Vol. 156, No. 1, 1988
October 14, 1988

BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
Pages 595-601

ISOLATION OF THE HUMAN PLATELET GLYCOPROTEIN IIb GENE AND CHARACTERIZATION OF THE 5' FLANKING REGION

Prandini M.H., Denarier E., Frachet P., Uzan G. and Marguerie G.

Laboratoire d'Hématologie, INSERM U.217,
Fédération des Laboratoires de Biologie, CENG. 85X, 38041 Grenoble Cedex, France

Received September 6, 1988

SUMMARY: Platelet membrane glycoprotein (GP) IIbIIIa complex functions as a receptor for fibrinogen, von Willebrand factor and fibronectin, and mediates adhesive reactions of platelets. The gene for the GPIIb subunit is only active in megakaryocytic cell type. We have isolated this gene from a genomic library. The GPIIb gene was characterized by restriction mapping and sequencing of the 5' and 3' regions containing the first and the last exons. The transcription start site and the polyadenylation signal were identified. From these data we deduced that the gene spans a region of 22 kb and that the mRNA contains a leader sequence of 32 nucleotides. At the 3' end the last exon encodes the 19 amino acids corresponding to the cytoplasmic domain of the GPIIb light chain. Upstream the transcription start site, two sequences are homologous to consensus binding sites of the nuclear factors SP1 and CP2. Two inverted repeats were also identified in this region.

Cell-cell contact and interaction between cells and extracellular matrices are mediated by a broadly distributed family of adhesion receptors which has recently been named integrins (1) . Each member of this superfamily is a non covalently bound heterodimer with α and β subunits containing membrane spaning domains. The superfamily of integrins, is composed of the different groups of adhesion receptors, including the leucocyte adhesion molecule (Leu-CAM), the VLA family and the cytoadhesins (for review see 1 and 2). Within each subfamily the same β subunit is associated with variable α subunits forming different heterodimers with specific structural and functional features. During the past two years most of the subunits of the different integrins have been cloned and their amino acid sequences were deduced from the cDNA sequence (3-8). A central question now emerging, concerns the identification of the mechanisms implicating the different integrins in cell-differenciation and phenotypic selection.

With available cDNA clones for these receptors it is now feasible to compare the structure of the corresponding genes and to identify the mechanisms that control their tissue specific expression. Two members of the cytoadhesin family have been identified : GPIIbIIIa from platelets and megakaryocytes and the vitronectin receptor which is expressed in endothelial cells (9,10). The two receptors have an RGD recognition specificity (11,12) and share the same β subunit (GPIIIa). While the β subunit is expressed in different cellular systems, the α subunits appear to be cell specific. Thus GPIIb

0006-291X/88 $1.50

expression seems to be restricted to the megakaryocytic lineage and consequently constitutes a suitable marker for the studies of the mechanisms governing megakaryocytopoiesis. In this paper we describe the isolation of genomic clones containing the entire GPIIb gene and the characterization of the 5' flanking domain of this gene.

METHODS

Screening of human genomic library and characterization of DNAs. A genomic library constructed in λEMBL4 phage from partially Mbo I digested human leucocyte DNA was generously provided by Dr A. Kahn (Paris). It was screened with different human GPIIb cDNA, that have already been described (4,13). Three specific cDNA clones were used for the screening : λIIb_2, corresponding to the 3' region, λIIb_3 corresponding to approximately 80% of the GPIIb cDNA and λIIb_4 corresponding to the 5' end of the cDNA (Fig. 1A). In addition an 18 bases oligonucleotide corresponding to the 5' end of the cDNA in position +37 to +54 (Fig. 3) was synthesized using the phosphoramidite method with an Applied Biosystem DNA synthesizer and used to further align the genomic clones. λIIb_3 was used as probe for Southern blot analysis of genomic DNA and of the DNA from the positive clones (14). DNA sequencing was performed by the dideoxynucleotide chain termination method (15). DNA fragments were subcloned in PGEM1 plasmid (Promega biotec, Madison WI. U.S.A.) and the sequence was obtained on both strands using synthetic oligonucleotides.
Primer extension. Total RNA was extracted from human megakaryocytes (16) using the thiocyanate-guanidinium method (17) and the poly A^+ mRNA were purified by chromatography on oligo (dT) cellulose column (18). The 5' end labelled oligonucleotide (7.10^6 cpm) was hybridized to poly A^+ mRNA and elongated using the protocol and reagents of the first strand cDNA synthesis kit (Amersham England).
RNase mapping. RNase mapping with Sp6 probes was carried out essentially as previously described (19). The templates used for the generation of the anti-sens RNA probes were obtained by cloning a 1.4 kb Nco I-Eco RI fragment and a 1 kb Bgl I-Bgl I fragment isolated from the 1.9 kb 5' subclone, in the appropriate orientation, in the PGEM1 plasmid. For synthesis of the Sp6 probes, the templates were linearized by Dde I and Pvu II respectively. Total RNA (50 μg) was hybridized overnight to 5.10^6 cpm of probe at 45°C and digested with RNase T1 and RNase A. The RNase resistant products were subsequently analysed on 6% acrylamide sequencing gel.

RESULTS

Isolation and characterization of the human GPIIb gene. Approximatively 10^6 recombinants from the amplified genomic library were screened using different human GPIIb cDNA clones as hybridization probes (Fig.1 A). Five independent clones were identified, isolated and purified. Alignement of these clones was achieved by restriction enzyme digestion and Southern blotting analysis using 3' and 5' restriction fragments from the cDNA probes. Two overlapping genomic clones spaning about 30 Kb were found to contain the boundaries of the gene. A partial restriction map of the GPIIb gene is shown in Fig. 1B. Comparison of the size of the restriction fragment predicted from the isolated genomic clones with that of the fragments obtained by the digestion of the genomic DNA indicates that there is only one copy of the GPIIb gene in the human genome (data not shown).
Characterization of the 3' end of the gene. A Hind III-Eco RI 125 bp fragment generated by the digestion of the λIIb_3 cDNA clone and corresponding to the 3' non coding region of

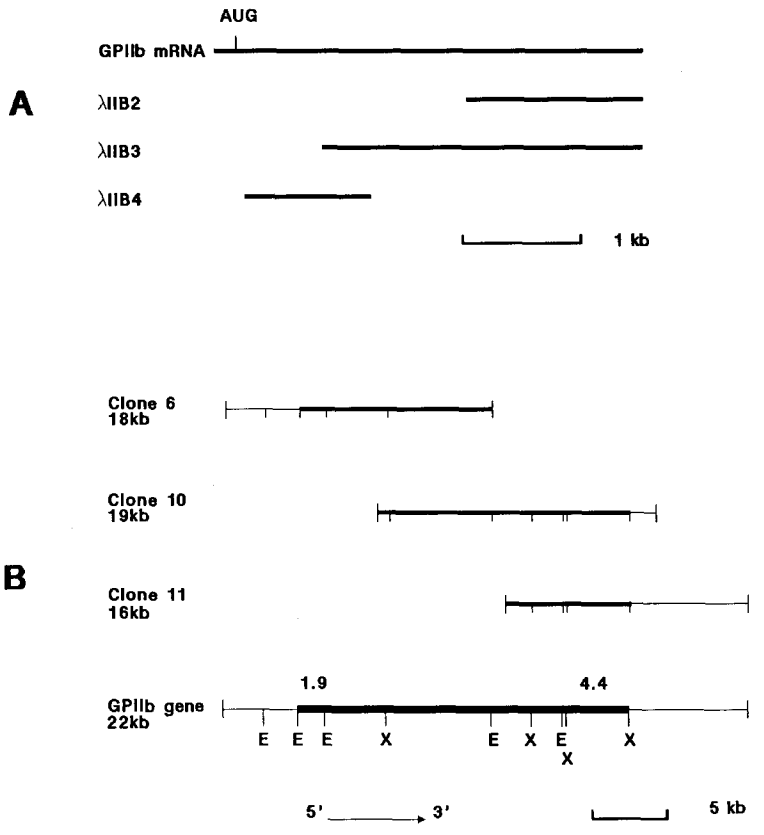

Fig. 1 A : Location of the different cDNA probes used for screening of the genomic library in relation to the GPIIb mRNA.
B : Position of three genomic DNA clones and restriction map of GPIIb gene. The thick line represents genomic fragments which hybridize with cDNA.

the mRNA was purified and used as probe to further localize the 3' end of the gene. This probe hybridized with an Xho I-Xho I 4.4 Kb fragment of genomic clones 10 and 11 (Fig. 1B). This fragment was subcloned in PGEM1 plasmid and sequenced. A partial sequence containing an exon encoding the 19 C-terminal amino acid residues of GPIIb and 196 nucleotides of the 3' untranslated region present in the mRNA is shown in Fig. 2. Only one polyadenylation site with the consensus sequence AATAAA was found 15 bases before the end of the exon. No other polyadenylation sites were found within 200 bases downstream.

Characterization of the 5' end of the gene and localization of the transcriptionnal start site. An 18 bases long oligonucleotide corresponding to the sequence of the 5' end of the GPIIb cDNA (20) was used to define more precisely the 5' region of the gene. This probe hybridized with a 1.9 Kb Eco RI-Eco RI fragment from the genomic clone 6 (Fig. 1B). This fragment was subcloned in the PGEM1 plasmid and sequenced. The complete sequence is shown in Fig. 3. Comparison with the cDNA sequence (20) and the N-terminal domain of GPIIb (21) indicates that this fragment contains the putative translation initiation codon

Fig. 2 DNA sequence of the 3' end of the human GPIIb gene.
The sequence shown covers the exon encoding the 19 C-terminal amino acids and a 3' untranslated region. The polyadenylation signal is boxed. The asterisk indicates the last nucleotide of the gene. Dotted line indicates the 3'splice site.

ATG, the entire signal peptide and 31 N-terminal amino acid residues of GPIIb. This exon is followed by an intron extending up to the end of this 1.9 Kb fragment. The exon-intron boundary sequence is consistent with the consensus sequence GT (A/G) AGT (22).

To identify the transcription initiation site for human GPIIb mRNA, a combination of primer extension and RNase mapping analysis was used. The primer extension experiments were performed with two oligonucleotides (A and B) corresponding to different positions on the exon (see Fig. 3). The extended products were resolved on a 6% denaturing polyacrylamide gel. After autoradiography, two fragments of 217 and 76 bases long respectively were visualized (Fig. 4). The RNase mapping analysis were performed using two anti-sens RNA probes hybridized with total RNA from megacaryocytes. The protected fragments extend 212 bases from the Nco I site and 170 bases from the Bgl I site (Fig. 5). All of the RNases mapping and primer extension analysis indicated similar 5' terminus, suggesting that the exon is the first one in the GPIIb gene and that its 5' end corresponds to the transcriptionnal start site indicated as +1 in Fig.3 .

DISCUSSION

The screening of a human genomic library has enable us to identify DNA segments containing the gene coding for the platelet glycoprotein IIb. From the positions of the transcription initiation site as determined by primer extension and RNases mapping, and the polyadenylation site corresponding to the cDNA, the GPIIb gene spans a region of 22 kb in length. The restriction map of this gene is in agreement with the size of the restriction fragment obtained from the digestion of genomic DNA indicating that there is only one copy of this gene. Results obtained from both primer extension and RNases mapping were compatible and identified a single transcription initiation start corresponding to an A in the 5' DNA sequence (Fig.3).

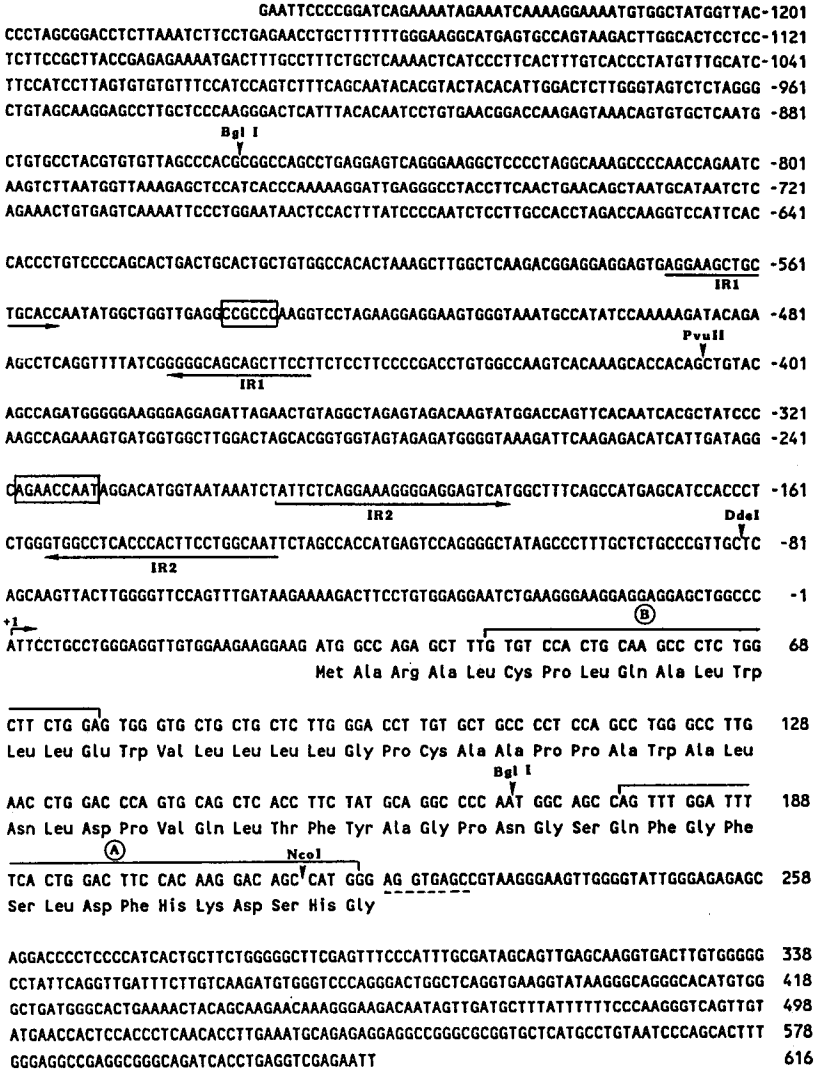

Fig. 3 DNA sequence of the first exon and 5'flanking region of the human GPIIb gene. The 1.9 kbp Eco RI fragment of genomic clone 6 was subcloned in PGEM1 and complete sequence was obtained using various synthetic oligonucleotides as sequencing primers. The A designated as +1 is the putative transcription start site, the arrow below shows the direction of transcription. Potential regulatory elements discussed in the text are boxed. Horizontal arrows show inverted repeats. The 5' splice site is indicated with a dotted line. A and B are oligonucleotides used for primer extension experiments.

The only ATG between +1 and the codon corresponding to the amino terminal Leucine of the mature protein is at position +33 and is in the correct reading frame. The deduced signal sequence is in total agreement with the sequence derived from the cDNA which was previously published by Poncz et al. (20). Thus the exon encoding the signal peptide is the first exon of the gene. In addition this exon contains 31 amino acid residues from the N-terminus of the mature protein. At the 3' end of the gene the last exon encodes

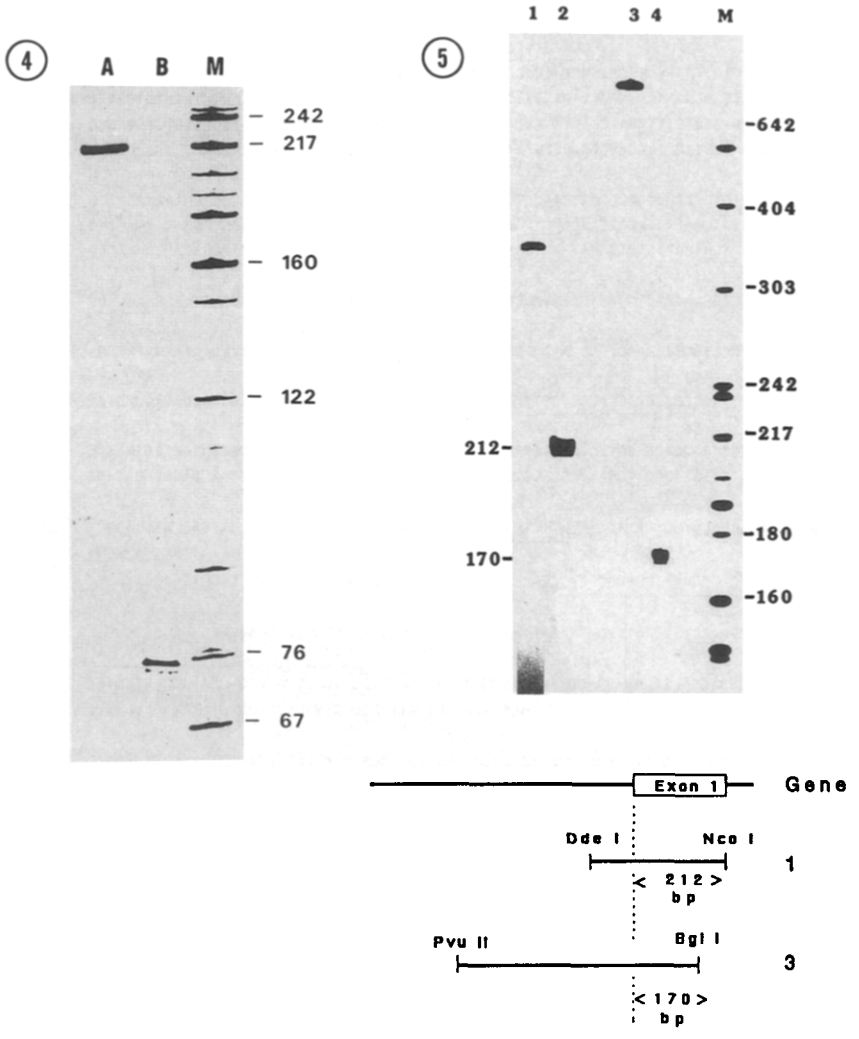

Fig. 4 Primer extension analysis.
Two oligonucleotides A and B (40 and 30 mers respectively) complementary to GPIIb mRNA were ^{32}P-end labelled and elongated with AMV reverse transcriptase after hybridization with poly A$^+$ mRNA from megakaryocytes.
The 40 mers oligonucleotide extended to produce a 217 bases product (lane A), whereas the 30 mers primer produced a 76 bases product (lane B). The molecular weight marker (lane M) is an Hpa II digest of pBR 322.

Fig. 5 RNase mapping experiments.
Complementary RNA transcript were synthesized with Sp6 RNA polymerase using genomic fragment subcloned in PGEM1. Two probes were used, the first one extending between the Nco I and Dde I sites (lane 1), the second extending between Bgl I and Pvu II sites (lane 3). After hybridization with total megakaryocyte RNA and digestion with ribonucleases, protected fragments were analysed on 6% sequencing gels (lane 2 and 4). The molecular weight marker (lane M) is an Hpa II digest of pBR 322.

the complete sequence of the cytoplasmic domain of the GPIIb light chain and the 3' untranslated region of the mRNA including the polyadenylation site. Thus the data presented confirm that the N-terminal methionin published by Poncz et al (20) is the first aminoacid of the pro-GPIIb and further indicate that the mRNA contains a leader

sequence of 32 bases. The limit of the gene is therefore compatible with a mRNA of 3336 bases long (poly-A tail excepted).

Analysis of the 5' flanking sequence indicates that there is no canonical consensus sequence corresponding to the TATA box (22) between -30 to -60, unless the sequence GAATCT at position -31 or the sequence AATAAA at position -55 may serve as TATA box for this gene. In addition there is no CAAT box in the usual domain between -60 to -120 (23). An AGAACCAAT sequence corresponding to the sequence identified as binding site for the CAAT binding protein CP2 (24) is however located at position -228. A sequence CCGCCC is found at position -534 which corresponds to the consensus sequence for the binding of the transcription factor SP1 (25). In addition several inverted repeats can be identified, one of them flanking the CCGCCC consensus sequence (Fig.3).

The knowledge of the sequence surrounding the transcription initiation start site will now allow the study of the regulation of GPIIb during the megakaryocytopoiesis.

REFERENCES

1. Hynes R.O. (1987) Cell 48, 549-554
2. Phillips D.R., Charo I.F., Parise L.V. and Fitzgerald L.A. (1988) Blood 71, 831-843
3. Fitzgerald L.A., Poncz M., Steiner B., Rall S.C., Bennett J.S. and Phillips D.R. (1987) Biochemistry 26, 8158-8165
4. Uzan G., Frachet P., Lajmanovich A., Prandini M.-H., Denarier E., Duperray A., Loftus J., Ginsberg M., Plow E. and Marguerie G. (1988) Eur. J. Biochem. 171, 87-93
5. Fitzgerald L.A., Steiner B., Rall S.C., Lo S.-S. and Phillips D.R. (1987) J. Biol. Chem.262, 3936-3939
6. Kishimoto T.K., O'Connor K., Lee A., Roberts T.M. and Springer T.A. (1987) Cell 48, 681-690
7. Corbi A.L., Miller L.J. O'Connor K., Larson R.S. and Springer T.A. (1987) EMBO J. 6, 4023-4028
8. Arnaout M.A., Remold-O'Donnel E., Pierce M.W., Harris P. and Tenen D.G. (1988) Proc. Natl. Acad. Sci. USA 85, 2776-2780
9. Ginsberg M.H., Loftus J., Ryckwaert J.-J., Pierschbacher M., Pytela R., Ruoslahti E. and Plow E.F. (1987) J. Biol. Chem. 262, 5437-5440
10. Zimrin A.B., Eisman R., Vilaire G., Schwartz E., Bennett J.S. and Poncz M. (1988) J. Clin. Invest. 81, 1470-1475
11. Pytela R., Pierschbacher M. and Ruoslahti E. (1985) Cell 40, 191-196
12. Ruoslahti E. and Pierschbacher M.D. (1986) Cell 44, 517-518
13. Frachet P., Uzan G. and Marguerie G. (1988) in Biochemistry and physiopathology of platelet membrane. Ed Marguerie G., Zwaal R.F.A. Colloque INSERM/John Libbey Eurotext Ltd. 158, 155-161
14. Southern E.M. (1975) J. Mol. Biol. 98, 503-517
15. Sanger F., Nicklen S. and Coulson A.R.(1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467
16. Berthier R., Marin M., Duperray A., Prenant M., Newton I., Schweitzer A., Hollard D. and Marguerie G. (1987) Exp. Hematol. 15, 750-758
17. Chomczynski P. and Sacchi N. (1987) Anal. Biochem. 162, 156-159
18. Aviv H. and Leder P. (1972) Proc Natl Acad Sci USA 69, 1408-1412
19. Melton D.A., Krieg P.A., Rebagliati M.R., Maniatis T., Zinn K. and Green M.R. (1984) Nucl. Acids Res. 12, 7035-7056
20. Poncz M., Eisman R., Heidenreich R., Silver S.M., Vilaire G., Surrey S., Schwartz E. and Bennet J.S. (1987) J. Biol. Chem. 262, 8476-8482
21. Charo I.F., Fitzgerald L.A., Steiner B., Rall S.C., Bekeart L.S. and Phillips D.R. (1986) Proc. Natl. Acad. Sci. USA 83, 8351-8355
22. Breathnach R. and Chambon P. (1981) Annu. Rev. Biochem. 50, 349-383
23. Benoist C. and Chambon P. (1981) Nature 290, 304-310
24. Chodosh L.A., Baldwin A.S., Carthew R.W. and Sharp P.A. (1988) Cell 53, 11-24
25. Briggs M.R., Kadonaga J.T., Bell S.P. Tjian R. (1986) Science 234, 47-52