

STRUCTURE OF THE HUMAN BLOOD PLATELET MEMBRANE GLYCOPROTEIN Ib α GENE

Roland H. Wenger*, Nelly Kieffer**, Andreas N. Wicki* and
Kenneth J. Clemetson*†

* Theodor Kocher Institute, University of Berne, Freiestrasse 1,
CH-3012 Berne, Switzerland

** Institut National de la Santé et de la Recherche Médicale
U.91, Hôpital Henri Mondor, F-94010 France

Received September 1, 1988

Summary: The gene for human platelet glycoprotein Ib α -chain has been cloned from a genomic cosmid library using a partial cDNA clone as probe. 3530 bp were sequenced including the entire transcribed part, as well as additional 5' and 3' regions. A single intron was found 6 bp upstream of the ATG initiation codon. An exceptionally long exon was identical to the recently published cDNA sequence (1). The 5' upstream promoter region is atypical for eukaryotic genes with only a weak homology to the characteristic promoter consensus sequences. The 3' region contains two repetitive Alu elements, belonging to distinct subfamilies, connected by an oligo(dA) linker. © 1988 Academic Press, Inc.

Human blood platelet glycoprotein (GP) Ib is the receptor on unactivated platelets for von Willebrand factor (vWf), a large multimeric plasma component. As such it plays a critical role in adhesion of platelets to the subendothelium of damaged blood vessels in the primary stages of haemostasis (reviewed in refs. 2 and 3). GPIb consists of two subunits, α and β , with masses of 150 kDa and 25 kDa, respectively, joined by disulphide bonds (4). cDNA coding for both these subunits has been cloned and sequenced (1,5) and both contained leucine-rich repeats. In platelet membranes GPIb exists as a 1:1 complex with another glycoprotein, GPIX (6). The α subunit of GPIb contains a

† To whom correspondence should be addressed.

Abbreviations: GP: glycoprotein; vWf: von Willebrand factor; bp: base pair; kb: kilobase pair.

thrombin binding site which is involved in the kinetics of platelet activation by thrombin (7). Several bleeding disorders are known in which there is an inherited deficiency or defect in GPIb. These include Bernard-Soulier syndrome (8,9), pseudo von Willebrand disease (10) and Bolin-Jamieson syndrome (11). In Bernard-Soulier syndrome both subunits of GPIb, GPIX and a further glycoprotein, GPV, are all missing or affected (12), suggesting an association between the genetic control or cellular expression of these components. As a step towards understanding the factors controlling expression of the GPIb complex in the megakaryocyte and in platelets we have isolated and sequenced the gene for GPIb α , which contains several unusual features.

MATERIALS AND METHODS

Preparation of a cDNA probe. Human platelet mRNA was isolated and used for the construction of a λ gt11 expression library as described (13,14). Polyclonal antibodies against GPIb α were used for screening this library. A partial cDNA clone for GPIb α , identical to nucleotide sequence 661 to 1444 of the full length cDNA (1), was isolated and sequenced (13). The probe was labelled with [α^{32} P]dCTP by the random primed labelling method (Amersham).

Screening of genomic library. This cDNA probe was used for screening a pCV105 cosmid library (15,16). Colonies that hybridized positively were rescreened and the cosmid DNA isolated by the alkaline extraction procedure (17). DNA (25 μ g) was digested with EcoRI (Boehringer Mannheim), separated by 0.8% agarose gel electrophoresis and blotted onto nitrocellulose membranes (BA 85, Schleicher & Schuell). The EcoRI fragment that hybridized with the cDNA probe on Southern blots, was isolated from agarose gels by electroelution in a membrane trap (Bio-Trap, Schleicher & Schuell), phenol extracted and precipitated with ethanol.

Nucleotide sequencing. This fragment was subcloned into M13 Bluescript (+) and (-) (Stratagene) by ligation to the linearized plasmid with T4 ligase (New England Bio-Labs) and transformation into *E. coli* strains JM 101, JM 107 and JM 109 by the CaCl_2 method (18). Single stranded DNA templates were prepared by superinfection with M13K07 helper phage. Alternatively, double stranded templates were used as described (19). The DNA sequence was determined by the dideoxy-chain-termination method (20) using [α^{35} S]dATP (New England Nuclear) and either the modified T7 DNA polymerase (Sequenase, United States Biochemicals) or the unmodified form (Pharmacia). The sequencing strategy involved synthetic primers and restriction mapping. Universal M13 and reversed M13 primer were purchased from Pharmacia, 18 to 24 bp oligodeoxynucleotides were prepared on a Applied Biosystems 381A DNA synthesizer and purified by gel filtration on NAP-10 columns (Pharmacia). Restriction enzymes were from Boehringer-Mannheim and New England Bio-Labs. Restriction fragments were isolated and subcloned for sequencing

as above. Sequences were assembled and analysed using the MicroGenie program (Beckmann).

RESULTS AND DISCUSSION

A pCV105 cosmid library (15,16) was screened (200,000 colonies) with a 783 bp partial cDNA probe (Fig. 1A) encoding human platelet GPIb α . Fifteen colonies gave a positive signal with this probe and secondary screenings identified 4 clones containing a 6.5 kb EcoRI fragment which hybridized to the same probe on Southern blots. Since genomic Southern blotting also revealed an EcoRI fragment of 6.5 kb (21), one of them (N10) was chosen for sequencing. 3530 bp were sequenced (Fig. 1B) including 261 bp of the 5' non-transcribed region, a single, 233 bp long intron 6 bp upstream of the ATG initiation codon and 646 bp downstream of the polyadenylation site at the 3' end (Fig. 2).

The 5' untranscribed region of the GPIb α gene has some unusual features. No typical TATA or CAAT consensus sequences, which play an important role in promotion of transcription in many eukaryotic genes (22), were found. However, there are two uncharacteristic sequences at relative positions where TATA and

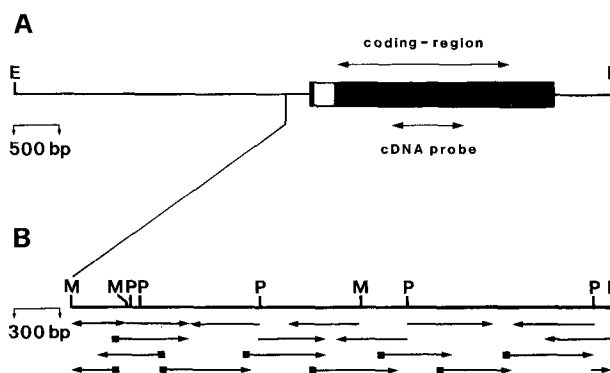


Figure 1: Cloning and sequencing strategy of the GPIb α gene.

Fig. 1A: The line represents the 6,5 kb EcoRI fragment of the GPIb α gene. 3530 bp were sequenced, including the transcribed part (filled bar) and a single intron (open bar). The lengths of the coding region and the cDNA probe used for screening the genomic library are indicated.

Fig. 1B: Restriction map and sequencing strategy. The extent and direction of sequencing are indicated by arrows. Solid boxes associated with the arrows indicate that the sequences were determined by using synthetic primers. The restriction sites are: E: EcoRI; M: MseI; P: PstI.

A

5'- TTAAAAGATGGCAGAAGGCTGTTTGGAGGAGTCCACCCCATCTCCCCTGTGTAAA 56
 AGGAAAGCGGAAGAGAGAACCACAAAGAGGGCCTGGGGGAAAGCCGTGGAGTGAGGCGAT 116
 AAGGGCTTGTGTCCAGGGGATTCCCGGTCACTGGAATCCCTATCAGGCTGCATTTCTCTC 176
 CTCACCCCCATCCCCTTCCCTTGCCACTGGCTTAGTCCCTCCATGGGGCTAGAAGAGAGAAG 236
 GACGGAGTCGAGTGGCACCCTAGAAGACGCTCTGTGCCTTCGGAGGTCTTTCTGCCTGCC 296
 Tgtaagccggggttggtgctggtgggagagaggggtctgagggaggggaaagagccaagg 356
 acctggagctagtagtttttaagttctgcaggcaagggtgggagatgggagtagggaggac 416
 agggaggtgtggatgctgtttctggaagcgaagctgcagggggaaggggctggggcctgg 476
 ggggatgcttccaggggatgcaggggatccactcaaggctcccttggccacagGTCTCTC 536

M P L L L L L L L L L P S P L H P H P I C 20
 ATGCCTCTCCTCCTCTGTGCTGCTGCTGCCAAGCCCTTACACCCACCCCATCTGT 596
 E V S K V A S H L E V N C D K R N L T A 40
 GAGGTCTCCAAAGTGGCCAGCCACCTAGAAAGTGAAGTGAAGAGGAATCTGACAGCG 656
 L P P D L P K D T T I L H L S E N L L Y 60
 CTGCCTCCAGACTGCCGAAAGACACAACCATCCTCCACCTGAGTGAGAACCTCCTGTAC 716
 T F S L A A T L M P Y T R L T Q L N L D R 80
 ACCTTCTCCCTGGCAACCCTGATGCCTTACACTCGCCTCACTCAGCTGAACCTAGATAGG 776
 C E L T K L Q V D G T L P V L G T A D L 100
 TGGGAGCTCACCAAGCTCCAGGTTCGATGGGACGCTGCCAGTGTGGGGACCTGGATCTA 836
 S H N Q L Q S L P L L G Q T L P A L T V 120
 TCCACAAATCAGCTGCAAGCCTGCCCTTGCTAGGGCAGACACTGCCTGCTCTCACCGTC 896
 L D V S F N R L T S L P L G A L R G L G 140
 CTGGACGTCTCCTTCAACCGCTGACCTCGCTGCCTCTGGGTGCCCTGCGTGGTCTTGGC 956
 E L Q E L Y L K G N E L K T L P P G L L 160
 GAACTCCAAGAGCTCTACCTGAAAGGCAATGAGCTGAAGACCTGCCCCACAGGGCTCCTG 1016
 T P T P K L E K L S L A N N N L T E L P 180
 ACGCCACACCAAGCTGGAGAAGCTCAGTCTGGCTAAACAACAATTGACTGAGCTCCCC 1076
 A G L L N G L E N L D T L L L Q E N S L 200
 GCTGGGCTCCTGAATGGGCTGGAGAATCTCGACACCCCTCTCCTCCAAGAGAAGCTCGCTG 1136
 Y T I P K G F F G S H L L P F A F L H G 220
 TATACAATACCAAGGGCTTTTTTGGGTCCCACCTCCTGCCTTTTGCTTTTCTCCACGGG 1196
 N P W L C N C E I L Y F R R W L Q D N A 240
 AACCCCTGGTTATGCAACTGTGAGATCCTCTATTTTCGTCGGCTGCAGGACAATGCT 1256
 E N V Y V W K Q G V D V K A M T S N V A 260
 GAAAATGTCTACGTATGGAAGCAAGGTGTGGACGTCAAGGCCATGACCTCTAACGTGGCC 1316
 S V Q C D N S D K F P V Y K Y P G K G C 280
 AGTGTGCACTGTGACAATTGACAAAGTTTCCCGTCTACAAATACCCAGGAAAGGGGTGC 1376
 P T L G D E G D T D L Y D Y Y P E E D T 300
 CCCACCTTGGTGTATGAAGGTGACACAGACCTATATGATTACTACCCAGAAGAGGACACT 1436
 E G D K V R A T R T V V K F P T K A H T 320
 GAGGGCGATAAGGTGCGTGGCCACAAGGACTGTGGTCAAGTTCCCCACCAAGGCCATACA 1496
 T P W G L F Y S W S T A S L D S Q M P S 340
 ACCCCCTGGGGTCTATTCTACTCATGGTCCACTGCTTCTCTAGACAGCCAAATGCCCTCC 1556
 S L H P T Q E S T K E Q T T F P P R W T 360
 TCCTTGCAATCCAACAAGAATCCACTAAGGAGCAGACCACATTCCCACCTAGATGGACC 1616
 P N F T L H M E S I T F S K T P K S T 380
 CCAAATTCACACTTCACATGGAATCCATCACATCTCCTCAAACTCCAAATCCACTACT 1676
 E P T P S P T T S E P V P E P A P N M T 400
 GAACCAACCCCAAGCCGACCACTCAGAGCCCGTCCCGGAGCCCGCCCAACATGACC 1736

Figure 2: Sequence of the GPIb α gene.

5' region: The intron is indicated by lower case letters. The stretch with a certain homology to the CAAT consensus is double underlined and the direct repeated stretches with weak homology to the TATA consensus are single underlined. The beginning of the cloned full length cDNA (1) is represented by an arrow.

3' region: The polyadenylation signal is double underlined and the polyadenylation site is marked with a triangle. The two Alu repetitive elements are underlined.

CAAT boxes have usually been found. These are the GGCCTGCAT and the repeated CTAGAAGA stretch starting at position -100, -39 and -6 respectively, relative to the estimated beginning of the cDNA

B

```

T L E P T P S P T T P E P T S E P A P S 420
ACCCTGGAGCCCACTCCAAGCCCGACCACCCAGAGCCCACTCAGAGCCCGCCCCAGC 1796
P T T P E P T P I P T I A T S P T I L V 440
CCGACCACCCCGGAGCCCAATCCCGACCATCGCCACAAGCCCGACCATCCTGGTG 1856
S A T S L I T P K S T F L T T T K P V S 460
TCTGCCACAAGCCTGATCACTCCA AAAAGCACATTTTAACTACCACAAAACCCGTATCA 1916
L L E S C T K K T I P E L D Q P P K L R G 480
CTCTTAGAATCCACCAAAAAACCATCCCTGAACTTGATCAGCCACCAAAGCTCCGTGGG 1976
V L Q G H L E S S R N D P F L H P D F C 500
GTGCTCCAAGGGCATTTGGAGAGCTCCAGAAATGACCCTTTCTCCACCCCGACTTTTGC 2036
C L L P L G F Y V L G L F W L L F A S V 520
TGCTCCTCCCCCTGGGCTTCTATGTCTTGGGTCTCTTCTGGCTGCTCTTTGGCTCTGTG 2096
V L I L L L S W V G H V K P Q A L D S G 540
GTCTCATCTGCTGCTGAGCTGGGTGGGCATGTGAAACACAGGCCCTGGACTCTGGC 2156
Q G A A L T T A T Q T T H L E L Q R G R 560
CAAGGTGCTGCTCTGACCACAGCCACACAAACACACACCTGGAGCTGCAGAGGGGACGG 2216
Q V T V P R A W L L F L R G S L P T F R 580
CAAGTGACAGTGCCCGGGCTGGCTGCTTCTTCGAGGTTCGCTTCCCACTTTCCGC 2276
S S L F L W V R P N G R V G P L V A G R 600
TCCAGCCTCTTCTGTGGGTACGGCCTAATGGCCGTGTGGGGCCTCTAGTGGCAGAAG 2336
R P S A L S Q G R G Q D L L S T V S I R 620
AGGCCCTCAGCTCTGAGTCAGGGTCGTGGTCAGGACCTGCTGAGCACAGTGAAGCATTAG 2396
Y S G H S L TER 626
TACTCTGGCCACAGCCTCTGAGGGTGGGAGGTTTGGGGACCTTGAGAGAAGAGCCTGTGG 2456
GCTCTCTATTGGAATCTAGTTGGGGGTGGAGGGGTAAGGAACACAGGGTGATAGGGGA 2516
GGGGTCTTAGTTCCTTTTCTGTATCAGAAGCCCTGTCTTCACAACACAGGCACACAATT 2576
TCAGTCCCGCCAAAGCAGAAGGGTAATGACATGGACTTGGCGGGGGACAAGACAAAG 2636
CTCCCGATGCTGCATGGGGCGCTGCCAGATCTCACGGTGAACCATTTTGGCAGAATACAG 2696
CATGGTTCACCATGCATCTATGCACAGAAGAAATCTGGAAGTGATTTTATCAGGATGT 2756
GAGCACPCGTTGTGTCTGGATGTTACAAATATGGGTGTTTTATTTTCTTTTCCCTGTT 2816
TAGCATTTTCTAGTTTTCACACTATTATTGTATATTATCTGTATAATAAAAAATAATTTTA 2876
GGGTTGGGAGTGATGGCTCATGCTGTAATCCTAGCACTTTGGGAGGCCGAGGCCGGGTGG 2936
AATCACCAGAGGTAGGGAGTTCAAGACCAGCCTGGCAAAACATGGTGAAACCCCTGGTCTCT 2996
ACTAAAAATACAAAAATTAGGCCAGGCGTGGTGGTGACACCTATAACCCAGCTACTCG 3056
GGAGGGTGGGGCAGGAGAATCGCTTGAACCTGGGAGGCGGAAGTTGCCGTGAGCCAAGAT 3116
CGTACCAGTGAACCTCCAGCCTGGGTAACAGAGTGAGACTCCGTCTAAAAA AAAAAAAAAA 3176
AAAAAAAAAACTTCTGGCCGGGTGCAGGGGCTCATGCCTGTAATTCCAGCACTCTGGAA 3236
GGCTGAGGCGGGTGGGTGCTTGAACCCAGGAGTTTGGCCCAGGCTTGGCAACATGGCAA 3296
AACCCGACCTCTACAAAAATACAAAACATTAGCCAGGTGTGGTGGCATGCACCTGTGGT 3356
CCCAGGTACCCGGGTGGCTGAGGAGGGAGGATCACCTGAGCCTGGGAGATGGAGGCTGCA 3416
GTGAGCCCTGAAGGTGCCACTGTACTCCAGCCTGGGTGACAGAGTGAGAGCCTGTCTCAA 3476
AACAACTTGGCTTCTTTTGGTGAAGAGTGGCTGGGGACCTGTGATGAGAATTC - 3' 3530

```

Figure 2 - Continued.

(1) (Fig. 2). Further investigations will show whether the homology to the TATA^A/T^A/T and GG^T/C^AATCA consensus sequences (22) is mere chance or if there exists a new family of promoter sequences controlling the expression of platelet GPIb complex proteins.

The exon-intron structure of the GPIb α gene is also rather uncommon for eukaryotic genes. The whole precursor of GPIb α (626 amino acids) is expressed by a single, exceptionally large exon of 2,4 kb. No difference in the nucleotide sequences was found between our genomic clone and the published cDNA (1). The gene

contains a single intron of 233 bp, 6 bp upstream from the ATG translation initiation codon in the 5' non-translated region. This intron has an unusually high content of guanosine (46% of total). A computer-assisted search of the GeneBank (DNASTar, release 55) revealed no significant homology to any other known sequence. The exon-intron splice junctions are in good agreement with the consensus sequences determined for correct splicing of the precursor mRNA (23). Interestingly, von Willebrand factor, for which GPIb α is the receptor on platelets, also contains an intron just before the initiation site (24,25). Whether this is important for cell specific expression of GPIb α , as has been suggested for vWf, remains to be shown.

The 3' region of the GPIb α gene contains two Alu repeats (Fig. 2). An alignment of the first of these (293 bp) with the published Alu-S consensus sequence (26,27) revealed 88% homology. The second (291 bp) is 83% homologous to the Alu-J subfamily, matching 11 of 15 diagnostic identification points determined for this subfamily, which is more similar to the 7SL DNA (28) in these positions (27). In contrast the Alu-S element did not match in any of these positions. These two repetitive Alu elements are connected by a 26 bp oligo(dA) linker, an element that is characteristic for Alu repeats. It should be noted that the first Alu consensus already starts 6 bp upstream from the polyadenylation site of the determined cDNA (Fig. 2), excluding the possibility of further mRNA processing signals downstream of the first polyadenylation signal.

While the origin of the genetic defect in Bernard-Soulier syndrome, which affects not only GPIb α , but also GPIb β , GPV and GPIX, is not yet established the information supplied here on the structure of the GPIb α gene is a necessary step on the way to understanding how the biosynthesis of the GPIb complex is regulated. The sequences of the genes for the other members of the complex will be required to determine whether there are common regulatory elements or whether concerted expression is a post transcriptional event as has been postulated for the GP IIb/IIIa complex (29).

ACKNOWLEDGMENTS

We wish to thank Dr. T. Seebeck, Dr. H.-P. Amstutz and Dr. A. Walz for helpful discussions. This work was supported by the Swiss National Science Foundation grant number 3.232-0.85.

REFERENCES

1. Lopez, J.A., Chung, D.W., Fujikawa, K., Hagen, F.S., Papayannopoulou, T. and Roth, G.J. (1987) *Proc. Natl. Acad. Sci. USA* 84, 5615-5619.
2. George, J.N., Nurden, A.T. and Phillips, D.R. (1984) *N. Engl. J. Med.* 311, 1084-1098.
3. Clemetson, K.J. and Lüscher, E.F. (1988) *Biochim. Biophys. Acta* 947, 53-73.
4. Phillips, D.R. and Poh-Agin, P. (1977) *J. Biol. Chem.* 252, 2121-2126.
5. Lopez, J.A., Chung, D.W., Fujikawa, K., Hagen, F.S., Davie, E.W. and Roth, G.J. (1988) *Proc. Natl. Acad. Sci. USA* 85, 2135-2139.
6. Du, X., Beutler, L., Ruan, C., Castaldi, P.A. and Berndt, M.C. (1987) *Blood* 69, 1524-1527.
7. Wicki, A.N. and Clemetson, K.J. (1985) *Eur. J. Biochem.* 153, 1-11.
8. Nurden, A.T. and Caen J.P. (1975) *Nature* 255, 720-722.
9. Jenkins, C.S.P., Phillips, D.R., Clemetson, K.J., Meyer, D., Larrieu, M.-J. and Lüscher, E.F. (1976) *J. Clin. Invest.* 57, 112-124.
10. Miller, J.L., Kupinski, J.M., Castella, A. and Ruggeri, Z.M. (1983) *J. Clin. Invest.* 72, 1532-1542.
11. Bolin, R.B., Okumura, T. and Jamieson, G.A. (1977) *Nature* 269, 69-70.
12. Clemetson, K.J., McGregor, J.L., James, E., Dechavanne, M. and Lüscher, E.F. (1982) *J. Clin. Invest.* 70, 304-311.
13. Wicki, A.N., Walz, A., Gerber-Huber, S., Vornhagen, R. and Clemetson, K.J., in preparation.
14. Wenger, R.H., Wicki, A.N., Walz, A., Kieffer, N. and Clemetson, K.J., submitted for publication.
15. Lau, Y.-F. and Kan, Y.W. (1983) *Proc. Natl. Acad. Sci. USA* 80, 5225-5229.
16. Lau, Y.-F. and Kan, Y.W. (1984) 81, 414-418.
17. Birnboim, H.C. and Doly, J. (1979) *Nucl. Acids Res.* 7, 1513-1522.
18. Maniatis, T., Fritsch, T.F. and Sambrook, J. (1982) *Molecular Cloning*, Cold Spring Harbor, New York.
19. Zhang, H., Scholl, R. and Somerville, C. (1988) *Nucl. Acids Res.* 16, 1120.
20. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
21. Wicki, A.N. and Drouin, J., personal communication.
22. Hejine, G. (1987) *Sequence Analysis in Molecular Biology*, Academic Press, New York.
23. Mount, S.M. (1982) *Nucl. Acids Res.* 10, 459-472.
24. Bonthron, D. and Orkin, S.H. (1988) *Eur. J. Biochem.* 171, 51-57.
25. Assouline, Z., Kerbirou-Nabias, D.M., Piétu, G., Thomas, N., Bahnak, B.R. and Meyer, D. (1988) *Biochem. Biophys. Res. Comm.* 153, 1159-1166.
26. Britten, R.J., Baron, W.F., Stout, D.B. and Davidson, E.H. (1988) *Proc. Natl. Acad. Sci. USA* 85, 4770-4774.
27. Jurka, J. and Smith, T. (1988) *Proc. Natl. Acad. Sci. USA* 85, 4775-4778.
28. Ullu, E. and Tschudi, C. (1984) *Nature* 321, 171-172.
29. Wautier, J.L. and Caen, J.P. (1987) *Nouv. Rev. Fr. Hematol.* 29, 277.