MOLECULAR AND CELLULAR BIOLOGY, Feb. 1987, p. 898–904
0270-7306/87/020898-07$02.00/0

Vol. 7, No. 2

Structure of the Rat Platelet Factor 4 Gene: a Marker for Megakaryocyte Differentiation†

TAKEFUMI DOI,[1] SHERYL M. GREENBERG,[1] AND ROBERT D. ROSENBERG[1,2]*

Department of Biology and Whitaker College of Health Sciences, Technology and Management, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139,[1] and Department of Medicine, Harvard University Medical School and Beth Israel Hospital, Boston, Massachusetts 02115[2]

Received 28 July 1986/Accepted 31 October 1986

A rat platelet factor 4 (PF4) cDNA has been isolated by immunoscreening a gλ11 rat megakaryocyte cDNA expression library. Sequence analysis of the rat PF4 cDNA revealed that this megakaryocyte protein is composed of a leader sequence of 29 amino acid residues and a mature protein sequence of 76 amino acid residues. The structure of rat PF4 derived from its cDNA shows a marked homology with the amino acid sequence of human PF4 obtained by classical protein chemistry techniques. This observation is particularly striking with regard to the carboxy-terminal region of rat and human PF4, where 28 of the last 31 C-terminal residues are identical. The rat PF4 gene was obtained from a rat genomic library by using rat PF4 cDNA as a hybridization probe. Sequence analysis showed that the gene is constructed of three exons and two short introns. The transcriptional start site is located 73 base pairs upstream of the translational start codon as judged by S1 nuclease mapping and primer extension. The 5′ noncoding region of the gene also exhibited a sequence homologous to the TATA box at −31, as well as a series of direct and inverted repeat sequences and a cluster of 26 T residues at −155 to −218. This latter domain may be involved in regulating PF4 gene expression during megakaryocytopoiesis.

Megakaryocytes originate from the pluripotential stem cells of the bone marrow (33). The earliest recognizable member of this lineage is the megakaryoblast, which possesses an immature nucleus imbedded in a basophilic cytoplasm having minimal numbers of specific granules (11, 14, 33, 36, 40–42). The conversion of the megakaryoblast into the mature megakaryocyte occurs in concert with a dramatic increase of DNA in the absence of cytoplasmic division (36, 37). After completion of DNA synthesis, a series of more complex events takes place which includes production of specific granules and development of an extensive membrane system (14, 36, 37, 40). The appearance of the α granule represents a distinctive hallmark of this process of cytoplasmic maturation. This unique organelle contains potent coagulation cofactors which accelerate the clotting of blood, adhesive proteins which facilitate the interactions of platelets and other blood cells with the injured blood vessel surface, and growth factors which stimulate the process of wound healing. These components include platelet factor 4 (PF4), the chondroitin sulfate proteoglycan carrier, platelet-derived growth factor, von Willebrand's factor, thrombospondin, etc. (1, 22, 39, 43, 49). Once cytoplasmic maturation has occurred, megakaryocytes fragment into large numbers of small anucleate platelets which possess α granules and circulate within the blood.

The molecular events which control differentiation of stem cells into mature megakaryocytes are poorly understood. We intend to investigate these phenomena by delineating specific plasma proteins and intracellular components which regulate the expression of genes that are characteristic of this cell lineage. The PF4 gene encodes an extremely abundant protein which is found only within the α granule of the megakaryocyte and should serve as an excellent marker for studying the differentiation of this cell type. In the sections below, we describe the cloning and sequencing of rat PF4 cDNA and the rat PF4 gene. The identification of the nucleotide sequence of the cDNA has allowed us to determine the entire primary structure of rat PF4, including the signal peptide. The determination of the nucleotide sequence of the gene has enabled us to ascertain the start site of the primary transcript, define the structures of exons and introns, and suggest areas within the promoter region that may be critical for regulating expression of PF4 during megakaryocytopoiesis.

MATERIALS AND METHODS

Experimentation materials. *Escherichia coli* DNA polymerase, *E. coli* DNA ligase, T4 DNA polymerase, polynucleotide kinase, *Eco*RI methylase, and all restriction enzymes were purchased from New England BioLabs, Inc., Beverly, Mass. Avian myeloblastosis virus reverse transcriptase and RNase H were obtained from Molecular Genetic Resources, Tampa, Fla., and Pharmacia Fine Chemicals, Piscataway, N.J., respectively. Biotinylated affinity-purified anti-rabbit immunoglobulin G (IgG), avidin DH, and biotinylated horseradish peroxidase H were provided as a Vectastain ABC kit from Vector Laboratories. Gold-labeled anti-rabbit IgG was purchased from Janssen Life Science Products, Beerse, Belgium. [α-^{32}P]UTP, [α-^{32}P]dCTP, [α-^{35}S]dCTP, and [γ-^{32}P]ATP were obtained from New England Nuclear Research Products, Boston, Mass. λgt11, the in vitro packaging extracts, and *E. coli* Y1088, Y1089, and Y1090 were provided by Vector Cloning Systems. The M13mp8 and M13mp9 were from New England BioLabs. Sepharose 4B was purchased from Pharmacia Fine Chemicals, and aminobutyl agarose was obtained from Bio-Rad Laboratories, Richmond, Calif. Heparin-Sepharose was prepared from cetylpyridinium chloride-precipitated heparin and cyanogen bromide-activated Sepharose 4B as previously described (31, 44). Other materials included nitrocellulose

* Corresponding author.

† This paper is dedicated to Morio Ikehara on the occasion of his retirement in March 1986 to Osaka University.

filters for immunoscreening and Northern, Southern, and Western blotting (Schleicher & Schuell, Keene, N.H.); nylon membranes for plaque hybridization (New England Nuclear Corp.); *Eco*RI linkers (New England BioLabs); and the universal sequencing primer (Amersham Corp., Arlington Heights, Ill.). The other oligonucleotides that were used for sequencing or primer extension were synthesized with β-cyanoethylphosphoamidites on a DNA synthesizer (model 8600; Biosearch, Inc., San Rafael, Calif.) according to the instructions of the manufacturer.

Production and affinity fractionation of rabbit anti-rat PF4 antisera. PF4 was released from rat platelets by freeze-thaw procedures and collected by precipitation at 4°C with 50% (vol/vol) ammonium sulfate. The harvested proteins were loaded onto a heparin-Sepharose column equilibrated with 0.15 M NaCl in 100 mM Tris hydrochloride (pH 8.3), and bound components were eluted with a linear gradient of 0.5 to 2.0 M NaCl containing 100 mM Tris hydrochloride (pH 8.3). The major protein peak emerged between 0.9 and 1.2 M NaCl and was shown by structural as well as immunochemical techniques to represent purified rat PF4 (S. M. Greenberg, D. J. Kuter, and R. D. Rosenberg, J. Biol. Chem., in press). Rabbits were initially injected intramuscularly with a 1:1 mixture of 100 μg of PF4 and complete Freund adjuvant and were subsequently injected intramuscularly at weekly intervals with the same mixture, except that incomplete Freund adjuvant was used (GIBCO, Laboratories, Grand Island, N.Y.). The antiserum directed against rat PF4 was processed to isolate rabbit anti-rat PF4 IgG by the method of Ey et al. (15). The latter material was then chromatographed on PF4-aminobutyl agarose to obtain affinity-purified rabbit anti-rat PF4 IgG. This procedure was accomplished by filtering 9 ml of rabbit anti-rat PF4 IgG at a concentration of 2.7 mg/ml through 2 ml of a PF4-aminobutyl agarose syringe column equilibrated with 0.15 M NaCl in 20 mM Tris hydrochloride (pH 8.0). After application of the sample, the column was washed with equilibrating buffer until A_{280} approached 0, and the specific rabbit anti-rat PF4 IgG was eluted with 0.15 M NaCl in 1 M acetic acid (pH 2.4). The pooled samples were dialyzed against 0.15 M NaCl in 20 mM Tris hydrochloride (pH 8.0) and were found to contain approximately 0.27 mg of specific rabbit anti-rat PF4 IgG.

Cell culture. Fibroblasts were isolated from the skin of female Fischer rats (Charles River Breeding Laboratories, Inc., Wilmington, Mass.) by the procedure of Hull et al. (23). Smooth-muscle cells, obtained from deendothelialized rat aortae, were grown from medial tissue explants (45). Microvascular endothelial cells were prepared from epididymal fat pads by the technique of Wagner et al. (50) and cloned as described by Marcum and Rosenberg (32).

Construction of a rat megakaryocyte cDNA library. Rat megakaryocytes were isolated from the bone marrow by elutriation and Percoll gradient centrifugation (Greenberg et al., in press). Total RNA was prepared from 5×10^7 megakaryocytes by guanidine hydrochloride extraction, and polyadenylated RNA was obtained by two cylces of chromatography on oligo(dT) cellulose (type 3; Collaborative Research, Inc., Waltham, Mass.). First-strand cDNAs were synthesized with reverse transcriptase by using oligo$(dT)_{12-18}$ as primer, and second-strand cDNAs were generated with *E. coli* DNA polymerase and RNase H by the method of Okayama and Berg (38) as modified by Gubler and Hoffman (19), except that actinomycin D was employed in the final reaction mixture at a concentration of 500 μg/ml. T4 DNA polymerase was used to produce blunt ends, and repaired cDNAs were coupled to synthetic *Eco*RI linkers after protection of internal sites with *Eco*RI methylase. The linker-ligated cDNAs were digested with *Eco*RI and separated from free linkers by Sepharose CL-4B chromatography (Pharmacia). The cDNA was ligated to the *Eco*RI-digested alkaline phosphatase-treated bacteriophage expression vector, λgt11, packaged in vitro, and plated on host *E. coli* Y1088 (19, 24, 52).

Northern blot analysis. Poly(A)-selected RNA was prepared from rat megakaryocytes and other cell types as described above for the construction of the cDNA library. RNA samples were denatured by heating for 15 min at 55°C, chilled, and electrophoresed through a 1.4% agarose gel containing 2.2 M formaldehyde (30). The electrophoresed samples were transferred to a nitrocellulose filter which was then treated at 42°C for 5 h with a prehybridization solution containing 50% formamide, 5× Denhardt solution, 0.9 M NaCl, 50 mM NaH_2PO_4 (pH 7.4), 5 mM EDTA (pH 8.0), 0.1% sodium dodecylsulfate (SDS), and 100 μg of denatured herring DNA per ml. The prehybridization solution was replaced, and the blots were hybridized for 16 h at 42°C with a probe prepared from rat PF4 cDNA and labeled either by nick translation (10^8 cpm/μg) or by in vitro transcription of pSP64 containing rat PF4 cDNA in the opposite orientation to the SP6 promoter (as described below) for the preparation of the riboprobe (4×10^8 cpm/μg). The filters were washed three times with 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) containing 0.1% SDS at room temperature for 20 min with gentle agitation, and twice with 0.1× SSC containing 0.5% SDS at 55°C for 15 min with gentle agitation. The washed filters were dried and exposed to X-Omat AR film (Eastman Kodak Co., Rochester, N.Y.) for 0.5 h.

S1 nuclease protection and primer extension. The S1 nuclease protection experiments were conducted with a single-stranded probe which was prepared as follows. A segment of the PF4 gene located between two *Pst*I sites (Fig. 2) was subcloned into an M13 phage vector in an orientation such that single-stranded DNA possessed the message sense for this gene. An oligonucleotide primer, dCCACAGCTG GCAGCAACAGCAGACCCAGAAGAAGCAGCT, which is homologous to the nucleotide sequence from +117 to +155 near the 5′ end of the PF4 mRNA, was synthesized for use as an S1 nuclease probe primer and was labeled at the 5′ end with [γ-^{32}P]ATP and polynucleotide kinase. The oligonucleotide (3 pmol) was annealed to 0.5 μg of the single-stranded viral DNA in 8.5 μl of 15 mM $MgCl_2$ in 15 mM Tris hydrochloride (pH 8.5) for 60 min at 60°C. Primer extension was initiated by the addition of 7 μl of 0.1 mM deoxynucleoside triphosphates and 3 U of Klenow fragment of DNA polymerase. After 60 min at 25°C, the reaction mixture was phenol extracted as well as ethanol precipitated, and *Bam*HI restriction enzyme (10 U) was added for 0.5 h to cleave the double-stranded vector. The resultant product was denatured in 8 M urea at 100°C for 3 min and then loaded onto a 6% polyacrylamide–8 M urea gel. The 0.34-kilobase (kb) labeled probe was excised and eluted. Poly(A)-selected RNA (2.5 μg) was annealed to 0.1 pmol of probe, treated with S1 nuclease (11,000 U/ml), and analyzed.

The primer extension studies were conducted with the synthetic 39-base oligonucleotide described above. The 5′ ^{32}P-labeled oligonucleotide was added to 2.5 μg of poly(A)-selected rat megakaryocyte RNA, and the mixture was incubated with 9 pmol of the primer in 10 μl of 100 mM KCl and 12 mM $MgCl_2$ in 100 mM Tris hydrochloride (pH 8.2) for 15 min at 90°C and then for 2 h at 60°C. The solution was slowly cooled to 41°C; equal volumes of 4 mM dATP, dCTP, dGTP, and dTTP plus 4 mM dithiothreitol, 2,000 U of

RNasin per ml, and 2,000 U of avian myeloblastosis virus reverse transcriptase per ml were added; and the annealed mixture was incubated at 41°C for 1 h. The reaction was quenched by the addition of EDTA to a final concentration of 10 mM, and the RNA was degraded by the admixture of DNase-free RNase (100 μg/ml) at 40°C for 30 min. The mixture was treated with phenol, the DNA was recovered by ethanol precipitation, and the final product was applied to a 6% polyacrylamide–8 M urea sequencing gel.

Preparation of the riboprobe. A 444-base-pair (bp) rat PF4 cDNA which has *Eco*RI sites at both ends was cut with *Pst*I to produce a 398-bp DNA fragment. This cDNA was ligated to a pSP64 vector which had been digested with *Eco*RI and *Pst*I and then purified by agarose gel electrophoresis. The recombinant plasmids transformed *E. coli* HB101 to ampicillin-resistant species. After recombinant plasmids were isolated plasmid DNA was linearized by *Eco*RI digestion and was used to carry out in vitro transcription. The transcription reaction was conducted in a 20-μl volume containing 1 μg of linear DNA, 0.5 mM ATP, 0.5 mM CTP, 0.5 mM GTP, 15 μM UTP, 50 μCi of [α-^{32}P]UTP, 40 mM Tris hydrochloride (pH 7.5), 6 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl, 10 mM dithiothreitol, 20 U of RNasin, and 15 U of SP6 RNA polymerase. After incubation at 40°C for 1 h, 1 U of RNase-free DNase was added, and this mixture was incubated at 37°C for 15 min. The enzymes were removed by phenol-chloroform and chloroform extractions, and RNA was recovered by sodium acetate-ethanol precipitation.

RESULTS

Isolation and identification of rat PF4 cDNA clones. We have constructed a λgt11 cDNA expression library from rat megakaryocyte poly(A)-selected RNA as outlined in Materials and Methods. The megakaryocyte cDNA library exhibited about 2×10^5 independent recombinant phage with a nonrecombinant background of about 10%, as determined by growth on isopropyl-β-D-thiogalactopyranoside (IPTG)–5 bromo-4 chloro-3 indolyl-β-D-galactopyranoside plates (30). The lengths of 10 randomly selected cDNA clones ranged from 0.2 to 2.1 kb. A rabbit anti-rat PF4 IgG heteroantibody was used to detect bacterial clones possessing fusion proteins of rat PF4 and β-galactosidase. Approximately 10^5 recombinant clones were screened as phage plaques on host *E. coli* Y1090 by the technique of Young and Davis (52). Four candidate plaques were detected by this approach and were taken through several rounds of plaque purification and immunochemical screening. The signals produced by the final screening showed that all plaques were strongly positive (data not shown).

The fusion proteins were further characterized by preparing lysogens of λgt-PF4$_a$, λgt-PF4$_b$, λgt-PF4$_c$, λgt-PF4$_d$, and nonrecombinant λgt11 in host *E. coli* Y1089 as outlined by Young and Davis (53). The phage were induced by shifting the temperature from 30 to 45°C to inactivate the repressor, and the production of β-galactosidase or fusion protein was initiated with IPTG at a final concentration of 5 mM. Cell lysates were precipitated at 4°C with 33% (vol/vol) saturated ammonium sulfate and then examined by SDS-polyacrylamide gel electrophoresis as well as by Western blot analyses with affinity-fractionated anti-rat PF4 IgG heteroantibody and then gold-labeled anti-rabbit IgG as outlined in the instructions of the manufacturer.

The synthesis of β-galactosidase and the fusion peptides were clearly dependent on the presence of IPTG (Fig. 1A, lanes 1 through 5). The various fusion proteins were about

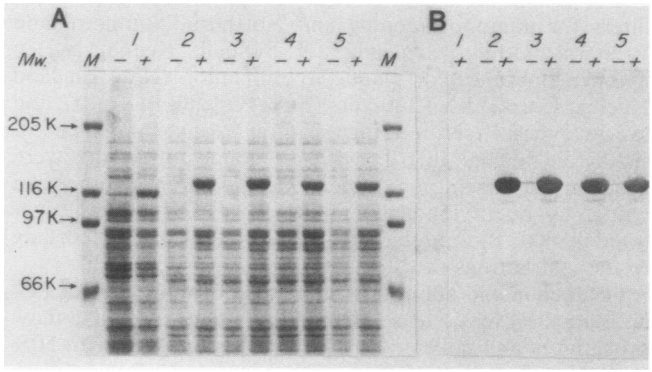

FIG. 1. Identification of PF4 cDNA clones and characterization of fusion peptides. (A) SDS-polyacrylamide gel electrophoresis with matrices of 7.5% acrylamide and 0.2% bisacrylamide for staining of bands with Coomassie brilliant blue. (B) Western blot analysis with affinity-purified rabbit anti-rat PF4 IgG. Molecular weight designations are shown on the left in thousands (205K, 205,000-molecular-weight protein). Lanes 1 through 5 show λgt11 lysogens: Y1089 (λgt11) (no insert), Y1089(λgt11-PF4$_a$), Y1089 (λgt11-PF4$_b$), Y1089 (λgt11-PF4$_c$), and Y1089(λgt11-PF4$_d$), respectively. + and − indicate incubation with and without IPTG, respectively.

9,000 to 11,000 daltons larger than β-galactosidase. This size would correspond to an insert of about 330 to 440 bp. The immunoblotting of the same cell lysates with affinity-purified rabbit anti-rat IgG heteroantibody showed that PF4 epitopes existed on all of the fusion proteins (Fig. 1B, lanes 2 through 5).

λgt11-PF4$_a$ and λgt11-PF4$_b$, which contained the largest inserts, were plaque purified, amplified to yield high-titer plate stocks, and used for large-scale preparation of phage (30). The isolated DNA was subjected to *Eco*RI digestion to obtain insert fragments, subcloned into M13mp8, and sequenced by the dideoxy chain termination method of Biggin et al. (3). The overall structure of the 444-bp cDNA is depicted in Fig. 2, and its nucleotide sequence is provided in Fig. 4. The cDNA exhibited a single open reading frame from an ATG translation initiation codon to a TAG stop codon, and it encoded a leader sequence of 29 amino acid residues as well as a mature polypeptide sequence of 76 amino acid residues. The structure of the mature polypeptide region deduced from the cDNA is strikingly homologous to the complete amino acid sequence of human PF4 and the N-terminal sequence of rat PF4 as outlined below (Greenberg et al., in press). The cDNA sequence also revealed that both recombinant phages possessed inserts that exhibit the same frame as the β-galactosidase gene, a finding which is consistent with the expression of a fusion protein.

Northern blot analyses. For Northern blot analyses, poly(A)-selected RNA was prepared from rat megakaryocytes, rat endothelial cells, rat smooth-muscle cells, and rat fibroblasts as outlined in Materials and Methods. The samples were electrophoresed on denaturing gels, transferred to nitrocellulose, and probed with rat PF4 cDNA (Fig. 3). It is readily apparent that the probe hybridized mainly to an RNA component of about 0.7 kb present within rat megakaryocytes but that traces of larger RNA species were also detected by the above-mentioned technique. The origins of these larger species will be considered in a subsequent communication. The other cell types examined, including bone marrow cells depleted of rat megakaryocytes (data not shown), contained essentially no RNA species capable of hybridizing to the PF4 cDNA probe. From this limited

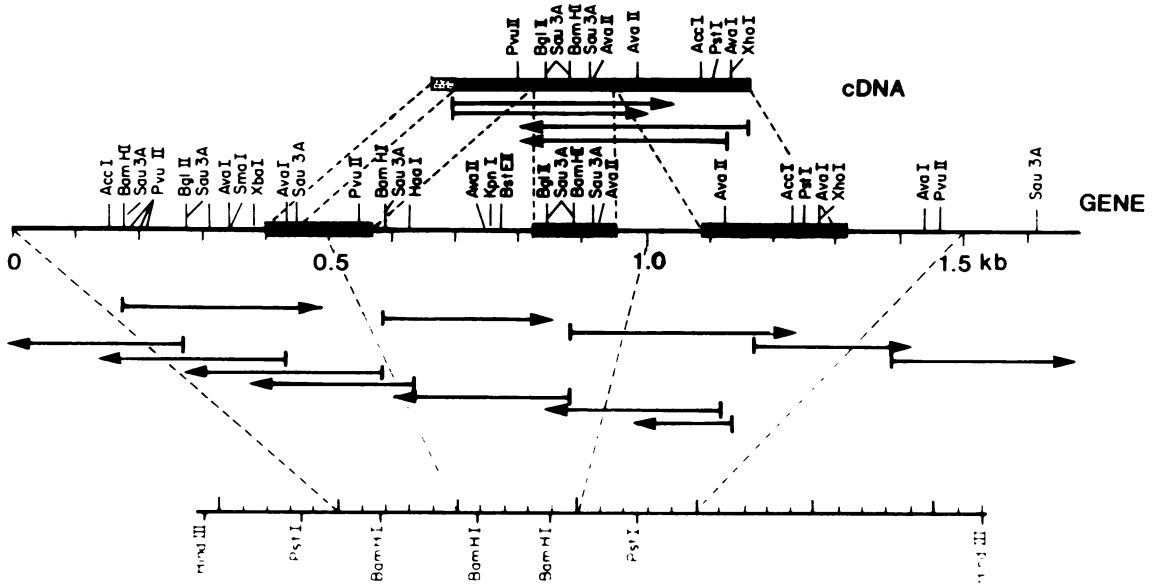

FIG. 2. Overall structure organization of rat PF4 cDNA and the rat PF4 gene. ■, Exons. Arrows show the extent and direction of nucleotide sequencing.

analysis it appears that megakaryocytes are the only cell type capable of transcribing the PF4 gene, a finding which is in accord with results obtained by other investigators with heteroantibodies directed against this platelet protein.

Isolation and structure of the PF4 gene. The genomic library employed was prepared by J. W Tamkun and associates (48) from a partial *Sau*3AI digest of Fischer rat DNA cloned into the EMBL3B vector. After screening 3×10^5 plaques with the PF4 cDNA probe, one positive recombinant clone was identified and was subsequently subjected to additional rounds of plaque purification. The cloned phage DNA containing the PF4 gene was isolated and cut with *Hin*dIII. Southern blot analysis of the digested products revealed that a 3.2-kb DNA fragment had hybridized with the PF4 cDNA probe. This restriction fragment was isolated from the agarose gel and then treated with either *Bam*HI or *Pst*I. Southern blot analysis of the *Bam*HI-digested products showed that 1.8-, 0.4-, and 0.3-kb fragments had hybridized with the PF4 cDNA probe. A similar examination of the *Pst*I-digested products detected two closely spaced bands of about 1.4 kb. These two sets of fragments were subcloned into M13mp8 or M13mp9 or both, and their sequences were determined. The overall structure of the PF4 gene is depicted in Fig. 2, and its nucleotide sequence is shown in Fig. 4.

The 5′ end of the PF4 gene was mapped by S1 nuclease and primer extension experiments (Fig. 5, inset). In the S1 nuclease mapping experiments, we used a 0.34-kb labeled DNA fragment which was prepared with a synthetic 39-base oligonucleotide whose 5′ end was located at position +155 as described in Materials and Methods. This probe contained 82 bases which encode a portion of the leader peptide as well as the translation initiation codon and about 260 bases of the 5′-flanking region of the PF4 gene. For the cDNA primer extension experiments, the same oligonucleotide was used for S1 nuclease mapping. The results of the S1 nuclease mapping and cDNA primer extension experiments with rat megakaryocyte poly(A)-selected RNA are depicted in Fig. 5. The rat megakaryocyte RNA protected a single-stranded DNA fragment of 155 bases, and a cDNA of 155 bases was obtained when the same RNA was used as template. Thus, the 5′ region of the rat megakaryocyte poly(A)-tailed RNA was colinear with the genomic sequence, and its cap site was 73 bp upstream of the translation initiation codon (see asterisk on the sequence of the PF4 gene, Fig. 5).

To show that the sequence of the rat PF4 gene obtained from cloned recombinant DNA is representative of genomic

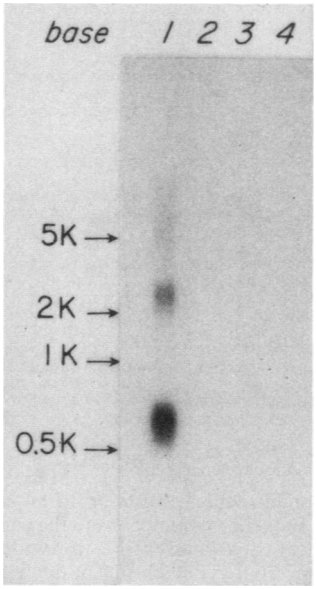

FIG. 3. Northern blot analyses of poly(A)-selected RNA from different rat cell types. Lanes: 1, megakaryocyte; 2, fibroblasts; 3, smooth-muscle cells; and 4, endothelial cells. Arrows denote (in descending order) the molecular weight standards (shown on left in thousands [5K, 5,000 bases]); for eucaryotic 28SrRNA, eucaryotic 18SrRNA, 1-kb DNA fragment (Bethesda Research Laboratories, Inc., Gaithersburg, Md.); and 0.5-kb λ *Hin*dIII DNA fragment.

DNA, we performed Southern blot analyses with *Hin*dIII and *Pst*I restriction enzymes on cellular DNA isolated from rat endothelial cells and recombinant phage DNA (data not shown; Fig. 2). The probe we used was derived from the riboprobe containing the 398 bases of the rat PF4 cDNA sequence. Both cloned and genomic DNAs gave an identical set of restriction fragments whose lengths could be predicted from the sequence of the rat PF4 gene. These results suggest that there is only one PF4 gene per rat haploid genome and that no significant rearrangements have occurred during cloning and phage propagation.

```
-400  GTGTTTGGGGTAGGGCAACCGGAAGTCGGGAAGGCAACAAATTGGTACTG   -351
-350  AAGGTGCATGTTCTGTAAACCGCATGGGGATAGCAGAAAATTCTCTGCCA   -301

-300  CATACAGCATACCTTCTGCGAAAATTCCAACTGTTTCTACCTCTGTAGAC   -251

-250  TGTTCACATAAATTCACATTGGGGACGTGGATCCTGCTGACAGCTGCTGA   -201
                                        ----------> <----
-200  CAGCTGGCCTCAGCTGCTTCTTTTTTTTTTTTTTTTTTTTTTTTTCCTG   -151
      ------>  <------
-150  TTCTATGTTGCTTTAATCTTGGCTGGCCAGATCTCAAGTACTGTTCCACA   -101
-100  AGTGTCATTGCTTCTGTGGATCACTTCCTCATCCCCTATCCCGGGTTTCC    -51

 -50  GGACTGGGCTGGCAGTGAAGATAAAACGTGTCTAGAAAGTCACAGGAGCC     -1
       *
   1  ACTGTCTGGCACTTAGAGCCCCAGACCCCAGTTTCCCCGAGGTAGCCCTT     50

                                  Met Ser Ala Ala Ala Val Phe Arg Gly
  51  GGATCTAGCAGCACCCTCTTGAGATGAGTGCCGCTGCGGTGTTTCGAGGC    100

      Leu Arg Pro Ser Pro Glu Leu Leu Leu Leu Gly Leu Leu Leu Leu Pro Ala
 101  CTCCGGCCCAGCCCTGAGCTGCTTCTTCTGGGTCTGCTGTTGCTGCCAGC    150

        Val Val Ala Val Thr Arg
 151  TGTGGTTGCTGTCACCAGGGGTGAGATTGGGTGAAAGGATCCTTGGGGCA    200
                                  ↳INTRON
 201  GCAAGACAGTTCCGAGGCCTCAGGCGCGGGGAGGTGAACTCTGACGACAC    250
 251  CAGTCCTACGGTATAGGCTTGCCAACCGCACCGCACTACCCATCTGCCTC    300
 301  TGGGGGCGGGGGGTTGATGATGCTCTAGCCAAGGGTTTGGAAAGGGTCCA    350
 351  GAGGTACCACTGGCCGATGAAAGGGTCACCACTATCTCTTCTACTCCCCT    400

                                  Ala Ser Pro Glu Glu Ser Asp Gly Asp
 401  CTTCCTCTCCCCCTTTGCCCACAGCTAGTCCTGAAGAAAGCGACGGAGAT    450
                                  ↳EXON
      Leu Ser Cys Val Cys Val Lys Thr Ser Ser Ser Arg Ile His Leu Lys Arg
 451  CTTAGCTGTGTGTGTGTGAAGACCAGTTCTTCCAGGATCCATCTCAAACG    500

        Ile Thr Ser Leu Glu Val Ile Leu Lys Ala Gly Pro His Cys Ala Val Pro
 501  CATCACCAGCCTGGAGGTGATCAAAGCAGGACCCCACTGTGCGGTTCCCC    550

      Gln Leu Ile
 551  AGCTCATGTGAGTCTTGCCCACACCCCACCCCCAGGCCTGTCCCTCTCAT    600
             ↳INTRON
 601  CCTCTGCCCAATCCTTCCAGTCTCTTTCCTCCTTCCAACCAAGTTTCTGA    650

                                                  Ala Thr
 651  AAAAGTAATTTCTCCACCCCACCCCCCCATGTCCCTCCCGACAGAGCCAC    700
                                                  ↳EXON
        Leu Lys Asn Gly Ser Lys Ile Cys Leu Asp Arg Gln Val Pro Leu Tyr
 701  GCTGAAGAATGGGAGCAAAATTTGCCTGGACCGGCAAGTACCTCTGTATA    750

      Lys Lys Ile Ile Leu Lys Lys Leu Leu Glu Ser
 751  AGAAAATAATCAAGAAACTCCTGGAGAGTTAGGTACGAGCTGCCTAAATG    800

 801  TGCAACCTGTTACATGGGATTCCTGGAGTCTTGTCTACTTTTAATGTAAC    850
 851  TGCAGTCTTCTGATGTTTTATTATCCTCGAGATTTAAATAAATGCATTGA    900

 901  ACCAAGTTGTCATAAAGCCAAGCTATTTTTTTTTGTAACGAAGCCAAAAG    950
 951  TCGTGGTGACAGAACACGAGTACGAAAACATCTCCAAGACTTTAAGCAAA   1000
1001  CATTATTTTGAAGGGAGGTTTGACCTAAGAACTGAATTTGCTCGGGATTA   1050
1051  TGGCTTACGTGGCAGCTGATGAATACTTGAATATTTATTTAATTCCCTTG   1100
1101  TCTCACAGTTACACTTTAGGTTTGCATTGTAATTAGCTTTGTAAAAAAAA   1150
1151  CATTAATGGCACTTAACATAGCCTTTGAAAGAAAGTACCTAAGTCAAACT   1200
1201  TTGATTCGATTTTCTTAAGATCGTCATAAAAATAAAAGTGATAGGTTAGA   1250
1251  ATGAATACACAGTGTTTATATGTTG
```

FIG. 4. Nucleotide sequence of the rat PF4 gene. *, Transcription start site; ~~, sequence homologous to the canonical TATA box and the potential hexanucleotide polyadenylation signal (AATAAA); =, the 26-thymine cluster in the 5′ noncoding region; – – –, inverted perfect 6-bp repeat sequences and inverted imperfect 12-bp repeat sequences in the 5′ noncoding region; The boxed sequence includes four types of 10-bp direct-repeat sequences in the 5′ noncoding region (start point shifts 1 bp to the 3′ side). →, Boundaries between introns and exons. The coding region is delineated with the appropriate amino acid residues derived from the nucleotide sequence. ●, Beginning of the structural domain of PF4. The translation initiation and termination codons are depicted with double boxes.

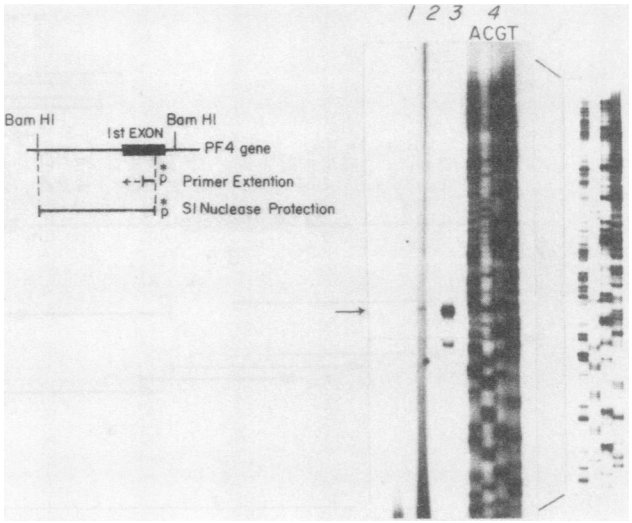

FIG. 5. S1 nuclease protection and primer extension mapping of the transcription start site of the PF4 gene. Poly(A)-selected rat megakaryocyte RNA was hybridized to the probes (inset). S1 nuclease analysis and cDNA synthesis were performed as described in Materials and Methods. The cDNA product and the S1 nuclease-protected fragment were electrophoresed on 6% polyacrylamide–8 M urea gels (lanes 2 and 3). The same unlabeled probes were used for the dideoxy chain termination reaction to indicate the sizes of the various products (lane 4). Lane 1 represents radiolabeled primer. The arrow indicates (from left to right) the positions of the S1 nuclease-protected fragment and the synthesized cDNA. The inset depicts the overall strategies of the S1 nuclease protection and primer extension studies.

DISCUSSION

PF4 is synthesized by bone marrow megakaryocytes and stored within their alpha granules (6, 25). The circulating blood platelet which arises from the megakaryocyte contains large amounts of this protein. PF4 is composed of four equivalent polypeptide chains of about 75 amino acid residues which are held together by noncovalent bonds. This component is able to neutralize the anticoagulant activity of heparan sulfate present on endothelial cells (1, 7); induce chemotaxis of neutrophils, monocytes, and fibroblasts (10, 47); inhibit the proteolytic activity of collagenase, as well as potentiate the action of elastase (21, 29); and may also play an important immunoregulatory function (27). Hence, the release of PF4 by the circulating blood platelet is thought to enhance formation of a blood clot at the site of injury and then induce many of the cellular reactions which eventually lead to healing of the wound.

We have isolated a full-length rat PF4 cDNA by screening a λgt11 rat megakaryocyte library with a specific rabbit anti-rat PF4 IgG heteroantibody. The primary structure of rat PF4 derived from the nucleotide sequence of the cDNA shows that this unique megakaryocyte protein is composed of a leader peptide of 29 amino acid residues and a mature protein domain of 76 amino acid residues. The leader peptide is extremely hydrophobic, with two clusters of a Leu-Leu-Leu-Leu sequence, and it may play an important role in targeting PF4 to the alpha granule within the megakaryocyte. This type of specificity has been observed for proteins that are exported to mitochondria (26). However, a comparison of the PF4 leader peptide with other such peptides did not reveal any particularly unique aspects of the structure which might allow it to perform this function. Examination of the sequences of the rat PF4 cDNA and human PF4 protein

showed that the two mature protein domains are quite homologous, with about a 75% identity of amino acid residues. The similarity of the two species of PF4 is most striking with regard to the carboxy-terminal area, where 28 of the last 31 residues are identical, and the observed differences in primary structure can be explained by only three single-base substitutions. This observation is of particular interest since the domain in question is involved in the binding of the polypeptide to heparin or heparan sulfate (9). The preservation of this region within both rat and human PF4 genes suggests that this function is of considerable biologic importance.

We have also isolated the rat PF4 gene in the screening of an EMBL3B genomic library with a probe derived from rat PF4 cDNA. The nucleotide sequence of the rat PF4 gene when compared with that of the cDNA, is seen to be composed of three exons of about 500 bp and two introns of about 390 bp, which have 5′ and 3′ untranslated regions of about 73 and 120 bp, respectively. The transcription-start site is reasonably homologous to a consensus transcription start site, PyPyCAPyPyPyPy (8), and is located 73 bp upstream of the AUG translation initiation codon as judged by S1 nuclease mapping and primer extension studies. The presence of a GATAAAA sequence is noted 31 bp upstream from the cap site, which is moderately homologous to the canonical TATA box often seen at this position in eucaryotic promoters (8). There are, however, no obvious candidates for the canonical CAAT box which is often present further upstream. This region of the gene, which is about 150 to 220 bp upstream of the cap site, has two additional features of interest. We noted a series of direct and inverted hexanucleotide and decanucleotide repeats and an imperfect inverted dodecanucleotide repeat at −185 to −218 as well as a cluster of 26 thymines at position −155 to −180. These types of sequences may be involved in the binding of *trans*-acting substances needed for the transcription of the PF4 gene (12, 13, 16, 18, 46).

The PF4 gene contains three exons and two introns. All of the introns begin with the dinucleotide G-T and end with the dinucleotide A-G (5). The splicing junctures of the introns and exons of the gene are also in reasonable accord with more detailed, consensus sequences of the donor G-T-A/G-A-G-T and the acceptor T/C-N-C/T-A-G sites (34). The CTGAC sequence which is thought to be a potential signal for splicing is found in the first intron at nucleotide +241. The second intron has several pentanucleotide sequences which are homologous to the CTGAC sequence at positions +588, +604, and +647. The pattern of interruption of the rat PF4 gene by introns is consistent with the proposal that functional domains of proteins are often encoded by individual exons (4, 17). Thus, exon 1 encodes the signal peptide as well as three N-terminal amino acid residues which are absent in human PF4. The leader sequence often plays an important role in the secretory process and is frequently isolated within the exons of genes coding for these types of proteins (28, 35). Exon 2 encodes 35 amino acid residues of rat PF4 and shows 15 differences in amino acid residues when compared with the protein sequence of human PF4 (11, 20, 51). The biologic role of this region remains enigmatic. Exon 3 encodes the remaining 28 amino acid residues of the carboxy-terminal region of rat PF4, which are believed to be important in the binding of this polypeptide to glycosaminoglycans and are virtually indentical to those found in human PF4 (as described above).

The PF4 gene also contains an AATAAA polyadenylation signal within its 3′ region at position +887, as well as a CAPyTC sequence 2 bp downstream which may play a role in the processing of the 3′ end of the primary transcript (2). We also noted the presence of a second AATAAA polyadenylation signal at position +1231. However, it appears likely that the first AATAAA sequence was used since polyadenylated PF4 mRNA exhibited a length of about 700 bp, while the length of the transcribed portion of the gene from the cap site to the first polyadenylation signal site was about 870 bp.

We have recently isolated a plasma protein with a molecular weight of 15,000, termed the megakaryocyte stimulatory factor, which is able to induce isolated, cytoplasmically immature megakaryocytes to synthesize PF4 when it is introduced at levels of less than 0.1 pM (Greenberg et al., in press). Under in vivo conditions, the plasma concentrations of megakaryocyte stimulatory factor are increased as platelet levels are reduced. Furthermore, preliminary experiments indicate that megakaryocyte stimulatory factor is capable of dramatically enhancing the amounts of PF4 mRNA via a transcriptional event in the earliest recognizable form of megakaryocytes and may be intimately involved in the regulation of megakaryocyte differentiation. Therefore, we believe that a careful investigation of the molecular events which are induced by the above-mentioned protein and which lead to PF4 gene expression will constitute a fruitful approach to defining the pathway by which bone marrow progenitors are committed to the megakaryocyte lineage. We are currently pursuing this line of inquiry.

ACKNOWLEDGMENTS

This work was supported by Public Health Service grant HL-33124 from the Heart, Lung, and Blood Institute and by funds from the National Foundation for Cancer Research.

We thank Monty Krieger and Robert Jackman for helpful discussions and James Marcum and Houry Youssoufian for providing rat endothelial cells, rat fibroblasts, and rat smooth-muscle cells.

ADDENDUM IN PROOF

We have recently determined the nucleotide sequence of the rat PF4 gene from about −1500 to about −400, and we have observed a second group of 26 thymines at position −954 to −979 as well as a block of 14 thymine-adenine dimers at position −980 to −1007. These thymine clusters, as well as the one outlined in this paper, may play an important role in PF4 gene expression.

LITERATURE CITED

1. **Barber, A. J., R. Käser-Glanzmann, M. Jakábová, and E. F. Luscher.** 1972. Characterization of a chondroitin 4-sulfate proteoglycan carrier for heparin neutralizing activity (platelet factor 4) released from human blood platelets. Biochim. Biophys. Acta **286**:312–329.
2. **Berget, S. M.** 1984. Are U4 small nuclear ribonucleoproteins involved in polyadenylation? Nature (London) **309**:179–182.
3. **Biggin, M. D., T. J. Gibson, and G. F. Hong.** 1983. Buffer gradient gels and 35S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA **80**:3963–3965.
4. **Blake, C. C. F.** 1978. Do genes-in-pieces imply proteins-in-pieces? Nature (London) **273**:267.
5. **Breathnach, R., C. Benoist, K. O'Hare, F. Gannon, and P. Chambon.** 1978. Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. Proc. Natl. Acad. Sci. USA **75**:4853–4857.
6. **Broekman, M. J., R. I. Handin, and P. Cohen.** 1975. Distribution of fibrinogen, and platelet factor 4 and XIII in subcellular fractions of human platelets. Br. J. Haematol. **31**:51–55.
7. **Busch, C., J. Dawes, D. S. Pepper, and A. Wasterson.** 1980. Binding of platelet factor 4 to cultured human umbilical vein endothelial cells. Thromb. Res. **19**:129–137.

8. **Corden, J., B. Wasylyk, A. Buchwalder, P. Sassone-Corsi, C. Kedinger, and P. Chambon.** 1980. Promoter sequences of eukaryotic promoter-coding genes. Science **209**:1406–1414.
9. **Deuel, T. F., P. S. Keim, M. Farmer, and R. L. Heinrikson.** 1977. Amino acid sequence of human platelet factor 4. Proc. Natl. Acad. Sci. USA **74**:2256-2258.
10. **Deuel, T. F., R. M. Senior, D. Chang, G. L. Griffin, R. L. Heinrikson, and E. T. Kaiser.** 1981. Platelet factor 4 is chemotactic for neutrophils and moncytes. Proc. Natl. Acad. Sci. USA **78**:4584-4587.
11. **DuPont, H., M.-A. DuPont, H. Bricaud, and M.-R. Boisseau.** 1983. Megakaryocyte separation in homogeneous classes by unit gravity sedimentation: physicochemical, ultrastructural and cytophotometric characterizations. Biol. Cell **49**:137–144.
12. **Dynan, W. S., and R. Tjian.** 1983. Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. Cell **32**:669–680.
13. **Dynan, W. S., and R. Tjian.** 1983. The promotor-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promotor. Cell **35**:79–87.
14. **Ebbe, S., and F. Stohlman, Jr.** 1965. Megakaryocytepoiesis in the rat. Blood **26**:20–35.
15. **Ey, P. L., S. J. Prowse, and C. R. Jenkin.** 1978. Isolation of pure IgG1, IgG2a and IgG2b immunoglobulins from mouse serum using protein A-sepharose. Immunochemistry **15**:429–436.
16. **Gidoni, D., W. S. Dynan, and R. Tjian.** 1984. Multiple specific contacts between a mammalian transcription factor and its cognate promoters. Nature (London) **312**:409–413.
17. **Gō, M.** 1983. Modular structural units, exons, and function in chicken lysozyme. Proc. Natl. Acad. Sci. USA **80**:1964–1968.
18. **Green, M. R., R. Treisman, and T. Maniatis.** 1983. Transcriptional activation of cloned human beta-globin genes by viral immediate-early gene products. Cell **35**:137–148.
19. **Gubler, U., and B. J. Hoffman.** 1983. A simple and very efficient method for generating cDNA libraries. Gene **25**:263–269.
20. **Hermodson, M., G. Schmer, and K. Kurachi.** 1977. Isolation, crystallization, and primary amino acid sequence of human platelet factor 4. J. Biol. Chem. **252**:6276–6279.
21. **Hiti-Harper, J., H. Wohl, and E. Harper.** 1978. Platelet factor 4: an inhibitor of collagenase. Science **199**:991–992.
22. **Huang, S. S., J. S. Huang, and T. F. Deuel.** 1982. Proteoglycan carrier of human platelet factor 4. Isolation and characterization. J. Biol. Chem. **257**:11546–11550.
23. **Hull, B. E., S. E. Sher, S. Rosen, D. Church, and E. Bell.** 1983. Structural integration of skin equivalents grafted to Lewis and Sprague-Dawley rats. J. Invest. Dermatol. **81**:429–436.
24. **Huynh, T. V., R. A. Young, and R. W. Davis.** 1984. Construction and screening cDNA libraries in λgt10 and λgt11, p. 49–78. *In* D. M. Glover (ed.), DNA cloning, vol. 1. A practical approach. IRL Press, Oxford.
25. **Kaplan, K. L., M. J. Broekman, A. Chernoff, G. R. Leschnick, and M. Drillings.** 1979. Platelet alpha-granule proteins: studies on release and subcellular localization. Blood **53**:604–618.
26. **Kaput, J., S. Goltz, and G. Blobel.** 1982. Nucleotide sequence of the yeast nuclear gene for cytochrome *c* peroxidase precursor. Functional implications of the pre sequence for protein transport into mitochondria. J. Biol. Chem. **257**:15054–15058.
27. **Katz, I. R., G. J. Thorbecke, M. K. Bell, J.-Z. Yin, D. Clarke, and M. B. Zucker.** 1986. Protease-induced immunoregulatory activity of platelet factor 4. Proc. Natl. Acad. Sci. USA **83**: 3491–3495.
28. **Larhammar, D., J. J. Hyldig-Nielsen, B. Servenius, G. Andersson, L. Bask, and P. A. Peterson.** 1983. Exon-intron organization and complete nucleotide sequence of a human major histocompatibility antigen DC beta gene. Proc. Natl. Acad. Sci. USA **80**:7313–7317.
29. **Lonky, S. A., J. Marsh, and H. Wohl.** 1978. Stimulation of human granulocyte elastase by platelet factor 4 and heparin. Biochem. Biophys. Res. Commun. **85**:1113–1118.
30. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
31. **March, S. C., I. Parikh, and P. Cuatrecasas.** 1974. A simplified method for cyanogen bromide activation of agarose for affinity chromatography. Anal. Biochem. **60**:149–152.
32. **Marcum, J. A., and R. D. Rosenberg.** 1985. Heparinlike molecules with anticoagulant activity are synthesized by cultured endothelial cells. Biochem. Biophys. Res. Commun. **126**: 365–372.
33. **Metcalf, D., G. R. Johnson, and T. E. Mandel.** 1979. Colony formation in agar by multipotential hemopoietic cells. J. Cell. Physiol. **98**:401–420.
34. **Mount, S. M.** 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **10**:459–472.
35. **Ny, T., F. Elgh, and B. Lund.** 1984. The structure of the human tissue-type plasminogen activator gene: correlation of intron and exon structures to functional and structural domains. Proc. Natl. Acad. Sci. USA **81**:5355–5359.
36. **Odell, T. T., Jr., and C. W. Jackson.** 1968. Polyploidy and maturation of rat megakaryocytes. Blood **32**:102–110
37. **Odell, T. T., Jr., and C. W. Jackson, T. J. Friday, and D. E. Charsha.** 1969. Effects of thrombocytopenia on megakaryocytopoiesis. Br. J. Haematol. **17**:91–101.
38. **Okayama, H., and P. Berg.** 1982. High-efficiency cloning of full-length cDNA. Mol. Cell. Biol. **2**:161–170.
39. **Oyo, R., A. Nakeff, S. S. Huang, M. Ginsberg, and T. F. Deuel.** 1983. New synthesis of a platelet-specific protein: platelet factor 4 synthesis in a megakaryocyte-enriched rabbit bone marrow culture system. J. Cell Biol. **96**:515–520.
40. **Paulus, J.-M.** 1970. DNA metabolism and development of organelles in guinea-pig megakaryocytes: a combined ultrastructural, autoradiographic and cytophotometric study. Blood **35**:298–311.
41. **Penington, D. G., and K. Streatfield.** 1975. Heterogeneity of megakaryocytes and platelets. Ser. Haematol. **8**:22–48.
42. **Penington, D. G., K. Streatfield, and S. M. Weste.** 1974. Megakaryocyte ploidy and ultrastructure in stimulated thrombopoiesis, p. 115–130. *In* M. Baldini and S. Ebbe (ed.), Production, function, transfusion and storage. Grune and Stratton, Inc., New York.
43. **Rabellino, E. M., R. B. Levene, L. L. K. Leung, and R. L. Nachman.** 1981. Human megakaryocytes. II. Expression of platelet proteins in early marrow megakaryocytes. J. Exp. Med. **154**:88–100.
44. **Rosenberg, R. D., and P. Damus.** 1973. The purification and mechanism of human antithrombin-heparin cofactor. J. Biol. Chem. **248**:6490–6505.
45. **Ross, R.** 1971. The smooth muscle cell. II. Growth of smooth muscle in culture and formation of elastic fibers. J. Cell Biol. **50**:172–186.
46. **Schöler, H. R., and P. Gruss.** 1984. Specific interaction between enhancer-containing molecules and cellular components. Cell **36**:403–411.
47. **Senior, R. M., G. L. Griffin, J. S. Huang, D. A. Walz, and T. F. Deuel.** 1983. Chemotactic activity of platelet alpha granule proteins for fibroblasts. J. Cell Biol. **96**:382–385.
48. **Tamkun, J. W., J. E. Schwarzbauer, and R. O. Hynes.** 1984. A single rat fibronectin gene generates three different mRNAs by alternative splicing of a complex exon. Proc. Natl. Acad. Sci. USA **81**:5140–5144.
49. **Vinci, G., A. Tabilio, J. F. Deschamps, D. VanHacke, A. Henri, J. Guichard, P. Tetteroo, P. M. Lansdorp, T. Hercend, W. Vainchenker, and J. Breton-Gorius.** 1984. Immunological study of in vitro maturation of human megakaryocytes. Br. J. Haematol. **56**:589–605.
50. **Wagner, R. C., P. Kreiner, R. J. Berrnett, and M. W. Bitnensky.** 1972. Biochemical characterization and cytochemical localization of a catecholamine-sensitive adenylate cyclase in isolated capillary endothelium. Proc. Natl. Acad. Sci. USA **69**: 3175–3179.
51. **Walz, D. A., V. Y. Wu, R. DeLamo, H. Dene, and L. E. McCoy.** 1977. Primary structure of human platelet factor 4. Thromb. Res. **11**:893–898.
52. **Young, R. A., and R. W. Davis.** 1983. Yeast RNA polymerase II genes: isolation with antibody probes. Science **222**:778–782.
53. **Young, R. A., and R. W. Davis.** 1983. Efficient isolation of genes by using antibody probes. Proc. Natl. Acad. Sci. USA **80**: 1194–1198.