Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 2313–2316, April 1984
Biochemistry

Potential basis for regulation of the coordinately expressed fibrinogen genes: Homology in the 5′ flanking regions

(genomic DNA sequences/signal peptides/transcription regulation/coagulation factors)

DANA M. FOWLKES*, NEAL T. MULLIS*, CLAUDETTE M. COMEAU†, AND GERALD R. CRABTREE†

*Department of Pathology, University of North Carolina, Chapel Hill, NC 27514; and †Laboratory of Pathology, National Cancer Institute, Bethesda, MD 20205

Communicated by K. M. Brinkhous, December 29, 1983

ABSTRACT **The three chains of fibrinogen are encoded by three separate genes whose transcription is coordinately regulated. The breakdown of fibrinogen during the acute-phase reaction leads to a simultaneous increase in α-, β-, and γ-fibrinogen mRNA in the liver. In a search for the basis of this coordinate increase in transcription, we have determined the sequences of the regions surrounding the points of transcriptional initiation of the three rat fibrinogen genes, 1490 nucleotides upstream and 730 nucleotides downstream. Two unique regions of homology have been found. One region consists of 15 nucleotides that have a common 6-nucleotide core lying between −116 and −160; the other is ≈100 nucleotides long and is in the −165 to −472 region. In this region, the β- and γ-fibrinogen genes are ≈65% homologous. α-Fibrinogen has somewhat less homology with both β- and γ-fibrinogen. In addition, the β-fibrinogen gene has 22 nucleotides at position −480 that are homologous to sequences that have been noted to occur in glucocorticosteroid-regulated genes in a similar position. We feel that these areas of conserved sequences play a role in the regulation of the transcription of fibrinogen. The fibrinogen chains are synthesized as precursor peptides, and the amino-terminal portion, the so-called signal peptide, is removed during the translocation of the peptide chain across the endoplasmic reticulum. We have determined those sequences that encode the signal peptides. Homology in the amino acid sequence between the rat and human signal peptides varies between 52% for α-fibrinogen and 66% for β-fibrinogen. This homology implies that there has been strong selective pressure on this portion of these genes.**

A model system to study coordinated gene expression should include a well-defined group of genes displaying similar induction kinetics. The α-, β-, and γ-fibrinogen genes in hepatocytes meet that criterion. Plasma levels of fibrinogen increase along with those of other acute-phase reactant proteins in response to a variety of stimuli, including defibrination (1). After defibrination, hepatic fibrinogen mRNAs accumulate preceding the increase in plasma fibrinogen (2). The increased fibrinogen biosynthesis has been shown to be at least in part regulated by another plasma protein (3, 4) and by glucocorticosteroids (2, 5), and it appears to be the result of increased rates of transcription (unpublished observations).

We have cloned cDNAs for all three fibrinogen transcripts (6) and have used these to isolate genomic DNA clones (7). Each of these clones has been analyzed at least 1490 bases upstream and more than 700 bases downstream from the point of mRNA initiation in all three rat fibrinogen genes, in order to determine whether there are structural features within the fibrinogen gene that may be responsible for coordinating the activity of the three genes.

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.

Since the fibrinogen genes diverged from each other more than 500 million years ago (8, 9) and rates of mutation have been estimated to be about 5×10^{-9} per site per year (10), we reasoned that any conservation of primary sequence outside coding sequences would most likely be of functional importance. Regulation of the expression of the genes' transcriptional units is such a function. We have found two significant areas of homology common to all three genes. These areas occur in sequences upstream from the sites of fibrinogen mRNA transcription initiation. We suggest that these conserved sequences are fibrinogen regulatory units aiding in the transcriptional response characterized by the acute-phase reaction. In addition, we show that the rat β-fibrinogen gene also has a sequence homologous to a highly conserved region noted in certain glucocorticosteroid-responsive genes.

Using primary rat hepatocyte cultures, Nickerson and Fuller showed that the fibrinogen chains are initially translated as large presecretory precursor peptides (11). We have determined the most likely amino acid structure of the NH_2-terminus of the rat fibrinogen peptide chains through examination of the genomic sequence. These rat fibrinogen "signal peptides" are compared to those recently determined for the human fibrinogen chains (12–15).

MATERIALS AND METHODS

Materials. The isolation of recombinant λ Charon 4A phages for all three rat fibrinogen genes has been described (7). Restriction endonucleases and ligase were from New England BioLabs; the Klenow fragment of *Escherichia coli* DNA polymerase was from New England Nuclear and New England BioLabs; unlabeled deoxynucleotide triphosphates, dideoxynucleotide triphosphates, and the M13 phage vectors (mp8 and mp9) were from P-L Biochemicals; [α-^{32}P]dATP (>400 Ci/mmol; 1 Ci = 37 GBq) was from Amersham; and chemicals for reagents were from Sigma.

DNA Sequence Analysis. Sections of the λ Charon 4A phage clones to be analyzed were previously determined by heteroduplex analysis and DNA blotting using the appropriate cDNA clones (7). DNA sequences were determined by the chain-termination method of Sanger *et al.* (16) after restriction endonuclease fragments were subcloned into M13 mp8 or mp9 phage vectors (17). More than 80% of the sequences presented here were determined on both strands. Sequences were analyzed and compared on an IBM system 370 computer using the programs described by Queen and Korn (18) and were analyzed graphically by dot-matrix plots with microcomputer programs.

RESULTS

Fig. 1 depicts the sequences of the regions of the rat genome that include and surround the 5′-initiation points of the α-, β-, and γ-fibrinogen genes. The transcriptional-initiation points were previously described for the α- (unpublished results) and γ-fibrinogen (19) genes. The point of initiation for

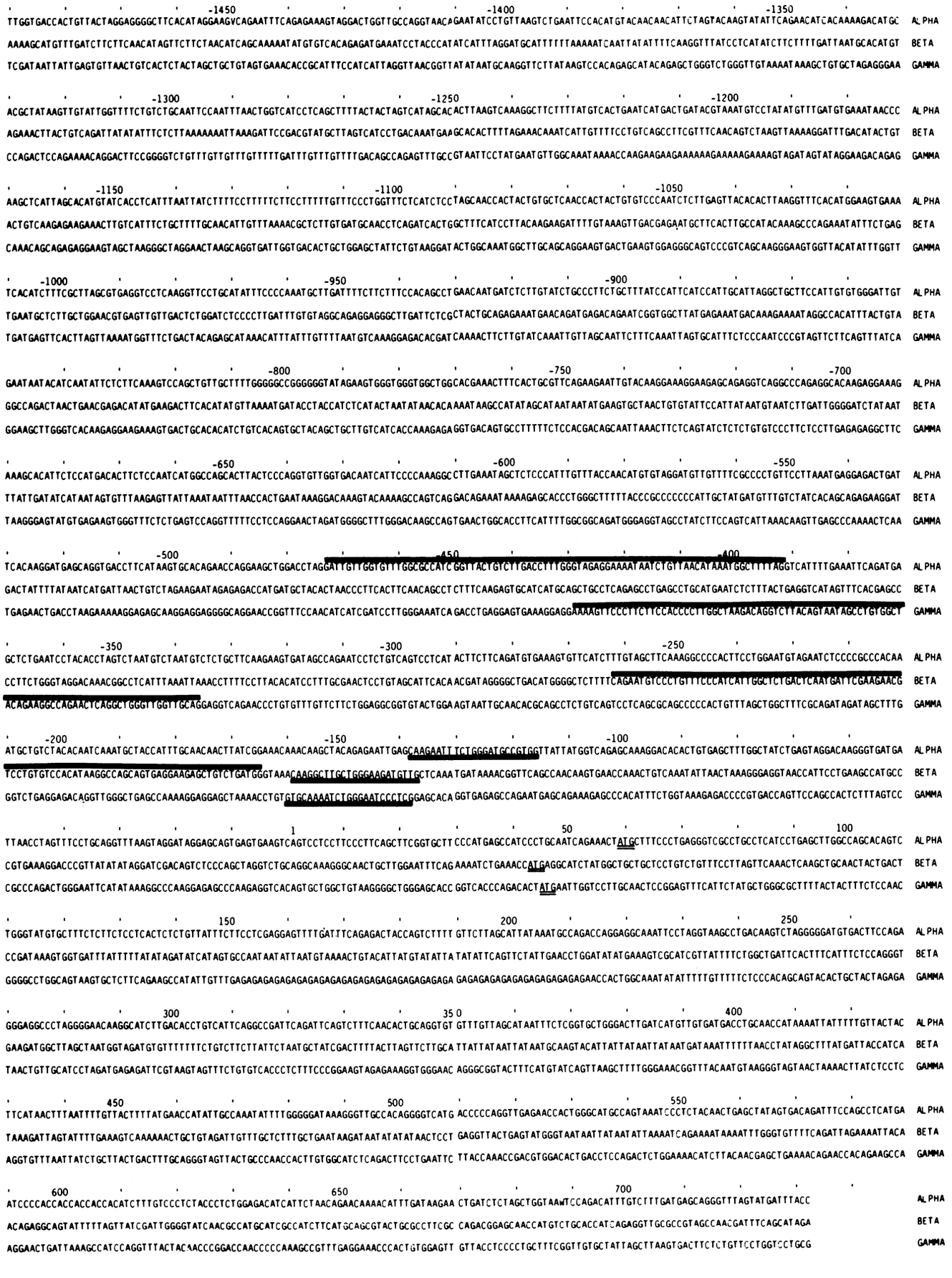

FIG. 1. The genomic DNA sequence flanking the 5′ end of the rat fibrinogen genes. The point of mRNA initiation is +1. The putative initiator methionine codon is underlined twice. One of the two areas of homology between the three genes is underlined. The other area is overlined. The best alignment of the overlined region is shown in Fig. 2.

the β-fibrinogen gene was determined initially by sequence analysis of a large segment of DNA upstream from the third exon. Computer analysis detected a region with the coding potential for the rat fibrinopeptide B (NH_2-terminal portion of the circulating β chain) (20). This is in a reading frame contiguous with the sequence that codes for amino acids typical for signal polypeptides. These sequences are also homologous with the 5′ end of a human β-fibrinogen cDNA clone and a region of the human β-fibrinogen gene reported by Chung *et al.* (14). In addition, we have examined this region for primary transcriptional units in isolated hepatocyte nuclei and found a 20-fold increase in transcription 3′ to the indicated position (unpublished observations).

There is no detectable homology of any significance within the 700 nucleotides downstream from the points of transcriptional initiation. However, there are several areas on the 5′ side that are significant. As is typical of genes transcribed by RNA polymerase II, we note in the β- and γ-fibrinogen genes a sequence within −30 nucleotides that resembles the so-called Goldberg–Hogness ("TATAA") box. Interestingly, a typical TATAA box is not present in the α-fibrinogen gene. We are certain that the assignment of the initiation point is correct for α-fibrinogen, as sequence of the 5′ end of the α mRNA has been determined (unpublished work). Possibly, the TTTAAG at −33 has a similar function. We also note that all three genes have the sequence A-G-G-A in the region from −10 to −30 and this 4-mer occurs twice in the α-fibrinogen gene.

In general there is very little homology among the three rat fibrinogen genes in noncoding and coding areas (unpublished observations). However, there is one major exception. The region −165 to −261 in the β-fibrinogen gene has 60% nucleotide homology with the region −334 to −429 in the γ-fibrinogen gene. This degree of homology is seen with near perfect alignment, requiring only a single, two-nucleotide gap. This homology is shown in Fig. 2 where portions of the α-fibrinogen gene are also aligned to show that a region of α-fibrinogen has similar primary sequence. Using the Queen and Korn sequence analysis programs, we have determined that these homologies between the fibrinogen genes are quite unlikely to have occurred by chance alone ($P < 0.01$). Another area of homology is found at −115, −137, and −138 in α-, β-, and γ-fibrinogen, respectively. This region has a common six-nucleotide core, C-T-G-G-G-A.

An interesting homology to possible glucocorticoid-regulatory sequences is found in the upstream portion of the β-fibrinogen gene. The sequence at −480 to −493 is 77% homologous to that previously noted to occur at about the same relative position as some genes whose transcription is responsive to glucocorticosteroids (21, 22). This high degree of homology implies that this too is unlikely to be a random occurrence. Throughout the upstream portions of all three fibrinogen genes are other areas with lesser degrees of homology to these putative glucocorticosteroid-regulatory sequences. The low degree of homology of these areas makes it possible that these are due to chance ($P > 0.01$).

Certain other sequences that one might expect to find in the 5′-flanking regions are not present. There is no clear area of homology to the so-called "CAAT" box which along with the TATAA box is thought to be a common feature of eukaryotic RNA polymerase II promoters. There are also no areas related to the rodent B1 (23) or B2 (24) repetitive sequences.

```
RAT (TOP) AND HUMAN ALPHA FIBRINOGEN SIGNAL PEPTIDES:
ATG CTT TCC CTG AGG GTC GCC TGC CTC ATC CTG AGC TTG GCC AGA ACA GTC TGG ACT
***     ***      ***          *** ***      *** ***                ***      *** ***
Met Leu Ser Leu Arg Val Ala Cys Leu Ile Leu Ser Leu Ala Ser Thr Val Trp Thr
Met Phe Ser Met Arg Ile Val Cys Leu Val Leu Ser Val Val Gly Thr Ala Trp Thr
ATG TTT TCC ATG AGG ATC GTC TGC CTA GTT CTA AGT GTG GTG GCC ACA GCA TGG ACT


RAT (TOP) AND HUMAN BETA FIBRINOGEN SIGNAL PEPTIDES:
ATG AGG CAT CTA TGG CTG CTG CTC CTG TCT GTT TCC TTA GTT CAA ACT CAA GCT
***     *** ***      *** *** *** ***      ***      *** ***
Met Arg His Leu Trp Leu Leu Leu Leu Ser Val Ser Leu Val Gln Thr Gln Ala
Met Lys His Leu Leu Leu Leu Leu Cys Val Phe Leu Val Lys Ser /// ///
ATG AAA CAT CTA TTA TTG CTA CTA TTG TGT GTT TTT CTA GTT AAG TCC /// ///


RAT (TOP) AND HUMAN GAMMA FIBRINOGEN SIGNAL PEPTIDES:
ATG AAT TGG TCC TTG CAA CTC CGG AGT TTC ATT CTA TGC /// TGG GCG CTT TTA CTA
***     *** *** ***          ***          *** ***                *** *** ***
Met Asn Trp Ser Leu Gln Leu Arg Ser Phe Ile Leu Cys /// Trp Ala Leu Leu Leu
Met Ser Trp Ser Leu His Pro Arg Asn Leu Ile Leu Tyr Phe Tyr Ala Leu Leu Phe
ATG AGT TGG TCC TTG CAC CCC CGG AAT TTA ATT CTC TAC TTC TAT GCT CTT TTA TTT

   CTT TCT CCA ACG GGC CTG GCA
   *** ***     ***          ***
   Leu Ser Pro Thr Gly Leu Ala
   Leu Ser Ser Thr Cys Val Ala
   CTC TCT TCA ACA TGT GTA GCA
```

FIG. 3. Comparison of the rat and human α-, β-, and γ-fibrinogen signal peptides. *** indicates conserved amino acids between the rat and human peptides. /// has been inserted to obtain the optimum alignment of the peptides. The location of the initiator methionines is indicated in Fig. 1.

The first exons are 109, 129, and 126 nucleotides long for α-, β-, and γ-fibrinogen, respectively. The first exons in α- and γ-fibrinogen code for the signal peptides. However, the first exon in β-fibrinogen codes not only for the signal peptide but also for an 8 amino acid portion of the fibrinopeptide B.

Fig. 3 shows the homology of the signal peptides with their apparent human counterpart. The first two methionines found in the human β-genomic sequence by Chung *et al.* (14) are not present in the rat β-genomic DNA. Thus, we have assigned the NH_2-terminal methionines as shown in Fig. 1. The homology between these peptides varies from ≈52% for α-fibrinogen to 66% for β-fibrinogen. Their overall composition maintains a charged amino acid near the NH_2-terminus and a highly nonpolar central domain. The 17th and 18th amino acids found in the rat β-fibrinogen peptide are not present in the human gene. Furthermore, the human γ-fibrinogen gene has an additional amino acid codon for phenylalanine at amino acid position 14. In both species, the γ-fibrinogen signal peptide is considerably larger than those for the other chains. The γ-signal peptides also have at least two internal polar amino acids, which are relatively atypical for signal peptide sequences in general.

DISCUSSION

The results presented above constitute two significant findings. The first regards the coordinate expression of genes. While the rat γ- and α-fibrinogen genes are physically linked on the genome, there are at least 12 kilobases between them.

```
BETA    GAATGTCCCTGTTTCCCATCATTGGCTCTGACTCAATGATTCGAAGAACGTCCTGTGTCC ACATAAGGCCAGCAGTGAGGAAGAGCTGTCTGATGG  -165
        || |||   |||||  |  ||| || |                    | ||| ||   | | |    |          | |  || ||  | |
ALPHA   GATTGTTGGTGTTTGGCGCCATCGGTTAC.............TGTCTTGACCTTTGGGT AGAGGAAAATAATCTGTTAACATAAATGGCTTTTAG  -392

        |   |   |  ||    || | || ||               ||  |  ||| ||| || || |    |         |    ||| |    |
GAMMA   AAAGTTCCCTTCTTCCACCCCTTGGCTAAGAC..AGGTCTTACAGTAATAGCCTGTGGCT ACAGAAGGCCAGAACTCAGGCTGGGTTGGTTGCAGG  -334

        ||  |||||  ||||   | ||||||  |||  |   ||  |  ||   |||||| | ||| |||||||| | | |||  | | ||  ||  ||
BETA    GAATGTCCCTGTTTCCCATCATTGGCTCTGACTCAATGATTCGAAGAACGTCCTGTGTCC ACATAAGGCCAGCAGTGAGGAAGAGCTGTCTGATGG  -165
```

FIG. 2. The best alignment of the largest region of homology between the α-, β-, and γ-fibrinogen genes is shown. This alignment was achieved by use of computer programs.

In addition, the β-fibrinogen gene is separated from the α- and γ-fibrinogen genes by at least 20 kilobases and possibly is unlinked to α- and γ-fibrinogen (7). Thus, it seems likely that each gene is independently regulated. We believe that the two strong stretches of homology seen between α-, β-, and γ-fibrinogen genes are significant, particularly because the time since the original duplication of these genes has been so long that areas of homology, even in coding regions, have almost disappeared. It is possible that these conserved regions have evolved as a part of coordinately induced gene networks as proposed by Britten and Davidson (25, 26). These homologous regions may represent *cis*-acting regulatory elements or "acceptor" sequences as Britten and Davidson have elaborated (27). The linkage of the α-fibrinogen gene to the γ-fibrinogen gene may be such that their transcription is regulated from the same *cis*-element. However, that need not be the case as there is still significant homology between α-fibrinogen and its areas of homology with β- and γ-fibrinogen. The great distance between the β-fibrinogen gene and the others quite likely requires that there be another nearby copy of any requisite *cis*-elements. Such areas of homology have been demonstrated in genes such as the *his3* (28) and *his4* (29) genes of yeast and the heat-shock genes of *Drosophila* (30). In these cases, functional assays demonstrated that the consensus sequence is necessary for physiological induction. It has been established that a major stimulus to fibrinogen induction is a medium-sized protein that is referred to as "hepatocyte-stimulating factor" (3, 4). It is attractive to think that in the fibrinogen genes these homologous regions serve as acceptors for the intracellular signals initiated by hepatocyte-stimulating factor. Furthermore, it is an interesting notion that the sequences noted to be homologous with the putative glucocorticosteroid-regulatory elements function as acceptor sequences for intracellular steroid complexes. One might speculate that once both classes of acceptors have been activated, then synergistic levels of transcription occur.

An alternative hypothesis might be that these genes are coordinately regulated by being in a region of transcriptionally active chromatin. However, even such regions must be able to respond in some manner to *trans*-active regulatory components. This might mean that only one or a few sites in the region of the fibrinogen genes would be necessary for coordinate regulation. Thus, the role of the fibrinogen genes' conserved-flanking sequences in the acute-phase response awaits detailed functional analysis.

Our second finding was the structure of the NH_2-terminal sequences of the rat fibrinogen precursors. One puzzle is the fact that in both the human and the rat the signal peptides of the γ-chains are unusual. Not only are they larger than those of the α- and β-chains, they also have an atypical primary structure consisting of four polar residues preceding the hydrophobic cluster of amino acids. Since the human and the rat internal polar residues have been conserved, we are led to speculate that the γ-fibrinogen signal peptides play a functional role in addition to the traditional *trans*-membrane transport of the polypeptide chain. This, too, needs detailed functional analysis.

This work was supported in part by grants to D.M.F. from the University Research Council of the University of North Carolina and a Grant in Aid from the American Heart Association. N.T.M. was supported by a National Research Service Award, HL07255, from the National Heart, Lung and Blood Institute.

1. Simon, T. L. & Grace, T. G. (1981) *N. Engl. J. Med.* **305,** 443–447.
2. Crabtree, G. R. & Kant, J. A. (1982) *J. Biol. Chem.* **257,** 7277–7279.
3. Ritchie, D. G., Levy, B. A., Adams, M. A. & Fuller, G. M. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 1530–1534.
4. Ritchie, D. G. & Fuller, G. M. (1983) *Ann. N.Y. Acad. Sci.* **408,** 490–500.
5. Grieninger, G., Hertzberg, K. M. & Pindyck, J. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 5506–5510.
6. Crabtree, G. R. & Kant, J. A. (1981) *J. Biol. Chem.* **256,** 9718–9723.
7. Kant, J. A. & Crabtree, G. R. (1983) *J. Biol. Chem.* **258,** 4666–4667.
8. Doolittle, R. F. (1983) *Ann. N.Y. Acad. Sci.* **408,** 13–26.
9. Henschen, A., Lottspeich, F., Kehl, M. & Southan, C. (1983) *Ann. N.Y. Acad. Sci.* **408,** 28–43.
10. Hayashida, H. & Miyata, T. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 2671–2675.
11. Nickerson, J. M. & Fuller, G. M. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 303–307.
12. Kant, J. A., Lord, S. T. & Crabtree, G. R. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3953–3957.
13. Rixon, M. W., Chan, W.-Y., Davie, E. W. & Chung, D. W. (1983) *Biochemistry* **22,** 3237–3244.
14. Chung, D. W., Que, B. G., Rixon, M. W., Mace, M. & Davie, E. W. (1983) *Biochemistry* **22,** 3244–3250.
15. Chung, D. W., Chan, W.-Y. & Davie, E. W. (1983) *Biochemistry* **22,** 3250–3256.
16. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
17. Messing, J., Crea, R. & Seeburg, P. H. (1981) *Nucleic Acids Res.* **9,** 309–321.
18. Queen, C. L. & Korn, L. J. (1980) *Methods Enzymol.* **65,** 595–609.
19. Crabtree, G. R. & Kant, J. A. (1983) *Cell* **31,** 159–166.
20. Doolittle, R. F. & Blomback, B. (1964) *Nature (London)* **202,** 147–152.
21. Davidson, E. H., Jacobs, H. T. & Britten, R. J. (1983) *Nature (London)* **301,** 468–470.
22. Cochet, M., Chang, A. C. Y. & Cohen, S. N. (1982) *Nature (London)* **297,** 335–339.
23. Krayev, A. S., Kramerov, D. A., Skryabin, K. G., Ryskov, A. P., Bayer, A. A. & Georgiev, G. P. (1980) *Nucleic Acids Res.* **8,** 1201–1215.
24. Krayev, A. S., Markusheva, T. V., Kramerov, D. A., Ryskov, A. P., Skryabin, K. G., Bayer, A. A. & Georgiev, G. P. (1982) *Nucleic Acids Res.* **10,** 7461–7475.
25. Britten, R. J. & Davidson, E. H. (1971) *Q. Rev. Biol.* **46,** 111–139.
26. Davidson, E. H. (1982) in *Evolution and Development*, ed. Bonner, J. T. (Springer, Berlin), pp. 65–97.
27. Britten, R. J. & Davidson, E. H. (1969) *Science* **165,** 349–353.
28. Donahue, T. F., Farabaugh, P. J. & Fink, G. R. (1982) *Gene* **18,** 47–59.
29. Donahue, T. F., Daves, R. S., Lucchini, G. & Fink, G. R. (1983) *Cell* **32,** 89–98.
30. Pelham, H. R. B. (1982) *Cell* **30,** 517–528.