*Proc. Natl. Acad. Sci. USA*
Vol. 79, pp. 739–743, February 1982
Biochemistry

# Cytochrome *c* gene-related sequences in mammalian genomes

(gene families/genomic hybridization/gene heterogeneity/pseudogenes)

RICHARD C. SCARPULLA, KATHLEEN M. AGNE, AND RAY WU

Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, New York 14853

*Communicated by Leon A. Heppel, October 6, 1981*

**ABSTRACT** **We use a rat cytochrome *c* gene that we previously isolated and determined the sequence of to estimate the number of related sequences present in the rat genome. Approximately 25 different *Eco*RI restriction endonuclease fragments from total rat DNA hybridize to the gene of known structure. Four of these correspond to homologous sequences present in four different λ Charon 4A–rat cytochrome *c* recombinants previously isolated. Intact or nearly intact genes appear to reside on almost all of the genomic fragments, because they hybridize strongly to gene subfragments representing both 5′ and 3′ portions of the coding sequence as well as to 3′ noncoding DNA that is found specifically associated with the coding region. A subgroup of about six of the fragments also shares homology within the 73 nucleotides immediately preceding the AUG codon. An intron-specific probe reveals only the *Eco*RI fragment from which it was derived and one other genomic fragment. On the basis of the temperature of complete dissociation of the coding region probe in 0.75 M NaCl/0.075 M $Na_3$ citrate/50% (vol/vol) formamide, the 25 fragments are separable into three stringency classes of 40–50°C, 50–55°C, and 55–60°C. The latter, high-stringency group of about seven fragments includes those cloned in the recombinant phage isolates, whose regions homologous to cytochrome *c* are shown to differ from the purified gene of known sequence by an amount equivalent to about 2% mismatched bases. Families of cytochrome *c* gene-related sequences are also found in the genomes of several other mammals, including humans.**

Cytochrome *c* is a phylogenetically ancient protein that functions as a carrier of electrons to their terminal acceptor, oxygen (1). Multiple forms of cytochrome *c* are physically distinguished in preparations from many organisms, but most of this heterogeneity is attributed to posttranslational or artifactual modifications rather than to genetic variability (1, 2). One notable exception is a protein purified from mouse testis that differs from the adult cytochrome in 13 of 104 amino acid residues, strongly implying the presence of a second gene in this organism (3). Yeast possesses two distinct forms of cytochrome *c*, iso-1 and iso-2, which are the products of different genetic loci (4, 5). The nucleotide sequence of genes encoding each of these proteins is known (6, 7).

In order to investigate the structure and organization of cytochrome *c* genes in higher organisms, we used the iso-1-cytochrome *c* gene of yeast as a hybridization probe to isolate cytochrome *c*-specific sequences from a λ Charon 4A–rat genomic library (8). Characterization and sequence analysis of Ch4A-RC4, one of eight recombinant clones homologous to the yeast probe, revealed a cytochrome *c* gene encoding the entire rat polypeptide chain and containing a 105-base pair intron dividing the coding region. Each of the eight isolates differed both in their restriction enzyme digest pattern and in the size of the restriction fragment homologous to the yeast gene, suggesting that several nonallelic cytochrome *c* genes or pseudogenes may be present in the rat genome. Because higher organisms are generally thought to possess only a single form of cytochrome *c* polypeptide (1, 2) it is of interest to examine in detail the number and structure of cytochrome *c*-specific sequences in mammalian genomes. We now use the cloned rat gene of Ch4A-RC4 to study the complexity of cytochrome *c* gene sequences in the genomes of rat and of several other mammals.

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

## MATERIALS AND METHODS

The plasmid pRC4 containing a rat cytochrome *c* gene was subcloned from the λ–rat recombinant phage Charon 4A-RC4 and the sequence of the 0.96-kilobase (kb) gene insert was determined (8). A second plasmid, having an additional 187 base pairs of the 5′-flanking DNA of the same gene extending to a *Hin*dIII site on the rat genomic fragment, was similarly constructed for these studies and designated pRC4-2. The other λ–rat recombinant clones Charon 4A-RC-1 through -8 were those previously described (8). Total rat DNA was isolated from the liver of a single Sprague–Dawley rat (9), and human placental DNA (Sigma) was a commercial preparation. DNAs from Russian baboon, gorilla muscle, and feline embryo cell line A were generous gifts of M. Tainsky, Frederick Cancer Research Center. Restriction enzymes were purchased from New England BioLabs and used according to their recommended assay procedures. Digested DNAs were electrophoresed on 1% agarose gels (electrode buffer: 40 mM Tris/20 mM sodium acetate/2 mM EDTA, pH 7.8) and restriction fragments were transferred to nitrocellulose filters by the Southern procedure (10).

Subfragments of the rat cytochrome *c* gene to be used as probes were purified from the 1.14-kb pRC4-2 *Hin*dIII/*Eco*RI cloned insert by 4% acrylamide gel electrophoresis after cleavage with the appropriate enzymes. The pure restriction fragments representing different portions of the gene were $^{32}P$ labeled (11) and separated from unreacted dNTPs by gel filtration on Sephadex G-50. Hybridization and washing conditions for nitrocellulose blots were the same as those described (8) except that the incubation temperature was 40°C unless otherwise indicated. Approximately $5 \times 10^5$ cpm (Cerenkov) of labeled DNA was used per filter and the washed filters were exposed to Kodak medical x-ray film for 5 to 7 days at −20°C. The insulin-specific fragment used as the probe in control experiments was purified from a rat insulin A-chain clone, pSIA12, containing a 63-nucleotide duplex that has the single base change required for synthesis of the human insulin A-chain polypeptide (unpublished).

Thermal dissociation curves were constructed by hybridizing $^{32}P$-labeled probes under low-stringency conditions (8) to approximately 20 μg of denatured phage DNA immobilized on nitrocellulose filters (1.3 cm diameter). The amount of radioactivity eluted from the filter by incubation in 0.39 M $Na^+$ was measured at various temperatures (12). Each point represents the average of duplicate determinations.

Abbreviation: kb, kilobase(s).

## RESULTS

**Detection of Multiple Rat Genomic Restriction Fragments Complementary to the Coding Region of a Rat Cytochrome *c* Gene.** When we screened a Charon 4A–rat genomic library by using a yeast iso-1-cytochrome *c* gene probe, eight recombinants that differed by restriction enzyme and hybridization analysis were isolated (8). The rat genomic insert of one of these phages, Ch4A-RC4, was shown to contain a single copy of a cytochrome *c* gene encoding the entire rat polypeptide sequence. Using the 0.54-kb *Bam*HI/*Acc* I coding region subfragment of this rat gene as a probe, only four of the eight original clones (Ch4A-RC4, -5, -6, and -8) could be detected by hybridization analysis (Fig. 1). The other isolates (Ch4A-RC1, -2, -3, and -7) hybridized only to the yeast cytochrome *c* gene, and the sequence of at least one appears to represent a highly diverged copy of a yeast-related gene in the rat genome (13).

The occurrence in the Charon 4A-rat library of four nonallelic cytochrome *c* gene sequences that hybridize to both yeast and rat genes suggests that they represent different rat genes or pseudogenes for cytochrome *c*. In order to estimate the number of different cytochrome *c* gene sequences, *Eco*RI-digested genomic DNA was hybridized after gel electrophoresis and transfer to nitrocellulose to the $^{32}$P-labeled *Bam*HI/*Acc* I coding region subfragment (Fig. 3) of pRC4-2. Fig. 1 (lane B) shows approximately 25 genomic *Eco*RI restriction fragments homologous to the rat gene sequence. These include segments comparable in size to those generated by *Eco*RI cleavage of the four recombinant phage DNAs, which each have a single copy of the cytochrome *c* sequence. Some of the bands exhibit somewhat lower hybridization intensity than others, suggesting they have less homology with the probe. It is unlikely that the pattern of multiple bands detected with the rat cytochrome *c* probe is the result of incomplete digestion of the genomic DNA, because an identical filter (lane A) that is hybridized to an insulin A-chain probe reveals only the four insulin-specific bands previously shown to be present in the genome of a Sprague–Dawley rat (14). Also, no change in the hybridization banding pattern is observed after extensive redigestion of identical aliquots of the *Eco*RI-cut DNA used in lanes A and B.

Although none of the four recombinant phage DNAs contain *Eco*RI cleavage sites dividing their cytochrome *c* homologous sequences, it may be possible that the sequence complexity is due in part to *Eco*RI sites in sequences not represented by the four phage isolates. This explanation does not account for a large number of these fragments, because Fig. 2 shows that *Bam*HI and *Pst* I, which are expected to cut genomic DNA infrequently and do not cleave within the cytochrome *c* sequences of the four recombinant phage DNAs (8), also generate an equivalent number of hybridizing bands in total DNA.

**Relationship of the Cytochrome *c*-Specific Rat Genomic Sequences to the Structure of the Cytochrome *c* Gene from pRC4-2.** In order to rule out that the large number of genomic restriction fragments homologous to the *Bam*HI/*Acc* I probe is the result of a sequence present in the intron or noncoding regions that is repeated in the genome but not necessarily related to cytochrome *c*, it was important to establish more precisely which portions of the rat gene are responsible for the hybridization pattern. This was done by subdividing the gene, using conveniently located restriction sites shown in Fig. 3, and purifying gene fragments A through G for use as probes in hybridization to identical Southern filters of *Eco*RI-digested rat DNA. Fig. 3, lanes D and F, shows that a banding pattern essentially identical to lane A is obtained by using either exon I (lane D) or exon II (lane F) specific probes. Thus, the vast majority of the fragments have sequences complementary to both the 5′ and 3′ portions of the coding region. The only obvious exception is a 0.7-kb fragment that contains only exon II.

In marked contrast, the intron-specific probe (E) reveals only the 2.5-kb fragment from which the rat gene sequence is derived and one other band at about 23 kb, which may represent a duplicate copy of this gene. If the other cytochrome *c* gene-related sequences contain introns, they have little or no homology with that of pRC4-2. When fragment B of the 5′ flanking region is used as a probe, the same two genomic fragments of 2.5 and 23 kb hybridize, indicating that, like the intron, this sequence is associated with two of the bands and is not a general feature of cytochrome *c* gene-specific sequences. However, adjacent probe C, which is the segment immediately upstream to the start AUG, hybridizes to six of the genomic fragments. The members of this subclass have sizes of 1.6, 1.9, 2.2, 2.5, 4.3, and 4.9 kb and do not include *Eco*RI fragments from Ch4A-RC5, -6, and -8.

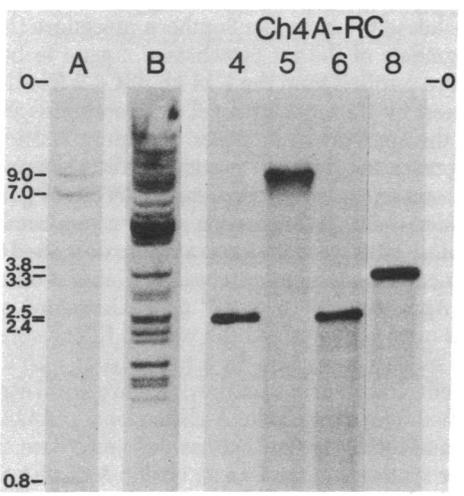


FIG. 1. Hybridization analysis of rat genomic DNA and Ch4A-RC recombinant phage DNAs digested with *Eco*RI. Hybridization was carried out at 40°C in 0.97 M $Na^+$ (0.75 M NaCl/0.075 M $Na_3$ citrate) and 50% (vol/vol) formamide. Lanes A and B contained 30 μg of *Eco*RI-digested DNA from a single Sprague–Dawley rat. Remaining lanes had 0.02 μg of *Eco*RI-digested DNA from the indicated Ch4A-RC recombinant clones. Filters were hybridized to the $^{32}$P-labeled *Bam*HI/*Acc* I cytochrome *c* gene fragment of pRC4-2 ($2.3 \times 10^7$ cpm/μg) except for lane A, where the insulin A chain gene from pSIA12 ($9.0 \times 10^6$ cpm/μg) was used as probe. On the left the lengths of DNAs are given in kb, as they are in following figures. O, origin.

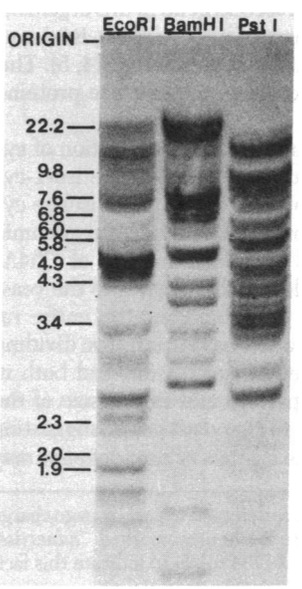


FIG. 2. Hybridization analysis of rat genomic DNA digested with *Eco*RI, *Bam*HI, or *Pst* I. Each lane contained 30 μg of total rat DNA hybridized to the $^{32}$P-labeled *Bam*HI/*Acc* I cytochrome *c* gene fragment of pRC4-2 ($2.3 \times 10^7$ cpm/μg). Size standards are restriction fragments from *Eco*RI- or *Hin*dIII-digested λ and *Hae* II-digested pBR322 DNA.

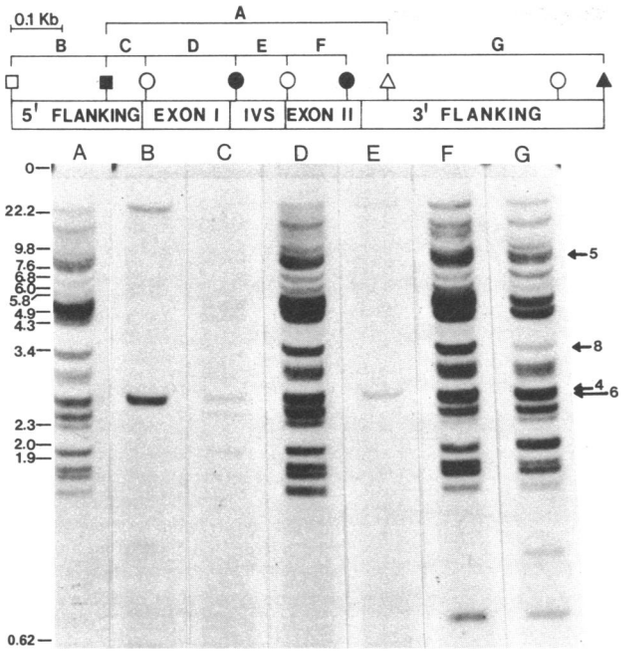


FIG. 3. Hybridization analysis of *Eco*RI-digested rat DNA by using flanking, exon, and intron specific probes from the cytochrome *c* gene of pRC4-2. Lanes A–G depict the autoradiographic patterns resulting from hybridization of corresponding $^{32}$P-labeled gene fragments *A* through *G*, shown in the restriction map at the top, to identical filters of *Eco*RI-digested rat DNA (30 $\mu$g). Specific activities for probes derived from fragments A through G were $2.3 \times 10^7$, $4.2 \times 10^7$, $8.0 \times 10^6$, $5.0 \times 10^7$, $2.0 \times 10^7$, $4.8 \times 10^7$, and $3.3 \times 10^7$ cpm/$\mu$g, respectively. Cleavage sites for restriction enzymes *Hin*dIII (□), *Bam*HI (■), *Hph* I (○), *Alu* I (●), *Acc* I (△), and *Eco*RI (▲) are shown. Standards are the same as in Fig. 2, and the positions of homologous fragments from Ch4A-RC4, -5, -6, and -8 are indicated by arrows. IVS, intervening sequence (intron).

The 3′ flanking region probe G gives essentially the same pattern observed with the coding region probes, with notable exceptions of a missing band(s) at about 5 kb and the presence of a new band at about 1.0 kb. Several other fragments display diminished hybridization intensity, suggesting less homology with the probe. Thus, sequences related to the 3′ flanking sequence are present in many copies in the rat genome and are almost always associated with the cytochrome *c* coding region. The presence of exon I, exon II, and 3′ flanking sequences in most, and the 5′ flanking sequence immediately adjacent to the start codon in many, of the cytochrome *c*-specific genomic fragments indicates that at least at this level of resolution intact or nearly intact cytochrome *c* gene copies reside on these segments.

**Thermal Dissociation of the Rat Cytochrome *c* Gene of pRC4-2 from Homologous Genomic Sequences.** The relatedness of the three homologous *Eco*RI fragments in total rat DNA, cloned in phages Ch4A-RC5, -6, and -8, to gene subdivisions C, D, F, and G (Fig. 3) was quantitatively estimated by thermal dissociation. Fig. 4*A* shows that 50% of the radioactivity from probe C dissociates from the DNA of clones 5, 6, and 8 at about 62–65°C in 0.39 M $Na^+$, whereas the same value for the perfectly homologous clone 4 is about 76°C. Fig. 4 *B*, *C*, and *D* shows that sequences homologous to exon segments D and F and 3′ flanking segment G of pRC4-2 are very closely related, with only a 2°C temperature difference separating the four clones at the point of 50% dissociation. In a control experiment, shown as the broken curves in Fig. 4*B*, where the insulin A-chain probe derived from pSIA12 was dissociated from rat insulin genes I and II (unpublished), the 2 out of 63 base-pairs difference between the two genes and the probe (i.e., 3% mismatch) is reflected in a 3°C temperature difference. Therefore the exon and 3′ flanking sequences common to the four cloned segments differ by an amount equivalent to about 2% sequence mismatch. The experiment, however, cannot distinguish mismatched bases within the sequence from, for example, a perfect homology followed by a deletion of bases at the end, which could shorten the complementary region by several nucleotides and thus slightly lower the melting temperature.

To compare these four closely related clones to the remainder of the cytochrome *c*-specific *Eco*RI fragments, filters identical to those of Fig. 3 were hybridized to fragment A probe at different temperatures in medium containing 0.97 M $Na^+$ and 50% formamide. Fig. 5 shows that as the hybridization stringency increases from 40°C to 50°C several bands are lost. Most notably missing are those at 0.7, 1.8, 2.3, 4.9, and 6.0 kb. As the temperature increases to 55°C, only about 7 of the 15 or so major bands detected at 50°C are still present. Their sizes are 2.2, 2.5, 3.3, 4.3, 4.8, 8.5, and 23 kb and include those present in recombinant phages Ch4A-RC4, -5, -6, and -8. At 60°C no hybridization is detected. The differences in the hybridization pattern at various temperatures are not due to loss of DNA from the filters, because after rehybridization at 40°C all of the original bands are visualized. Thus, the genomic *Eco*RI fragments complementary to the rat cytochrome *c* gene of pRC4-2 fall into three homology classes based upon the temperature of complete dissociation of the *Bam*HI/*Acc* I coding region probe A. These include a group of about 5 fragments dissociating at 40–50°C, a group of about 7–10 fragments dissociating at 50–55°C, and a high-stringency group of about 7 fragments dissociating at 55–60°C. The latter appear to have cytochrome *c*-specific sequences nearly identical to those present in Ch4A-RC4, which is known to contain a complete cytochrome *c* protein coding sequence. It is also of interest that four of the six bands having homology with the 73 nucleotides upstream from the AUG (probe C) are included in the high-stringency class.

**Detection of Families of Cytochrome *c* Gene Sequences in the Genomes of Other Mammals.** In order to examine whether multiple copies of sequences related to cytochrome *c* occur only in rat or are generally present in mammalian genomes, hybridization analyses were performed with total DNA from various mammals. Fig. 6 demonstrates that many bands complementary to rat gene probe A (see Fig. 3) are detected in *Eco*RI- or *Pst* I-digested DNA from cat, baboon, gorilla, and human. Patterns of similar complexity were also found in mouse, dog, bovine, and monkey DNA (data not shown). None of the mammalian DNAs tested were found to have only one or a few homologous fragments. Therefore, families of these sequences appear to be a general feature of mammalian genomes. The hybridization pattern of the cytochrome *c*-specific restriction fragments is distinctive for each organism, with differences found even between closely related primates.

We have recently screened $10^6$ plaques from a Charon 4A–human genomic library, using rat gene probe A (unpublished). Approximately 175 recombinant phages were isolated that had sequences homologous to both yeast iso-1 and rat cytochrome *c* genes. Assuming an average human DNA insert size of 15 kb and a genome size of $3 \times 10^6$ kb, a single-copy sequence would be expected to occur once in every 200,000 phage plaques. The frequency of these sequences is about 35 times higher than that expected for a single-copy gene. These results are consistent with the observation of multiple cytochrome *c* genes or pseudogenes in human DNA.

## DISCUSSION

In this work we show that many mammalian genomes contain families consisting of as many as 20–30 sequences that are

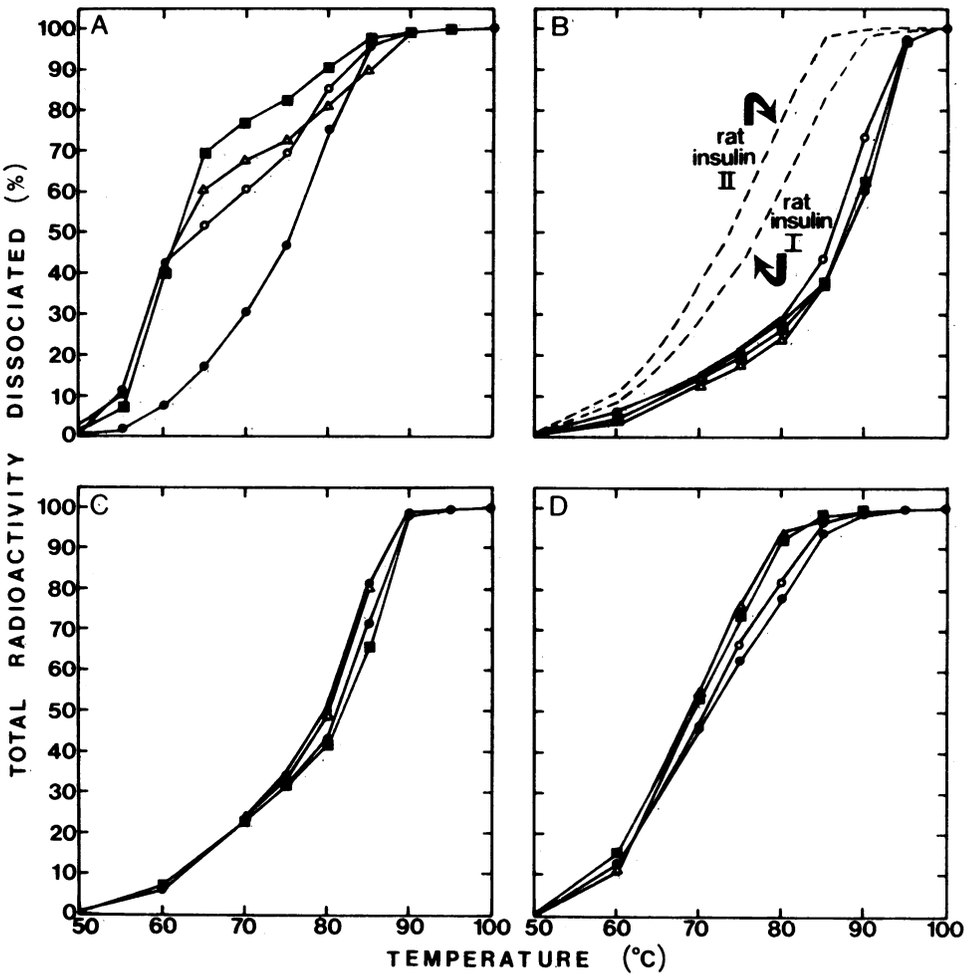


FIG. 4. Thermal dissociation of $^{32}$P-labeled cytochrome *c* gene subfragments from Ch4A-RC recombinant phage DNAs. Curves show the temperature dependence of dissociation in 0.39 M $Na^+$ (0.30 M NaCl/0.03 M $Na_3$ citrate) of various probes from about 20 $\mu$g of phage DNA from Ch4A-RC4 (●), -5 (■), -6 (○), and -8 (△) denatured and immobilized on nitrocellulose discs. *A*, *B*, *C*, and *D* represent the family of curves obtained by using rat cytochrome *c*-specific probes C (5′ noncoding), D (exon I), F (exon II), and G (3′ noncoding), respectively. Initially each filter hybridized approximately 10,000–25,000 cpm. Broken curves in *B* show dissociation of insulin A-chain probe of pSIA12 from cloned rat insulin I and II genes.

closely related to a cytochrome *c* gene of rat that has an intact protein coding region. Essentially all of the rat fragments detected with the complete gene hybridize to probes purified from

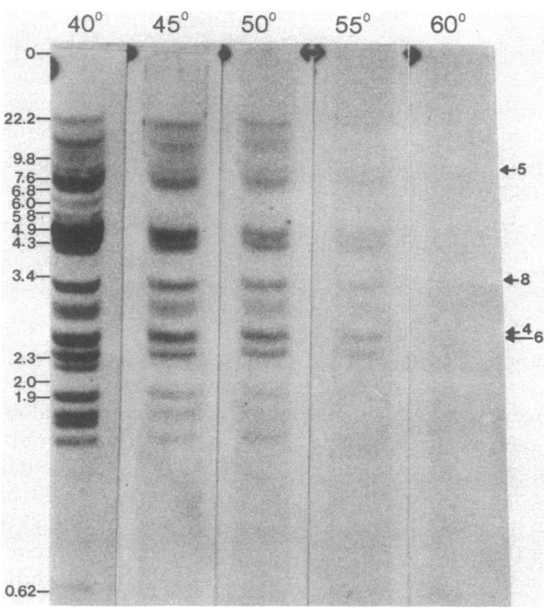


FIG. 5. Temperature dependence of hybridization of the rat cytochrome *c* gene to rat genomic *Eco*RI fragments. Hybridization to identical Southern filters of *Eco*RI-digested rat genomic DNA (30 $\mu$g) using $^{32}$P-labeled probe A (Fig. 3) was carried out in 0.97 M $Na^+$/50% formamide at the indicated Celsius temperatures.

both 5′ and 3′ portions of the coding sequence. The high frequency of these sequences in the genome therefore is not the result of a repeated sequence unrelated to cytochrome *c* or to cleavage within the complementary regions of individual genomic fragments. A probe derived from the 3′ noncoding portion of the rat gene hybridized exclusively to nearly all of the cytochrome *c*-specific bands, suggesting it may have some function related to the biological activity or organization of these sequences in rat. The 3′ flanking DNA, however, does not appear to be extensively conserved between organisms, because in preliminary experiments we detected only modest homology with some human cytochrome *c* gene clones (unpublished). Homology in the 3′ noncoding region of related genes is observed in $\beta$-globin (15, 16) and interferon gene families (17).

Although the rat genomic fragments homologous to the cytochrome *c* gene are closely related, they are not identical. Only two genomic bands are complementary to the intron-specific probe, indicating that the intron sequence is not conserved within the family or, alternatively, that the other fragments are devoid of introns. Extensive sequence conservation in introns is observed between the two $\beta$-globin genes of mouse (15) but not between $\alpha$- and $\beta$-globin genes (16). A subgroup of about six bands was distinguished by ability to hybridize to 73 nucleotides of the 5′ noncoding sequence immediately upstream from the AUG codon (probe C). These, however, did not include fragments corresponding in size to the complementary regions of recombinant phages Ch4A-RC5, -6 and -8, which are shown by thermal dissociation experiments to be otherwise nearly indistinguishable from the cloned gene of Ch4A-RC4. Because in most cases transcriptional and translational control signals are thought to precede the coding sequence, it is intrigu-

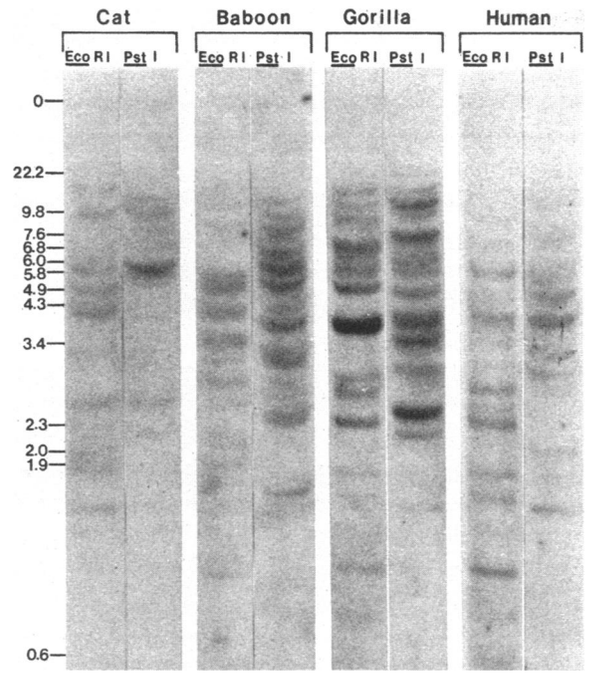


FIG. 6. Hybridization analysis of *Eco*RI- or *Pst* I-digested genomic DNA from various mammals. Total DNA (30 $\mu$g) from cat, baboon, gorilla, or human, digested with *Eco*RI or *Pst* I, was hybridized to rat cytochrome *c* gene probe A (Fig. 3) at 40°C in 0.97 M $Na^+$/50% formamide.

ing that there are at least two types of cytochrome *c*-specific fragments differing in the 5′ noncoding DNA adjacent to the AUG codon.

The rat genomic fragments also differ with respect to the stringency conditions required for their detection with coding region probe A. Three different classes are resolved, including a high-stringency group of about seven bands that includes all but two of the probe C-specific subgroup. The protein coding sequences of at least three members of this class, corresponding to those cloned in Ch4A-RC5, -6, and -8, differ by an amount equivalent to about 2% sequence mismatch from the complete coding region of Ch4A-RC4. If at least some of these represent active genes, they may give rise to polypeptides very similar or identical to the adult form of the protein. Such a group of related genes may serve a regulatory role by having different requirements for activation.

The discovery of families of cytochrome *c* gene-related sequences in mammalian genomes suggests that cytochrome *c* may be the product of more than a single gene. Several other examples of families of related genes are known to exist in eukaryotes (for references see refs. 18 and 19). Pseudogenes, which have incomplete coding sequences and no known product, are also found in gene families (20–24). If a sizable number of mammalian genes have related sequences present in 10–20 copies per cell, this multiplicity can account partly for the puzzling observation that the amount of DNA in higher organisms appears to be about 20 times more than the number of genes. The other part of the excess DNA can be accounted for by introns and certain highly repetitive sequences (25).

With the exception of the mouse testis protein (3), variants in the amino acid sequence of cytochrome *c* from a given mammal have not been described. Many of the cytochrome *c*-specific sequences we observe in rat fall into the two groups that have less homology with the intact gene and may encode genetic variants of the adult protein or may represent pseudogenes. For example, a gene for the mouse testis protein would differ from the coding region of the adult gene by a minimum of 19 out of 312 bases. This corresponds to about 6% mismatch and would account for the difference of 5–15°C separating low- and high-stringency categories. By comparison, a similar mismatch of approximately 18% exists between the rabbit $\beta$-globin gene and linked pseudogene in the sequence complementary to the coding region (24).

Having identified the complexity of cytochrome *c* gene-related sequences in the DNA of mammals, we may now turn our attention to some of the interesting questions raised by these observations. How precisely do these sequences differ from one another? Why are so many DNA sequences present when there appears to be only one adult form of the protein? Are there variants of the gene that are expressed in specific tissues, physiological conditions, or developmental stages? Finally, what is the significance of these sequences for the process of cytochrome *c* evolution?

We thank Jenifer Barthel for excellent technical assistance. The work was supported by Research Grant GM29179 from the National Institutes of Health and Grant DRG-323-F (to R.C.S.) from the Damon Runyon–Walter Winchell Cancer Fund.

1. Margoliash, E. & Schejter, A. (1966) *Adv. Protein Chem.* **21**, 114–286.
2. Sherman, F. & Stewart, J. W. (1971) *Annu. Rev. Genet.* **5**, 257–296.
3. Hennig, B. (1975) *Eur. J. Biochem.* **55**, 167–183.
4. Nunnikhoven, R. (1958) *Biochim. Biophys. Acta* **28**, 108–119.
5. Sherman, F., Taber, H. & Campbell, W. (1965) *J. Mol. Biol.* **13**, 21–39.
6. Smith, M., Leung, D. W., Gillam, S. & Astell, C. R. (1979) *Cell* **16**, 753–761.
7. Montgomery, D. L., Leung, D. W., Smith, M., Shalt, P., Faye, G. & Hall, B. D. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 541–545.
8. Scarpulla, R. C., Agne, K. M. & Wu, R. (1981) *J. Biol. Chem.* **256**, 6480–6486.
9. Blin, N. & Stafford, D. W. (1976) *Nucleic Acids Res.* **3**, 2303–2308.
10. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
11. Rigby, P. W. J., Diekmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113**, 237–251.
12. Wallace, R. B., Shaffer, J., Murphy, R. F., Bonner, J., Hirose, T. & Itakura, K. (1979) *Nucleic Acids Res.* **6**, 3543–3557.
13. Scarpulla, R. C., Agne, K. M., Barthel, J. & Wu, R. (1981) *Recombinant DNA* **1**, 96 (abstr.).
14. Cordell, B., Bell, G., Tischer, E., DeNoto, F. M., Ullrich, A., Pictet, R., Rutter, R. J. & Goodman, H. M. (1979) *Cell* **18**, 533–543.
15. Konkel, D. A., Maizel, J. V. & Leder, P. (1979) *Cell* **18**, 865–873.
16. Nishioka, Y. & Leder, P. (1979) *Cell* **18**, 875–882.
17. Goeddel, D. V., Leung, D. W., Dull, T. J., Gross, M., Lawn, R. M., McCandliss, R., Seeburg, P. H., Ullrich, A., Yelverton, E. & Gray, P. W. (1981) *Nature (London)* **290**, 20–26.
18. Firtel, R. A. (1981) *Cell* **24**, 6–7.
19. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50**, 349–383.
20. Jacq, C., Miller, J. R. & Brownlee, G. G. (1977) *Cell* **12**, 109–120.
21. Nishioka, Y., Leder, A. & Leder, P. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2806–2809.
22. Vanin, E. F., Goldberg, G. I., Tucker, P. W. & Smithies, O. (1980) *Nature (London)* **286**, 222–226.
23. Proudfoot, N. J. & Maniatis, T. (1980) *Cell* **21**, 537–544.
24. Lacy, E. & Maniatis, T. (1980) *Cell* **21**, 545–553.
25. Orgel, L. E. & Crick, F. H. C. (1980) *Nature (London)* **284**, 604–607.