Vol. 178, No. 1, 1991 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
July 15, 1991 Pages 38-44

The Human Liver and Reticulocyte Cytochrome b_5 mRNAs are Products from a Single Gene

Sara J. Giordano and Alan W. Steggles

Department of Biochemistry and Molecular Pathology
Northeastern Ohio Universities
College of Medicine
Rootstown, Ohio 44272

Received May 3, 1991

Using a combination of standard cDNA library screening techniques and the Polymerase Chain Reaction (PCR) we have isolated and sequenced DNA fragments corresponding to the human reticulocyte cytochrome b_5 mRNA. The reticulocyte specific sequence codes for amino acids 97 and 98 only, with a TAA stop codon. The reticulocyte specific 3'non-translated sequence has 15 new nucleotides then utilizes the liver mRNA sequence from amino acid 97 onward. This indicates that the reticulocyte specific exon has 24 base pairs (bp). In addition, we have isolated sequences that are derived from a transcribed cytochrome b_5 pseudogene. This transcript contains multiple mutations which prevent the synthesis of any functional protein.

Cytochrome b_5 (b_5) is a small amphipathic protein that exists in two forms. One with 98 amino acids (aa), counting the initial Met as 1, found in the reticulocyte, the other with 134 aa found in the liver and other tissues (1). The essential differences between the two forms are a) the extra 36 aa in the liver b_5 occur at the C terminal end and serve as a membrane binding domain, and b) the 98th aa for all species except bovine is different, i.e., aa 1-97 are always the same in the two forms. This has suggested that the two forms of b_5 could arise from the tissue specific post-transcriptional processing of one mRNA. It is possible that the human and rabbit, etc. (1) erythrocyte b_5's could be derived by post-translational processing of the liver b_5, but this would

0006-291X/91 $1.50

entail the addition of at least one new aa to the C-terminal end of the protein. This rare novel alteration has been reported for tubulin (2). However, for bovine post-translational processing of the liver b_5 to the reticulocyte b_5 is a real possibility.

Only one reticulocyte b_5 mRNA has been analyzed. In the chicken, the liver and the reticulocyte mRNA's have exactly the same coding sequence (3) differing only in the length of their 3' non-translated regions. There is also one report of a mitochondrial membrane specific b_5 which has only 58% amino acid identity with the liver form (4).

MATERIALS AND METHODS

The human reticulocyte λZAP cDNA library was a generous gift from Dr. J. Prchal (Birmingham, AL). Human total RNA was isolated from reticulocytes (5) obtained from blood taken from normal individuals. Blood was drawn following NEOUCOM IRB approved procedures.

A total of $\geq 1.4 \times 10^6$ plaques were screened using standard procedures (6) and Stratagene (La Jolla, CA) protocols. Fourteen plaques were purified corresponding to four independent clones containing inserts of 0.52, 0.54, 1.9, and 3.9 kb, respectively. The inserts were isolated in Bluescript KS- according to the manufacturer's instructions (Stratagene). DNA sequencing was carried out using the chain termination method with Sequenase (USB, Cleveland, OH) and α^{35}S dATP (NEN, Boston, MA).

Single stranded cDNA was prepared from reticulocyte total RNA (6) using random primers and MLTV (BRL, Gaithersburg, MD), and used for PCR without further purification. For the amplification of human reticulocyte cDNA, oligonucleotide primers A, B, and C derived from the liver and reticulocyte b_5 sequences (8 and Fig. 1) were obtained from National Biosciences (Hamel, MN) and used without further purification. The PCR amplification conditions were 30 cycles of 1.5 min at 94°, 1 min at 60°, 2.25 min at 72° with a final 7 min step at 72°. The initial PCR amplification used one tenth volume of the cDNA synthesis reaction. When samples were subjected to a second set of 60 cycles of amplification, they were purified through UltraFree columns (Millipore, Bedford, MA). PCR products were analyzed by 4% NuSieve agarose gel electrophoresis and selected bands of interest were subcloned into Bluescript KS- (7).

RESULTS AND DISCUSSION

From the screening of the human reticulocyte λZAP cDNA library, 4 independent recombinant phage were purified. The two

```
ATG GCA GAG CAG TCG GAC GAG GCC GTG AAG TAC TAC ACC   39
Met Ala Glu Gln Ser Asp Glu Ala Val Lys Tyr Tyr Thr   13

CTA GAG GAG ATT CAG AAG CAC AAC CAC AGC AAG AGC ACC   78
Leu Glu Glu Ile Gln Lys His Asn His Ser Lys Ser Thr   26

TGG CTG ATC CTG CAC CAC AAG GTG TAC GAT TTG ACC AAA  117
Trp Leu Ile Leu His His Lys Val Tyr Asp Leu Thr Lys   39

TTT CTG GAA GAG CAT CCT GGT GGG GAA GAA GTT TTA AGG  156
Phe Leu Glu Glu His Pro Gly Gly Glu Glu Val Leu Arg   52

GAA CAA GCT GGA GGT GAC GCT ACT GAG AAC TTT GAG GAT  195
Glu Gln Ala Gly Gly Asp Ala Thr Glu Asn Phe Glu Asp   65

GTC GGG CAC TCT ACA GAT GCC AGG GAA ATG TCC AAA ACA  234
Val Gly His Ser Thr Asp Ala Arg Glu Met Ser Lys Thr   78

TTC ATC ATT GGG GAG CTC CAT CCA GAT GAC AGA CCA AAG  273
Phe Ile Ile Gly Glu Leu His Pro Asp Asp Arg Pro Lys   91
                  G
TTA AAC AAG CCT CCA GAA CCT TAA AGGCGGTGTTTCAAGGAAA  316
Leu Asn Lys Pro Pro Glu Pro xxx                      98

CTCTTATCACTACTATTGATTCTAGTTCCAGTTGGTGGACCAACTGGGTGA  367

TCCCTGCCATCTCTGCAGTGGCCGTCGCCTTGATGTATCGCCTATACATGG  418

CAGAGGACTGAACACCTCCTCAGAAGTCAGCACAGGAAGAGCCTGCTTTGG  469

ACACGGGAGAAAAGAAGCCATTGCTAACTACTTCAACTGACAGAAACCTTC  520

ACTTGAAAACAATGATTTTAATATATCTCTTTCTTTTTCTTCCGACATTAG  571

AAACAAAACAAAAAGAACTGTCCTTTCTGCGCTCAAATTTTTCGAGTGTGC  622

CTTTTTATTCATCTACTTTATTTTGATGTTTCCTTAATGTGTAATTTACTT  674

ATTATAAGCATGATCTTTTAAAAATATATTTGGCTTTTAAAGTAAAAAAAA  725

AAAAAAAAAAAAAAA                                      740
```

Figure 1. The complete nucleotide sequence of human reticulocyte cytochrome b_5 cDNA. The reticulocyte specific sequence is underlined, the alternative polyA site triple underlined, and the polymorphism at nucleotide 288 indicated. The PCR primer A was AAAACAGCTGGCTCGCGGCGAACCG and based on nucleotides -1 to -19 from the human liver b_5 sequence (8). Primers B and C were AAAACAGCTCTTCCTGCGCTGACTTCT and CACCGCCTTTAAGGTC were complementary to nucleotides 459→439 and 306→289 of the reticulocyte b_5 sequence.

smallest (A,B; 524 and 540 bp, respectively) were initially chosen for analysis. DNA sequencing showed that clone A corresponded to liver b_5 from aa 86 to the 3' polyA tail but contained an additional 24 bp inserted between aa 96 and aa 97. This extra sequence was GAACCTTAAAGGCGGTGTTTCAAG corresponding to glu-pro-stop-non-translated (Fig. 1, single underline). The remainder of the liver b_5 sequence was now part of the 3' non-translated sequence for the reticulocyte mRNA. The 5' end of clone A was not a recognizable sequence and the break point between the 5' sequence

and the coding sequence at aa 86 did not resemble an intron-exon junction. Analysis of clone B showed that the 3' polyA tail had been attached at a new position (triple underline, Fig. 1). The new polyA adenylation site was different to the alternative site previously described for liver b_5 mRNA (9). Analysis of the 5' end of clone B again gave an unrecognizable sequence, but the coding information started at aa 72. The reticulocyte specific 24 bp sequence was present in the same relative position as clone A. Again the breakpoint at aa 72 did not resemble an intron-exon junction. The sequencing data indicated that the reticulocyte specific sequence is contained in one exon of 24 bp coding for only two amino acids. By analogy, one would expect rabbit and most other species to have a similarly organized reticulocyte b_5 cDNA, but bovine reticulocyte could still be the exception.

In order to isolate the rest of the human reticulocyte cDNA, we obtained oligonucleotide primers based on the 5' and 3' sequence of the liver b_5 cDNA (8). The PCR products using primers A and B were purified by passage through an UltraFree column. A 5 μl aliquot (ca 1/10) was used for a second amplification with primers A and C, and gave rise to an apparently homogenous product of ca 350 bp. Subcloning and sequencing of this product gave the nucleotide sequence corresponding to aa 1-98, and confirmed a polymorphism in two clones (P1 and P2) at codon 96. The overall organization of the reticulocyte specific b_5 cDNA was confirmed. Clone P1 is CCA and clone P2 is CCG. We do not think that this is a PCR generated artifact because we have found both sequences in the analysis of genomic DNA clones (data not shown). The CCA seems to be less common, because in the analysis of 6 human b_5 cDNAs (this paper, 8, 9, 10), 5 were CCG, only one CCA. If the previous exon ends at aa 96, then there would be a CCA/GT or a CCG/GT junction. The restriction enzyme HpaII recognizes CCGG, the presence or absence of this restriction site may be of some future value in the analysis of the b_5 gene in human genomic DNA.

The difference between our data and that obtained from the chicken is not surprising (3). Chicken erythrocytes are nucleated and possess endoplasmic reticulum as do hepatocytes. Therefore, the finding of a b_5 with a membrane binding domain, i.e., a liver type b_5, would not be unexpected.

Northern analysis of human reticulocyte RNA showed only one hybridizing signal of ca 720bp, suggesting that the individual studied had either little or no transcribed pseudogenes (data not shown).

Clones C and D (1900, 3900 bp) were clearly too long to represent the liver or reticulocyte forms of b_5, however, they could have been partially processed mRNA's (9). Both C and D had the same 5' terminal sequence as clone B (Fig. 2), but differed from clone A. However, both C and D had the same 3' sequence. In

```
                •   •                                  1   2   3   4   5
GGGGCTGTGTGAGCTGGGACTGGCTCGCAGCCCTCCGAG ATG GCC CAG CAG TCG
                                        * ****           * *

6   7   8   9   10  11  12  13  14  15  16  17  18  19  20
GAC AAG GCC GTG AAG TCG TGA GCC CTG GAG GAA ATT CA- AAG CAC
    *                 **  **  *   *

21  22  23  24  25  26  27  34  35  36  37  38  39  40  41
AAT CAT AGC AAA AGC ACC TGG GTG TAC GAT TTG ACC AAA TTT CTG
  *   *     *                                       *

42  43  44  45  46  47  48  49  50  51  52  53  54  55  56
GCA GAG CAT CCT GGT GGG AAA GAA GTC TTA AGG GAA TAA GCT GGA
  *                     *       *               *

57  58  59  60  61  62  63  64  65  66  67  68  69  70  71
GA- GAT GCT ACT GAA AAC TTT GAG GAT GTC GGG CAC TCT ACA GAT
    *               *

72  73  74  75  76  77  78  79  80  81  82  83  84  85  86
GCC AGA GAA TTA TTC AAA TCA TAA ATC ATT GGG GTT -T- CAT CTG
      *     * *  *     *   **                **          **

87  88  89  90  91  92  93  94  95  96  97  98  stop
GAT GAC AGA -TA AAG TTA ACT AAG CCT TCA GAA TTT TTT
                        **          * *     **  **

TTTTTTTTT
```

Figure 2. The partial nucleotide sequence of the reticulocyte cytochrome b_5 pseudogene transcript. Only the 5' and 3' end sequences and the hybridizing region are shown. The numbers 1-98 correspond to the reticulocyte b_5 codons (Fig. 1). * represents a single bp change, - indicates a 1 bp deletion, and • indicates a 1 bp insertion. The 18bp deletion occurs between codons 27 and 34.

order to analyze clone D in more detail, the hybridizing sequence was isolated as a HinfI fragment, subcloned and sequenced (Fig. 2). Analysis showed that the sequence represented that of a b_5 pseudogene spanning amino acids 1-98. In order to obtain the rest of the pseudogene sequence, two oligonucleotides, TTGGTGAAATCGTAC-ACC and TTTGAGGATGTCGGGCAC, were synthesized and used as sequencing primers. The data (Fig. 2) indicates that clone D, and clone C, represent transcripts derived from a reticulocyte b_5 pseudogene. The pseudogene transcript contains an 18bp deletion, 5 single bp deletions, 2 single bp insertions, plus 43 one bp changes. Overall, there is a ca 87% homology with the authentic reticulocyte b_5 cDNA sequence (Fig. 1).

The transcription of a pseudogene has been described (11,12) but it is a rare event. The sequence of this pseudogene is different from the three liver b_5 pseudogenes previously isolated (9,13). As yet, we do not know exactly how many b_5 pseudogenes are present in any individual's genome.

ACKNOWLEDGMENTS

We thank Dr. Prchal for the human reticulocyte library. This work was supported by funds from the American Heart Association, Ohio Affiliate, and the United Way of Stark County, Ohio. The accession number for the reticulocyte b_5 and the pseudogene sequences are M60174 and M64714, respectively.

REFERENCES

1. Kimura S, Abe K, and Sugita Y (1984) FEBS Letters, 169, 143-146.
2. Raybin D, and Flavin M (1977) Biochemistry, 16, 2189-2194.
3. Zhang H, and Somerville C (1990) Arch. Biochem. Biophys., 280, 412-415.
4. Lederer F, Ghrir R, Guiard B, Cartial S, and Ito A (1983) Eur. J. Biochem., 132, 95-102.
5. Chomczynski, P, and Sacchi, N (1987) Anal. Biochem., 162, 156-159.
6. Maniatis T, Fritsch EF, and Sambrook J (1982) in Molecular Cloning. A Laboratory Manual. Cold Spring Harbor.
7. Struhl K (1985) BioTechniques, 3, 452-453.
8. Yoo M, and Steggles AW (1988) Biochem. Biophys. Res. Commun., 156, 576-600.

9. Yoo M, and Steggles AW (1989) Biochem. Biophys Res. Commun., 163, 18-24.
10. Miyata M, Nagata K, Yamajoe Y, and Kata R (1989) Pharmacol. Res., 21, 513-519.
11. Sorge J, Gross E, West C, and Beutler E (1990) J. Clin. Investig., 86, 1137-1141.
12. McCarrey JR, and Thomas K (1987) Nature, 326, 501-505.
13. Yoo M, and Steggles AW In preparation.