

Alternative Expression of Platelet Glycoprotein Ib β mRNA from an Adjacent 5' Gene with an Imperfect Polyadenylation Signal Sequence

Barbara Zieger, Yoshimi Hashimoto, and Jerry Ware

Roon Research Center for Arteriosclerosis and Thrombosis, Division of Experimental Hemostasis and Thrombosis, Department of Molecular and Experimental Medicine, and Department of Vascular Biology, The Scripps Research Institute, La Jolla, California 92037

Abstract

Glycoprotein (GP) Ib is a major component of the platelet membrane receptor for von Willebrand factor, designated the GP Ib-IX-V complex. GP Ib is composed of two subunits (GP Ib α and GP Ib β) each synthesized from separate genes. The 206 amino acid precursor of GP Ib β is synthesized from a 1.0-kb mRNA expressed by megakaryocytes and was originally characterized from cDNA clones of human erythroleukemic (HEL) cell mRNA, a cell line exhibiting megakaryocytic-like properties. The cell line CHRF-288-11 also exhibits megakaryocytic-like properties, but synthesizes two related GP Ib β mRNA species of 3.5 and 1.0 kb. We performed cDNA cloning experiments to identify the origin of the 3.5-kb transcript and determine its relationship to the 1.0-kb GP Ib β mRNA found in megakaryocytes, platelets, and HEL cells. Our cloning experiments demonstrate that the longer transcript results from a nonconsensus polyadenylation recognition sequence, 5' AACAAAT^{3'}, within a separate gene located upstream to the platelet GP Ib β gene. In the absence of normal polyadenylation the more 5' gene uses the polyadenylation site within its 3' neighbor, the platelet GP Ib β gene. This newly identified 5' gene contains an open reading frame encoding 369 amino acids with a high degree of sequence similarity to an expanding family of GTP-binding proteins. (*J. Clin. Invest.* 1997; 99:520-525.) Key words: gene expression regulation • cloning, molecular • blood platelets • platelet membrane glycoproteins • cell division cycle proteins

Introduction

The GP Ib-IX-V complex is an essential platelet membrane receptor supporting platelet adhesion through its interaction with von Willebrand factor, an adhesive ligand of the blood and subendothelium (1, 2). The expression of the GP Ib-IX-V complex is dependent upon the coordinated assembly of at least three gene products, the α - and β -subunits of GP Ib, and GP IX (3). cDNA cloning experiments for these three subunits of the complex have utilized megakaryocytic-like cell lines and without exception the reported coding sequences have been representative of the mRNA species present in the megakaryo-

cyte and platelet (4-6). However, the megakaryocytic cell line CHRF-288-11 (7) is an exception containing two related mRNA species from the locus encoding the β -subunit of GP Ib (GP Ib β) (8). A short 1.0-kb transcript is consistent with the size of the megakaryocytic mRNA encoding platelet GP Ib β , whereas a longer transcript approximating 3.5 kb appears to be present in a wide variety of cell types but noticeably absent in megakaryocytes and other megakaryocytic-like cell lines (8, 9).

Kelly et al. reported the cloning of a cDNA hybridizing to the platelet GP Ib β cDNA and suggested the 3.5-kb transcript expressed by endothelial cells possesses the potential to synthesize a novel 45-kD protein containing a carboxyl terminus of platelet GP Ib β polypeptide sequence (8). However, a number of fundamental concerns were raised by their report. First, the cDNA sequence lacked a 3' polyadenylated sequence and contained \sim 600 base pairs of sequence 3' to the polyadenylation recognition signal utilized by megakaryocytes. It was unclear why endothelial cells would fail to recognize a consensus polyadenylation signal sequence unless the reported sequence was derived from a pre-mRNA species. A second concern was the presence of the single GP Ib β intron sequence within the proposed open reading frame of the related endothelial cell mRNA. This situation suggested an intron of a gene expressed by megakaryocytes is a translated exon when expressed by endothelial cells, an unusual form of alternative processing.

We sought to investigate the origin of the longer transcript in CHRF-288-11 cells. Now we report a number of key differences from the sequence reported by Kelly et al. (8) and in the fundamental conclusions derived from their work. We show that a unique gene located 5' to the platelet GP Ib β gene utilizes two different polyadenylation sites. The more 5' polyadenylation site results from an imperfect polyadenylation recognition sequence, whereas the more 3' polyadenylation event results from a utilization of the consensus polyadenylation sequence within the platelet GP Ib β gene. The use of the platelet gene's polyadenylation sequence results in a 3.5-kb transcript containing an open reading frame encoding a polypeptide related to GTP-binding proteins and a downstream open reading frame encoding platelet GP Ib β polypeptide. The potential to express platelet GP Ib β by cells other than megakaryocytes and platelets is discussed along with the identity of a unique gene residing 5' to the platelet GP Ib β gene.

Methods

RNA isolation and northern transfer. CHRF-288-11 cells were obtained from Dr. Michael Lieberman (Children's Hospital Medical Center, Cincinnati, OH) and were cultured as described in the original description of the cell line (7). Total RNA was isolated by ultracentrifugation through cesium chloride cushions using an established procedure (10). Poly (A)⁺ RNA was prepared by affinity chromatography using oligo-dT cellulose columns purchased from Life Technologies, Inc. (Bethesda, MD).

Address correspondence to Jerry Ware, SBR8/The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037. Phone: 619-784-9939; FAX: 619-784-2030; E-mail: jware@scripps.edu

Received for publication 24 June 1996 and accepted in revised form 20 November 1996.

J. Clin. Invest.

© The American Society for Clinical Investigation, Inc.

0021-9738/97/02/0520/06 \$2.00

Volume 99, Number 3, February 1997, 520-525

Gel electrophoresis of poly (A)⁺ RNA was performed through denaturing formaldehyde gels prepared as described by Lehrach et al. (11). Northern transfer and hybridization were performed using standard methodologies (10). The hybridization probe for the Northern analyses illustrated in Figs. 1 and 4 was a full-length platelet GP Ib β cDNA fragment isolated from a human erythroleukemia cell cDNA library prepared by the authors. The coding sequence for platelet GP Ib β within this cDNA fragment is identical to that described by Lopez et al. (5). After hybridization, the nitrocellulose filter was washed three times in 0.3 M sodium chloride/0.03 M sodium citrate/0.1% SDS (10 min, RT) and 1 time in 0.15 M sodium chloride/0.015 M sodium citrate/0.1% SDS (30 min, 55°C). Less stringent washing conditions resulted in a relatively high background due to crosshybridization with 28S rRNA. After washing the nitrocellulose was analyzed by autoradiography using Kodak X-OMAT film.

cDNA library construction. CHRF-288-11 poly (A)⁺ RNA was converted into cDNA and cloned into the bacteriophage λ ZAP/Express available from Stratagene Cloning Systems (La Jolla, CA). First-strand cDNA was synthesized using Stratascript RNase H-deficient reverse transcriptase (Stratagene) primed by poly(dT). The unamplified cDNA library consisted of 1.5×10^6 plaque-forming units. Random analysis of 20 independent clones revealed cDNA inserts in 18 clones ranging in size from 0.8-kb pairs to ~ 4 kb. 1×10^6 plaque-forming units were expanded into a high-titer stock of the CHRF-288-11 cDNA library.

Two additional cDNA libraries were also used in this project. The first was a library prepared by the authors from poly(A)⁺ mRNA isolated from human endothelial cells. These cells were a gift from Dr. Eugene Levin (The Scripps Research Institute, La Jolla, CA) and were third-passage cells prepared from a donor umbilical vein. This library was also constructed in the λ ZAP/Express vector and contained more than 1×10^6 recombinant bacteriophage. An additional fetal brain cDNA library was purchased from Stratagene.

Library screening and clone analysis. Screening of the library was performed using standard methodologies. A fragment of platelet GP Ib β cDNA was radiolabeled using a random priming system available from Stratagene. Positive clones were plaque-purified via four rounds of screening before performing analyses on the cloned cDNA fragments. Each cDNA insert was rescued from its recombinant λ ZAP/Express vector via the *in vivo* excision protocol provided by Stratagene. DNA sequence analysis was performed on single-stranded DNA templates from the rescued filamentous phagemids using dideoxynucleotides and Sequenase (U.S. Biochemical, Cleveland, OH; 12). All DNA sequence reported in this manuscript was determined by analysis of both strands and from multiple runs of individual sequencing reactions. The majority of sequence was generated using sequence-specific primers synthesized after each individual DNA sequencing gel. In a few cases some confirming sequence data was obtained after subcloning individual restriction fragments in the filamentous phage, M13mp18 or M13mp19, and sequencing using standard dideoxy methodologies (12).

Computer analysis of the DNA sequence was performed using the PC-based software, Lasergene (DNASTAR Inc., Madison, WI) and sequence comparison programs available through the Internet from the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>).

Results

Origin of the 3.5-kb platelet GP Ib β -related transcript expressed by CHRF-288-11 cells. CHRF-288-11 cells were established from a metastatic tumor in an infant with acute megakaryoblastic leukemia (7). These cells contain two mRNA species hybridizing with a full-length (1.0 kb) platelet GP Ib β cDNA probe (Fig. 1). The smallest transcript is ~ 1.0 kb and most likely corresponds to the mature GP Ib β mRNA transcript seen in other megakaryocytic cell lines, megakaryocytes, and

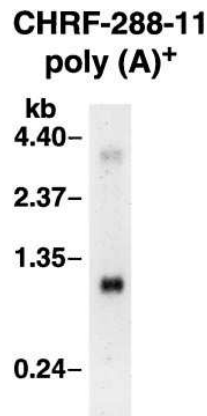


Figure 1. Northern analysis of CHRF-288-11 poly(A)⁺ RNA. Poly(A)⁺-selected RNA was purified from the megakaryocytic cell line, CHRF-288-11 (7), and electrophoresed through a 1.2% formaldehyde-denaturing gel. After electrophoresis the RNA was transferred to nitrocellulose and hybridized to a full-length platelet GP Ib β cDNA probe. The autoradiograph after hybridization and washing documents two hybridizing mRNA species of 3.5 and 1.0 kb. The observed 1.0 kb transcript is consistent with the size of GP Ib β mRNA in megakaryocytes and platelets. The migrating position of RNA molecular weight standards is shown to the left.

platelets (5, 9). A larger, less abundant species is ~ 3.5 kb and is not an expressed transcript seen in megakaryocytic RNA (8). To investigate the origin of the 3.5-kb transcript we prepared a CHRF-288-11 cDNA library and screened the library with the full-length platelet GP Ib β cDNA. Characterization of 17 hybridizing clones revealed that 14 contained cDNA fragments ranging from 0.9 to 1.0 kb corresponding to near full-length copies of the platelet GP Ib β mRNA. Of the remaining three clones, two contained inserts ~ 1.9 kb and another was ~ 3.2 kb. The clone containing the longest insert, bacteriophage λ 6C, was chosen for further DNA sequence analysis.

DNA sequence analysis of λ 6C revealed a cDNA insert of 3,172 nucleotides (Fig. 2). The 3' portion of the insert, nucleotides 1,866–3,172 corresponds to the GP Ib β genomic gene sequence as reported by Yagi et al. (9), yet lacking the single GP Ib β intron. Within this cDNA sequence is the major megakaryocytic transcription start site of the platelet GP Ib β gene (nucleotide 2,218; 9), the complete coding sequence for the GP Ib β precursor protein (nucleotides 2,245–2,862), the platelet GP Ib β polyadenylation recognition sequence (nucleotides 3,153–3,158) and a polyadenylated tail (Fig. 2).

The more 5' sequence within λ 6C contained an open reading frame encoding a 385 residue polypeptide sequence and represented the only significant open reading frame within the more 5' sequence. Thus, this fragment contained an unusual arrangement of cDNA sequence beginning with a unique 5' sequence contiguous with the platelet GP Ib β gene lacking its single intron. The possibility that the unique 5' sequence might be a cloning artifact resulting from the analysis of a single clone was dismissed with the characterization of the two remaining clones purified during the initial screening of the same library, λ 11C, and λ 17C. Each of these clones contains varying lengths of the same cDNA sequence beginning at nucleotides 1,224 and 1,521 of λ 6C, respectively. Clone λ 17C contained an additional uncommon feature, the complete intron sequence of the platelet GP Ib β gene, suggesting that in the context of the longer transcript, the single intron within the platelet GP Ib β gene was not efficiently removed.

A gene residing 5' to the platelet glycoprotein Ib β gene utilizes two alternative polyadenylation recognition sequences. To obtain additional information on the newly identified 5' sequence, the CHRF-288-11 cDNA library, a fetal brain cDNA library and a human endothelial cell cDNA library were each

118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

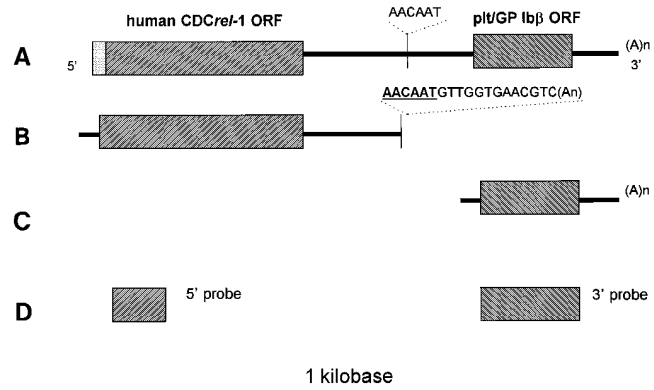


Figure 3. Diagrammatic representation of the GP Ib β -related transcripts. Three GP Ib β -related transcripts were identified by cDNA cloning experiments. Each of the three transcripts is schematically illustrated as they align to each other. (A) The 3.5-kb GP Ib β -related transcript expressed by CHRF-288-11 cells contains two open reading frames. Based on the characterization of the cDNA clones λ 12F and λ 8EN (see Fig. 5) the 5' open reading frame has been designated hCDCrel-1. (B) The isolation of multiple clones containing a polyadenylated tail after an imperfect polyadenylation sequence (*underlined*) suggests the longer transcript depicted in (a) is a result of the utilization of the consensus polyadenylation recognition sequence within the downstream platelet GP Ib β gene. The transcript depicted in (B) is \sim 2.3 kb and was identified by Northern analysis (see Fig. 4) and by the characterization of cDNA clones (Results and see Fig. 5). (C) The 1.0-kb platelet GP Ib β transcript is expressed by megakaryocytes and megakaryocytic-like cells lines and is shown as it aligns within the longer transcript. (D) Two radiolabeled probes representing 5' sequence (5' probe) and 3' sequence (3' probe) were used to confirm the presence of the three transcripts (A–C) by Northern analysis (Fig. 4). The location of each probe depicts its position relative to each transcript.

Figure 2. Nucleotide sequence of the platelet GP Ib β related-transcript of CHRF-288-11 cells. The complete DNA sequence of a bacteriophage λ clone designated, λ 6C, is shown and contains two long open reading frames (ORFs). The first ORF, nucleotides 2–1,156, encodes 385 amino acids sharing homology with a family of GTP-binding proteins. The second ORF, nucleotides 2,245–2,862, is the complete ORF encoding the platelet GP Ib β precursor polypeptide. A poly (A)⁺ tail resulting from the polyadenylation signal sequence in the platelet GP Ib β gene (*underlined*, nucleotides 3,153–3,158) represents the most 3' sequence. The transcription initiation site of the platelet GP Ib β precursor transcript is highlighted by a double arrow (\Uparrow) at nucleotide 2,218 (9). Thus, the longer transcript expressed by CHRF-288-11 cells contains a unique 5' sequence adjoining the platelet GP Ib β gene. An imperfect polyadenylation site within the more 5' sequence is double-underlined (nucleotides 1,976–1,981). Accession number U59632 has been assigned by GenBank.

screened with a 5' fragment of λ 6C. The fetal brain library was chosen based on the identification of more than 50 near-identical matches of brain-expressed cDNA fragments within the expressed sequence tag (EST)¹ database (13). None of the fragments identified in the EST database contained more 5' cDNA sequence than λ 6C. The endothelial cDNA library was chosen based on the previous identification of a longer (\sim 3.5

kb) transcript within endothelial cells hybridizing with the platelet GP Ib β cDNA (8). This screening yielded 8 more unique cDNA clones.

Several of the characterized clones contained a polyadenylated tail in which the last nucleotide preceding the polyadenylated tail corresponded to nucleotide 1,994 within the sequence presented in Fig. 2. Since the 3' boundary of these new clones lacked adenines, these clones were not the result of oligo d(T) mispriming in the first-strand reaction, and most likely represented cDNA products from an imperfect polyadenylation sequence (5'AACAAT^{3'}) 13 nucleotides upstream from the 3' end of each clone (nucleotides 1,976–1,981 of Fig. 2). Fig. 3 conceptually illustrates the origin of these clones as they align with the derived sequence from individual cDNA clones.

The utilization of this imperfect polyadenylation site suggested a related mRNA transcript of \sim 2.3 kb would be present in the poly (A)⁺ mRNA from CHRF-288-11 cells. Indeed, Northern analysis identified a 2.3-kb transcript utilizing the 5' probe demonstrating that post-transcriptional modification of the primary transcript for the more 5' gene involves two different polyadenylation recognition sequences, its own imperfect site or the polyadenylation recognition site within its 3' neighbor, the platelet glycoprotein Ib β gene (Figs. 3 and 4). Utilizing a 3' probe transcripts of 3.5 and 1.0 kb were observed, each presumably the product of different gene promot-

1. Abbreviation used in this paper: EST, expressed sequence tag.

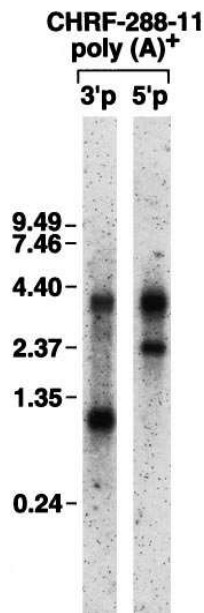


Figure 4. Northern analysis illustrates the alternative usage of two polyadenylation sequences. The utilization of an imperfect polyadenylation site is confirmed by Northern analysis of CHRF-288-11 poly(A)⁺ RNA and probes from the 5' region or 3' region of the longer transcript (see Fig. 3 d). As described in Fig. 1, the poly(A)⁺ RNA was electrophoresed through a denaturing formaldehyde gel, transferred to nitrocellulose, and hybridized with a probe representing with the more 5' sequence, or the 3' platelet GP Ib β probe. As predicted from the analysis of multiple bacteriophage λ clones, the more 5' probe hybridizes with 3.5- and 2.3-kb species, resulting from the alternative usage of two different polyadenylation recognition sequences. The more 3' probe hybridizes with 3.5- and 1.0-kb species resulting from polyadenylation via the recognition sequence within the platelet GP Ib β gene. A photograph of the autoradiograph is shown along with the migrating positions of RNA molecular weight standards.

ers, yet utilizing the same polyadenylation recognition sequence (Figs. 3 and 4).

Sequence analysis of transcripts polyadenylated from the imperfect recognition sequence. A common cDNA sequence representing the 2.3-kb transcript is presented in Fig. 5. The presented sequence is a composite of two cDNA clones, designated λ 12F and λ 8EN, isolated from fetal brain and endothelial cell cDNA libraries, respectively. Clone λ 8EN contained the sequence from nucleotide 83 to 2,032 (Fig. 5). Nucleotide sequence 1–82 was derived from clone λ 12F. The sequence contains an open reading frame encoding 369 amino acids with an initiating methionine codon (nucleotide 88–90) immediately preceded by a consensus Kozak sequence ($5'$ CCACC $^{3'}$; 14). Thus, unlike the incomplete sequence derived from λ 6C this sequence contains the translational machinery to synthesize the conceptual translation product presented in Fig. 5. The most 5' sequence illustrated in Fig. 5 differs from the 5' sequence of λ 6C (Fig. 3). Specifically, the sequences are identical in the 3' direction starting at nucleotide 104 of the λ 6C sequence (Fig. 2) and nucleotide 142 of the composite sequence presented in Fig. 5. Concerns that the more 5' sequence of λ 6C (nucleotides 1–103 of Fig. 2) might be a cloning artifact were dismissed by two additional clones containing 5' ends spanning nucleotides 1–103 of λ 6C. Thus, the presented cDNA sequences most likely represent two different processing variants of the same gene.

Discussion

Our results provide evidence that a gene residing in close 5' proximity to the platelet GP Ib β gene directly influences the expression of the platelet GP Ib β gene. The utilization of the platelet GP Ib β polyadenylation sequence by the more 5' gene results from an imperfect polyadenylation recognition sequence ($5'$ ACAAT $^{3'}$) in the more 5' gene. Indeed, in vitro studies have confirmed that mutation of a consensus polyadenylation sequence ($5'$ AATAAA $^{3'}$ \rightarrow $5'$ AACAAA $^{3'}$) dramati-

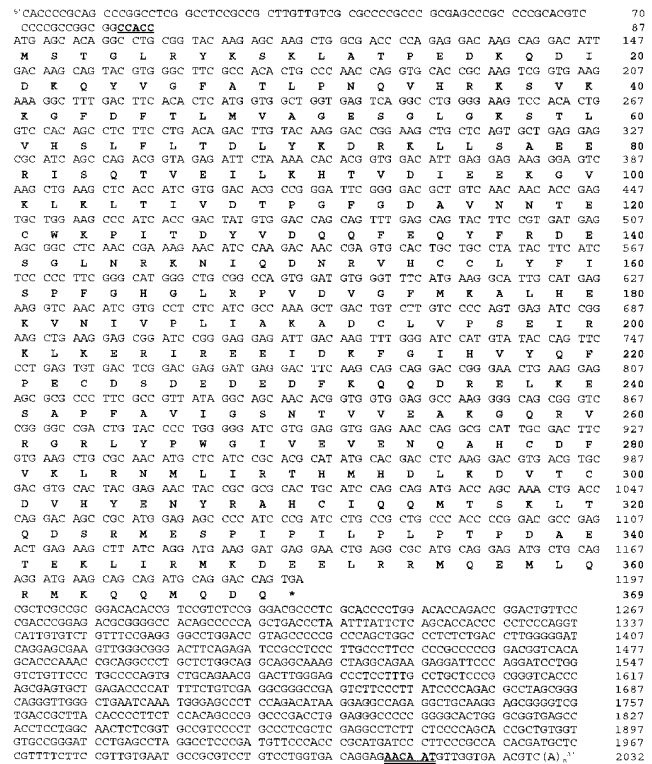


Figure 5. Sequence analysis of an mRNA transcript originating from an imperfect polyadenylation signal sequence. Bacteriophage λ clones, designated λ 12F and λ 8EN, were isolated from fetal brain and endothelial cell cDNA libraries. The sequence is schematically depicted as transcript *b* in Fig. 3. The nucleotide sequence contains an open reading frame encoding 369 amino acids with an initiating AUG codon preceded by a consensus Kozak sequence (underlined). The conceptual translation product has been designated, hCDCrel-1. An imperfect polyadenylation recognition sequence is the hallmark of this gene and is double underlined. Accession number U74628 has been assigned by GenBank.

cally reduces the efficiency of polyadenylation (15, 16). In the present study, an imperfect polyadenylation sequence results in the alternative usage of a polyadenylation recognition site within a nearby 3' gene. The net result is a longer transcript, approximating 3.5 kb, and containing two open reading frames (Fig. 3).

In the course of our analysis, we isolated multiple clones from the CHRF-288-11, endothelial cell and fetal brain cDNA libraries. In addition, we isolated and characterized more than 40 clones identified from the EST database and available through the cDNA consortium. It is clear from this analysis that the 5' sequence varies with a number of alternatively processed variants. In fact, a shorter exposure of the Northern blots presented in Figs. 1 and 4 suggests the 3.5 kb mRNA transcript is composed of several transcripts of slightly varying length. The sequences presented in Figs. 2 and 5 represent two of the more common processing variants identified. Whether these processing variants are specific for a particular cell type was not obvious from the characterization of our clones. Additional length heterogeneity within the longer transcript was provided by our isolation of a single cDNA clone containing the complete intron within the GP Ib β gene (data not shown). The presence or absence of the GP Ib β intron would produce a 274 ribonucleotide length difference among the mature

transcripts. Future studies characterizing genomic clones of this locus may aid in clarifying the identity of each processing variant.

A search of the current GenBank database with the polypeptide sequence derived from λ 8EN (Fig. 5) suggests the encoded polypeptide would be an additional member of a growing family of proteins related to the cell division cycle (CDC) proteins of yeast (17). We have designated this polypeptide as human cell division cycle-related (hCDCrel) protein 1, anticipating additional variants and related human sequences will be identified. Homologous proteins have been found in mouse and *Drosophila* and an alignment with the deduced amino acid sequence of hCDCrel-1 is presented (18–20; Fig. 6). Experimental evidence for the *Drosophila* homologue confirms its role in cytokinesis suggesting that a unifying functional attribute of these proteins may be a role in some aspect of cell division (20, 21). Structurally, the unifying feature of these proteins is the presence of GTP- or phosphate-binding motifs (22, 23; Fig. 6). Whether the 5' gene and the platelet GP Ib β gene are related remains unknown. The imperfect polyadenylation sequence within the more 5' gene is less than 250 nucleotides upstream of the megakaryocytic transcription initiation site within the platelet GP Ib β gene (Fig. 2). Indeed, in our own preliminary studies a fragment extending 262 base pairs upstream of the platelet glycoprotein Ib β transcription initiation site can support reporter gene expression in human erythroleukemia (HEL) cells suggesting it may represent a megakaryocytic promoter sequence.

Our analysis of λ 6C (Fig. 2) suggests some unidentified 5' sequence approximating 300 nucleotides remains to be identified based on a mature transcript size of 3.5 kb. This analysis will be necessary to determine if the longer transcript contains the potential to encode a polypeptide differing at the amino terminus from hCDCrel-1. Certainly, a 3.5-kb transcript containing the potential to code for two proteins would be unusual. Polycistronic units and internal ribosome binding sites in prokaryotes are common, but the translation of eukaryotic mRNA is most commonly initiated by ribosome binding to a 5' terminal cap structure and scanning for an initiation codon (24). According to this mechanism, we might predict for the 3.5-kb transcript illustrated in Fig. 3 efficient protein synthesis

from the more 5' coding sequence, and likewise, no protein synthesis from the more 3' open reading frame encoding platelet glycoprotein Ib β . However, the initiation of translation in eukaryotic cells via a cap-independent method, or via an internal ribosome entry site, is an efficient process for the synthesis of viral proteins dependent upon cellular not viral gene products (25–27). Indeed, it has been suggested that eukaryotic internal ribosome binding sites may be more widespread than previously recognized but lacking, to date, the description of a consensus motif (27). The synthesis of platelet glycoprotein Ib β from the longer transcript may represent an example of internal ribosome binding and will be examined in future studies to determine if synthesis of this protein can occur via a cap-independent mechanism.

Kelly et al. reported the cDNA cloning of an endothelial cell-expressed form of the 3.5-kb GP Ib β transcript (8). However, their cloning did not identify the 3' or 5' boundaries of the endothelial cDNA fragments. In fact, their reported sequence contained \sim 600 nucleotides 3' to the platelet polyadenylation site and was a composite of three overlapping cDNA fragments. They concluded that the longer transcript potentially expresses a 45-kD protein that is essentially a fusion protein composed of a unique amino terminus coupled to the platelet GP Ib β polypeptide sequence. This conclusion was based on immunoreactivity of a 45-kD protein present in endothelial cell extracts and recognized by a polyclonal anti-GP Ib β antibody. Their open reading frame for this protein assumed translation through the single intron within the platelet GP Ib β gene (9). Unlike the reported sequence of Kelly et al., our derived sequence does not contain an open reading frame capable of synthesizing the 45-kD polypeptide due to a number of insertions and deletions differing from their reported sequence. Including the present study, three different reports contain DNA sequence through the regions encoding the potential 45-kD protein. A comparison of the reported genomic sequence of Yagi et al. (9) to the cDNA sequence reported by Kelly et al. (8) reveals four discrepancies. Using the sequence of Kelly et al. as a reference, the differences are an inserted cytosine base at position 675, the absence of a cytosine/guanine dinucleotide pair at position 844 and a silent transversion at position 1,275. Three of these differences are critical since they

human CDCrel-1	MSTGL-----RY-----KSKLATPE-----	16
murine H5	MDHSLGWQGSVPEDGTGAGIKHFLE-----DSSDDAELSKFVKDFPGSEPYHSAESKTRVARPQILEFRPQSPDLCDDDVEFRGSLWFQP	86
murine D6	MSXQ-----	5
Dros. pnut	MNSFRSNVAGSGSGAISALPSTLAQLALRLKQQAASASASATNGSSGSESLVGVGRRFPNPQPSV---PVAASGKL---DTSSGGASNGDS	87
human CDCrel-1	-----DKQIDKQYGFATLPNQVHRKSVKKGDFDTLMVAGSGLGKSTLVHSLFLTDLYKDRKLLSA	78
murine H5	SDSQQYSPAPLSPSSRRSPWGLDIPYDSSDDKVEYGFATLPNQVHRKSVKKGDFDTLMVAGSGLGKSTLVHSLFLTDLYKDRKLLSA	270
murine D6	-QPTQFINPETP-----GYVGFANLPNQVHRKSVKKGDFDTLMVAGSGLGKSTLVHSLFLTDLYKDRKLLSA	71
Dros. pnut	NKLTLDLQEKSHQQAQKPKPF---LPVRQKPMIAGYVGFANLPNQVHRKSVKKGDFDTLMVAGSGLGKSTLVHSLFLTDLYKDRKLLSA	176
human CDCrel-1	EERISQTVELIKHTVDSIEKGVKLKLTIVDTFGFGDAVNNTCEWKPITDVTVDQFQYFRDESSGLNRKNIQNRVHCCLYFISFPFGHGLRPV	170
murine H5	EERIMQTVELIKHTVDSIEKGVKLKLTIVDTFGFGDAVNNTCEWKPITDVTVDQFQYFRDESSGLNRKNIQNRVHCCLYFISFPFGHGLRPV	270
murine D6	AEKIERVQIEASTVEIERGVKLKLTIVDTFGFGDAVNNTCEWKPITDVTVDQFQYFRDESSGLNRKNIQNRVHCCLYFISFPFGHGLRPV	163
Dros. pnut	SLRKKKTVAVEATKVMKENGVLTLTVDTFGFGDAVNNTCEWKPITDVTVDQFQYFRDESSGLNRKNIQNRVHCCLYFISFPFGHGLRPV	268
human CDCrel-1	DVGFMKALHEKVNIVPLIAKADCLVPSEIRKLKERIREEDIKFGIHVYQFPECDSDEDEDFKQDRELKESAPFAVIGSNTVVEAKGQVRVG	262
murine H5	DVEFMKALHQRVNIVPLIAKADCLVPSEIRKLKERIREEDIKFGIHVYQFPECDSDEDEDFKQDRELKESAPFAVIGSNTVVEAKGQVRVG	362
murine D6	DVAFMKAIHNKVNIVPLIAKADCLVPSEIRKLKERIREEDIKFGIHVYQFPECDSDEDEDFKQDRELKESAPFAVIGSNTVVEAKGQVRVG	255
Dros. pnut	DIACMQSLSKVNIVPLIAKADCLVPSEIRKLKERIREEDIKFGIHVYQFPECDSDEDEDFKQDRELKESAPFAVIGSNTVVEAKGQVRVG	359
human CDCrel-1	RLYPWGIIVEVNPQHCDPVKLRNMLIRTHMDLKDVTCDVHYENYRAHCICQMT-----SKLTQDS---RMESPIP-----IL-	333
murine H5	RLYPWGIIVEVNPQHCDPVKLRNMLIRTHMDLKDVTCDVHYENYRAHCICQMTRLVVK---RNRNKLRES---GTFPIP-----AV-	442
murine D6	RLYPWGIIVEVNPQHCDPVKLRNMLIRTHMDLKDVTCDVHYENYRAHCICQMTRLVVK---RNRNKLRES---GTFPIP-----AV-	346
Dros. pnut	RRYPWGLVEVENLTHCDPIALNNMIVIRHLQDLKDVNTNVHYENYRAHCICQMTRLVVK---RNRNKLRES---GTFPIP-----AV-	451
human CDCrel-1	-----PLPTPDASTTEKL-----IRKDEBELRR-----MQEMLQRMK-----QQMDQ-Q	369
murine H5	-----P-PGTFDETEKL-----IREKDEBELRR-----MQEMLQRMK-----QQMDQ-Q	478
murine D6	-----QMGGGDDGGALGHTCKVMCTYKSKSLTW-----IKKTKFQMHGPAVCFQSLGGCHPHFNSTCA	406
Dros. pnut	FDMKVKEKMQKLKRDSELELARRHEERKKALEQLIRELEKRREREREKKEWEDVNHVTLSELEKRRSLGANSSTQDVNDGKKKKKKGLF	539

Figure 6. Comparison of the conceptual translation product, hCDCrel-1, to similar GenBank entries. The polypeptide, designated hCDCrel-1, shares homology with a family of GTP-binding proteins. Two murine proteins, H5 and D6 (GenBank accession numbers X61452 and D28540, respectively) and a *Drosophila* peanut homologue (accession number U08103) are aligned illustrating the conservation of sequence through more than 300 amino acids of each sequence. hCDCrel-1 is 75, 56, and 46% identical to H5, D6, and peanut polypeptides, respectively. Each of the aligned proteins contains motifs involved in GDP/GTP exchange (GX₂GK[S/T]) or in GTP hydrolysis (DX₂G), residues 51–59 and 108–111 of hCDCrel-1.

result in frame-shifts within the proposed open reading frame encoding the 45-kD polypeptide (8). Our sequence analysis agrees with that of Yagi et al. at these positions further supporting the conclusion that the synthesis of this 45-kD polypeptide is not possible. Thus, the immunoreactivity of a 45-kD endothelial cell protein most likely represents cross-reacting material distinct from GP Ib β peptide. However, based on this previous study there are \sim 50 EST database entries containing gene sequence identical to that presented in Figs. 2 and 5, yet designated platelet GP Ib β precursor polypeptide. Our data would support a renaming of these sequences to hCDCrel-1 protein.

The current studies provide two directions for future research. First, we have identified a new gene within the 22q11.2 chromosomal locus. Initiatives to identify novel genes from this locus have been fueled by studies linking DiGeorge syndrome and other cardio-facial abnormalities to this region of chromosome 22 (28, 29). The close relationship of DiGeorge syndrome and the platelet GP Ib β gene has been verified by chromosomal deletion of 22q11.2 in an individual diagnosed with both the Bernard-Soulier syndrome (congenital absence of the platelet GP Ib-IX receptor) and the DiGeorge syndrome (30). Indeed, the sequence similarities between the *Drosophila* peanut protein and hCDCrel-1 suggests a function for hCDCrel-1 in development since the expression of the peanut protein is linked to proliferation of the imaginal disc tissues of *Drosophila* (20). Thus, the identity of the 5' gene and its product should be characterized as another candidate gene for normal development. Finally, the identification of platelet GP Ib β coding sequence in cells of nonhematopoietic origin suggests the presence of this protein in cells other than megakaryocytes and platelets. Thus, a previously unrecognized functional consequence for the expression of this protein may exist and should be examined.

Acknowledgments

The authors thank Dr. Mayumi Yagi (University of Washington, Seattle, WA) for sharing unpublished genomic DNA sequence flanking the platelet GP Ib β gene. The authors are also grateful to the Sam and Rose Stein Trust Fund for support establishing the DNA Core Facility within the Department of Molecular and Experimental Medicine at The Scripps Research Institute.

This work was supported in part by grant HL50545 from the National Heart, Lung, and Blood Institute of the National Institutes of Health. B. Zieger is a recipient of a DFG Scholarship from the Deutsche Forschungsgemeinschaft. J. Ware is a recipient of an Established Investigator Award from the American Heart Association.

References

1. Ruggeri, Z.M. 1991. The Platelet Glycoprotein Ib-IX Complex. In *Progress in Hemostasis and Thrombosis*. B.S. Collier, editor. W.B. Saunders Company, Philadelphia. 35-68.
2. Lopez, J.A. 1994. The platelet glycoprotein Ib-IX complex. *Blood Coagul. Fibrinolysis*. 5:97-119.
3. Lopez, J.A., B. Leung, C.C. Reynolds, C.Q. Li, and J.E.B. Fox. 1992. Efficient plasma membrane expression of a functional platelet glycoprotein Ib-IX complex requires the presence of its three subunits. *J. Biol. Chem.* 267:12851-12859.
4. Lopez, J.A., D.W. Chung, K. Fujikawa, F.S. Hagen, T. Papayannopoulou, and G.J. Roth. 1987. Cloning of the alpha chain of human platelet gly-

coprotein Ib: A transmembrane protein with homology to leucine-rich alpha2-glycoprotein. *Proc. Natl. Acad. Sci. USA*. 84:5615-5619.

5. Lopez, J.A., D.W. Chung, K. Fujikawa, F.S. Hagen, E.W. Davie, and G.J. Roth. 1988. The alpha and beta chains of human platelet glycoprotein Ib are both transmembrane proteins containing a leucine-rich amino acid sequence. *Proc. Natl. Acad. Sci. USA*. 85:2135-2139.

6. Hickey, M.J., L.L. Deaven, and G.J. Roth. 1990. Human platelet glycoprotein IX: Characterization of cDNA and localization of the gene to chromosome 3. *FEBS Lett.* 274:189-192.

7. Fugman, D.A., D.P. Witte, C.L. Jones, B.J. Aronow, and M.A. Lieberman. 1990. In vitro establishment and characterization of a human megakaryoblastic cell line. *Blood*. 75:1252-1261.

8. Kelly, M.D., D.W. Essex, S.S. Shapiro, F.J. Meloni, T. Druck, K. Huebner, and B.A. Konkle. 1994. Complementary DNA cloning of the alternatively expressed endothelial cell glycoprotein Ib β (GPIb β) and localization of the GPIb β gene to chromosome 22. *J. Clin. Invest.* 93:2417-2424.

9. Yagi, M., S. Edelhoff, C.M. Distech, and G.J. Roth. 1994. Structural characterization and chromosomal location of the gene encoding human platelet glycoprotein Ib β . *J. Biol. Chem.* 269:17424-17427.

10. Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

11. Lehrach, H., D. Diamond, J.M. Wozney, and H. Boedtker. 1977. RNA molecular weight determination by gel electrophoresis under denaturing conditions, a critical reexamination. *Biochemistry*. 16:4743-4751.

12. Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*. 74:5463-5467.

13. Adams, M.D., J.M. Kelley, J.D. Gocayne, M. Dubnick, M.H. Polymeropoulos, H. Xiao, C.R. Merril, A. Wu, B. Olde, et al. 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science (Wash. DC)*. 252:1651-1656.

14. Kozak, M. 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res.* 12:857-872.

15. Fitzgerald, M., and T. Shenk. 1981. The sequence 5'-AAUAAA-3' forms parts of the recognition site for polyadenylation of late SV40 mRNAs. *Cell*. 24:251-260.

16. Sheets, M.D., S.C. Ogg, and M.P. Wickens. 1990. Point mutations in AAUAAA and poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Res.* 18:5799-5805.

17. Hartwell, L.H., J. Culotti, J.R. Pringle, and B.J. Reid. 1974. Genetic control of the cell division cycles in yeast. *Science (Wash. DC)*. 183:46-51.

18. Kato, K. 1990. A collection of cDNA clones with specific expression patterns in mouse brain. *Eur. J. Neurosci.* 2:704-711.

19. Nottenburg, C., W.M. Gallatin, and T. St. John. 1990. Lymphocyte HEV adhesion variants differ in the expression of multiple gene sequences. *Gene*. 95:279-284.

20. Neufeld, T.P., and G.M. Rubin. 1994. The *Drosophila* peanut gene is required for cytokinesis and encodes a protein similar to yeast putative bud neck filament proteins. *Cell*. 77:371-379.

21. Hartwell, L.H. 1971. Genetic control of the cell division cycle in yeast. IV. Genes controlling bud emergence and cytokinesis. *Exp. Cell Res.* 69:265-276.

22. Saraste, M., P.R. Sibbald, and A. Wittinghofer. 1990. The P-loop—a common motif in ATP- and GTP-binding proteins. *Trends Biochem. Sci.* 15:430-434.

23. Bourne, H.R., D.A. Sanders, and F. McCormick. 1991. The GTPase superfamily: conserved structure and molecular mechanism. *Nature (Lond.)*. 349:117-127.

24. Kozak, M. 1989. The scanning model for translation: an update. *J. Cell Biol.* 108:229-241.

25. Macejak, D.G., and P. Sarnow. 1991. Internal initiation of translation mediated by the 5' leader of cellular mRNA. *Nature (Lond.)*. 353:90-94.

26. Jang, S.K., and E. Wimmer. 1990. Cap-independent translation of encephalomyocarditis virus RNA: structural elements of the internal ribosomal entry site and involvement of a cellular 57-kD RNA-binding protein. *Nature (Lond.)*. 353:90-94.

27. Mountford, P.S., and A.G. Smith. 1995. Internal ribosome entry sites and dicistronic RNAs in mammalian transgenesis. *Trends Genet.* 11:179-184.

28. Driscoll, D.A. 1994. Genetic basis of DiGeorge and velocardiofacial syndromes. *Curr. Opin. Pediatr.* 6:702-706.

29. Lindsay, E.A., R. Goldberg, V. Jurecic, B. Morrow, C. Carlson, R.S. Kucherlapati, R.J. Shprintzen, and A. Baldini. 1996. Velo-cardio-facial syndrome: frequency and extent of 22q11 deletions. *Am. J. Med. Genet.* 57:514-522.

30. Budarf, M.L., B.A. Konkle, L.B. Ludlow, D. Michaud, M. Li, D.J. Yamashiro, D. McDonald-McGinn, E.H. Zackal, and D.A. Driscoll. 1995. Identification of a patient with Bernard-Soulier syndrome and a deletion in the DiGeorge/Velo-cardio-facial chromosomal region in 22q11.2. *Hum. Mol. Genet.* 4:763-766.