci. U. S. A.*, in press.

[3] C. Blatt and D. Roop, unpublished data.

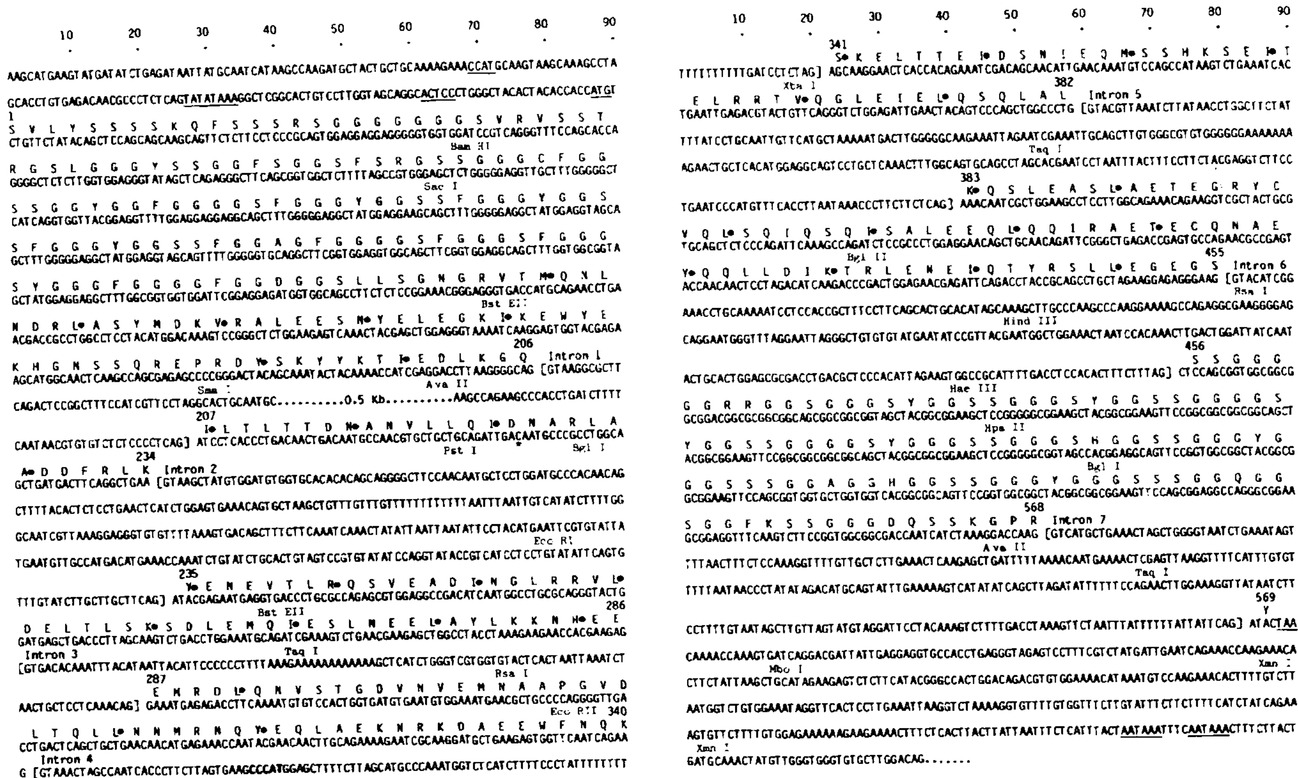

FIG. 2. **Sequence of the 59-kDa mouse keratin gene.** Appropriate regions of the 14-kb λ59K-4 clone were sequenced using chemical (24) as well as dideoxy chain termination sequencing methods (22, 23). The locations of restriction sites used for sequence analysis are indicated. Introns are given in *brackets*; exons were identified by comparison with the known cDNA sequence of this keratin (9) and by S1 nuclease mapping (not shown). The amino acid sequence encoded by the exons is given in the *single letter code above* the respective *codons*. The *spots adjacent* to certain *amino acids* demarcate the *a* or first position of the characteristic heptad repeat of the sequences that form a coiled-coil structure (9). Note the occasional interruption in this heptad repeat due to (i) the noncoiled-coil linkers and (ii) the reversal of heptad polarity or stutter (see Fig. 3). All introns contain consensus donor and acceptor sequences (32). The CAT-box, the TATA-box, cap-site, and the ATG-start codons are *underlined*. At the 3′ end of the gene, the stop codon TAA is followed by a long stretch of noncoding sequences, which contain two polyadenylation signals (AATAAA). An *asterick* indicates a difference between the sequence of the genomic clone and the cDNA (9).

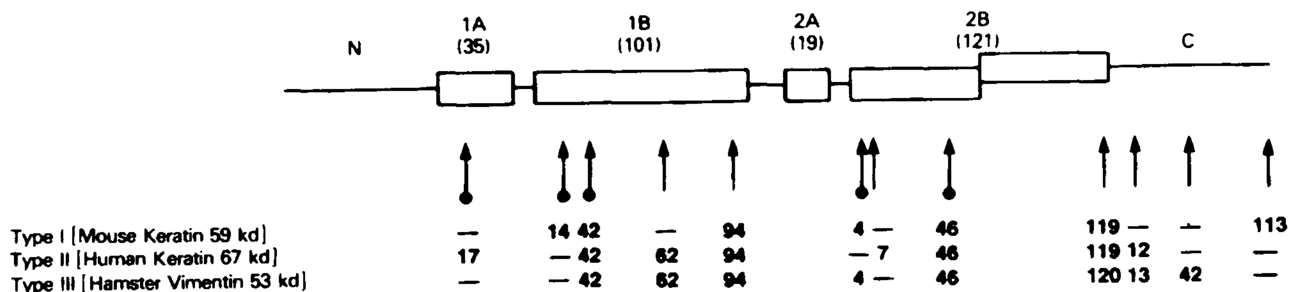

FIG. 3. **Localization of introns in Type I, II, and III IF genes.** All IF are composed of N- and C-terminally located non-α-helical domains and a large central α-helical domain, which can be subdivided into coiled-coil subdomains (*1A, 1B, 2A*, and *2B*). These are interrupted by short non-α-helical linkers (2, 10). The size (number of amino acids residues) of each subdomain is given in *brackets*. Domain *2B* contains a reversal of the heptad polarity ("stutter"), which is schematically indicated (2, 10). *Arrows* indicate the location of introns within each gene. *Arrows* containing *solid circles* are used to mark those introns which occur after position 7 of the 7-residue repeat (see Fig. 2). The *number below* each *arrow* indicates the exact intron position (amino acid residue) within each subdomain. The hamster vimentin (Type III) genomic structure has been published (12) and the location of introns in the gene coding for the human 67-kDa keratin (Type II) has recently been determined.[2]

genes is presumably important in maintaining the common secondary structure of α-helical domains of the IF proteins and reflects a more conservative evolution of this region as has been observed for other genes coding for rigid structural proteins such as collagens (31). It will be interesting to see if the unique features observed for each IF gene type compared in this study are characteristic of other members of that particular IF type.

Acknowledgments—We thank Dr. Philip Leder for supplying the

mouse genomic library, Dr. Dan Kuebbing for advice on dideoxy sequencing, Dr. Gerald Crabtree for communicating results prior to publication, and Dr. Stuart Yuspa for his enthusiastic support.

REFERENCES

1. Lazarides, E. (1980) *Nature* **283,** 249–256
2. Steinert, P. M., Jones, J. C. R. & Goldman, R. D. (1984) *J. Cell Biol.* **99,** 22s–27s
3. Fuchs, E. & Hanukoglu, I. (1983) *Cell* **34,** 332–334
4. Moll, R., Franke, W. W., Schiller, D. L., Geiger, B. & Krepler, R. (1982) *Cell* **31,** 11–24
5. Eichner, R., Bonitz, P. & Sun, T. T. (1984) *J. Cell Biol.* **98,** 1388–1396
6. Roop, D. R., Toftgard, R., Kronenberg, M. S., Clark, J. H. & Yuspa, S. H. (1985) in *The Molecular Biology of the Cytoskeleton* (Borisy, G. G., Cleveland, D. W., and Murphy, D. B., ed) pp. 409–414, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
7. Hanukoglu, I. & Fuchs, E. V. (1982) *Cell* **31,** 243–252
8. Hanukoglu, I. & Fuchs, E. V. (1983) *Cell* **33,** 915–924
9. Steinert, P. M., Rice, R. H., Roop, D. R., Trus, B. L. & Steven, A. C. (1983) *Nature* **302,** 794–800
10. Steinert, P. M., Parry, D. A. D., Racoosin, E. L., Idler, W. W., Steven, A. C., Trus, B. L. & Roop, D. R. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81,** 5709–5713
11. Steinert, P. M., Idler, W. W., Zhou, X. M., Johnson, L. D., Parry, D. A. D., Steven, A. C. & Roop, D. R. (1985) *Ann. N. Y. Acad. Sci.*, in press
12. Quax, W., Egberts, W. V., Hendricks, W., Quax-Jeuken, Y. & Bloemendal, H. (1983) *Cell* **35,** 215–223
13. Geisler, N. & Weber, K. (1982) *EMBO J.* **1,** 1649–1656
14. Lewis, S. A., Balcarek, J. M., Krek, V., Shelanski, M. L. & Cowen, N. J. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81,** 2743–2746
15. Roop, D. R., Hawley-Nelson, P., Cheng, C. K. & Yuspa, S. H. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 716–720
16. Benton, W. L. & Davis, R. W. (1977) *Science* **196,** 180–182
17. Seidman, J. G. & Leder, P. (1978) *Nature* **176,** 790–795
18. Zehner, Z. E. & Paterson, B. M. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 911–915
19. Schafer, M. P., Boyd, C. D., Tolstoshev, P. & Crystal, R. G. (1980) *Nucleic Acids Res.* **8,** 2241–2253
20. Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C. & Proudfoot, N. J. (1980) *Cell* **21,** 653–668
21. Proudfoot, N. J. & Brownlee, G. G. (1976) *Nature* **263,** 211–214
22. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
23. Messing, J., Crea, R. & Seeburg, P. (1980) *Nucleic Acids Res.* **9,** 309–321
24. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 500–560
25. Gilbert, W. (1978) *Nature* **271,** 501
26. Craik, C. S., Sprang, S., Flettenick, R. & Rutter, W. J. (1982) *Nature* **299,** 180–182
27. Yamada, Y., Avvedimento, V. E., Mudryj, M., Ohkubo, H., Vogeli, G., Irani, M., Pastan, I. & deCrombrugghe, B. (1980) *Cell* **22,** 887–892
28. Roop, D. R., Cheng, C. K., Titterington, L., Meyers, C. A., Stanley, J. R., Steinert, P. M. & Yuspa, S. H. (1984) *J. Biol. Chem.*, **259,** 8037–8040
29. Crabtree, G. R., Comeau, C. M., Fowkles, D. M., Fornace, A. J., Malley, J. D. & Kant, J. A. (1985) *J. Mol. Biol.*, in press
30. Karn, J., Brenner, S. & Barnett, L. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 4253–4257
31. Yamada, Y., Lian, G., Mudryj, M., Obici, S. & deCrombrugghe, B. (1984) *Nature* **310,** 333–337
32. Mount, S. M. (1982) *Nucleic Acids Res.* **10,** 459–472