Vol. 163, No. 1, 1989 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
August 30, 1989 Pages 18-24

The Characterization of Three Types of Partially Processed mRNA and Two Pseudogenes for Human Liver Cytochrome b_5

Min Yoo and Alan W. Steggles

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

Received July 4, 1989

SUMMARY. We have isolated cDNA clones corresponding to partially processed human liver cytochrome b_5 mRNAs. All the clones contained poly(A) sequences, and one clone had a shorter 3' non-translated sequence, indicating the use of an alternative poly(A) addition signal. In addition, all the clones contained the coding information for amino acids 87-134; however, there were two types of intron junction adjacent to the coding sequence. Detailed analysis of the Type I clones showed that the Type II intron sequence was contained within the Type I sequence, but approximately 1000 bp 5' of the Type I intron-exon junction showed alternative splicing within this intron. In addition, we have isolated two pseudogenes which lack introns, suggesting the retroviral insertion of human liver cytochrome b_5 mRNA sequences into the human genome.

Cytochrome b_5 (b_5) is involved in the stimulation of many cytochrome P-450 catalyzed liver reactions (1) and in the erythrocyte acts with cytochrome b_5 reductase to convert methemoglobin back to hemoglobin (2), thereby maintaining the normal physiological function of blood. There is also one reported instance of a form of inherited methemoglobinemia due to defective b_5 protein (3), although many other instances are thought to exist but either have not been fully reported or accurately diagnosed (4). Cytochrome b_5 is a small amphipathic protein that has either 134 or 98 amino acids in the liver or erythrocyte, respectively. In most species the first 97 amino acids are the same for both forms of the protein (5), however, in bovine all 98 are conserved (6). It is not clear whether there are two b_5 genes or whether there are two b_5 mRNAs derived from

0006-291X/89 $1.50

one gene by alternative splicing. In bovine there is also the possibility of post-translational processing. Recently, nucleotide sequences have been reported for bovine, chicken, human and rabbit liver b_5 mRNA's (7-10), and there is considerable sequence homology between them, although the chicken sequence showed five extra amino acids at the NH_2 and one less at the COOH termini, respectively (8). There is no data on the gene organization other than the rat b_5 gene is believed to contain at least five exons (11), and the chicken gene at least two exons (8). In rat there is an indication that there are two b_5 mRNAs with different lengths of 3' non-translated sequence (11).

In order to understand the exact relationship between the liver and erythrocyte b_5's, and to study the molecular lesions responsible for one form of methemoglobinemia, it is necessary to understand the exact organization of the b_5 gene(s).

MATERIALS AND METHODS. The human liver λgt11 cDNA libraries were either purchased from Clontech, Ltd. (Palo Alto, CA) or obtained as a gift from Dr. S.L.C. Woo (Baylor, TX). The human genomic DNA libraries were purchased from Clontech, or a gift from Dr. T. Maniatis (Harvard, MA) (12). The libraries were screened using standard procedures (7,9,10,13) with full-sized rabbit liver b_5 or human liver b_5 (Hb_5) cDNA's as probes (9,10). The probes were labelled with [^{32}P]dCTP to a specific activity of ca $1x10^9$cpm/μg using random hexamer priming (14). Restriction and DNA modification enzymes were purchased from Pharmacia (Piscataway, NJ). DNA fragments were subcloned into the pBluescript KS(-) vector (Stratagene, La Jolla, CA) and sequenced using oligonucleotide primers (15). In some instances, deletions were prepared using exonuclease III, and mung bean nuclease (16) in order to facilitate DNA sequencing.

RESULTS AND DISCUSSION. Eight different cDNA clones for partially processed mRNAs were isolated from the two cDNA libraries and all contained the coding information corresponding to amino acids 87-134 of Hb_5 (Fig. 1). They were put into two main groups based on DNA sequence analysis. The Type I group (five clones) all had a 3' non-translated sequence exactly the same as that of Hb_5 mRNA (9), but with variable length poly(A)

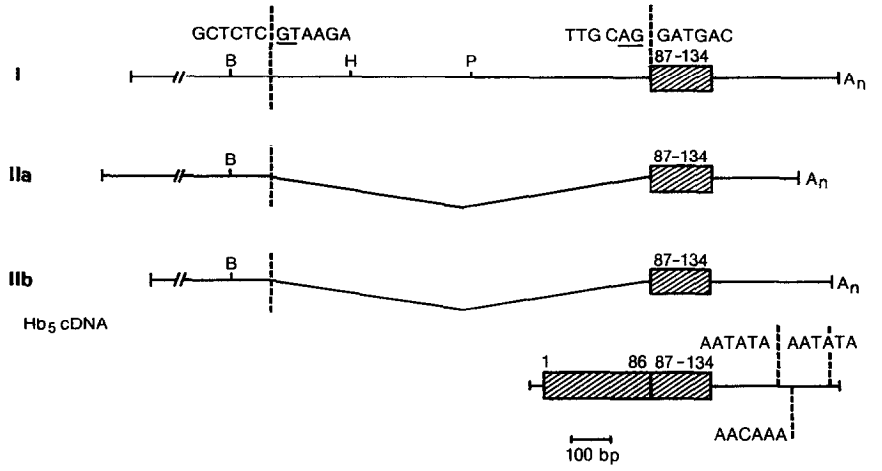

Figure 1. The comparison of three types of cDNA clones derived from partially processed Hb_5 mRNAs to full size Hb_5 cDNA. (I). In this group (5 clones) the intron sequence ranged from ca 1350 to ca 360 bp in length, the poly(A) tails (A_n) contained from 18 to ca 70 adenylic acid residues. B, H, P represent BamHI, Hind III and Pst I sites. (IIa). In this group (one clone), the intron sequence was ca 1250 bp, and the 3' non-translated sequence was only 220 bp long with a ca 70 bp poly(A) tail. (IIb). These two clones had intron sequences of ca 550 and 350 bp, ca 330 bp 3' non-translated sequences, and 18 and ca 70 bp poly(A) tails. Hb_5 cDNA represents full-sized human liver cytochrome b_5 cDNA (9). The presence of the alternative poly(A) splicing sequences in the 3' non-translated region of Hb_5 is indicated (17,18).

tails (18 to ca 70 bp). The Type II group (3 clones) also had variable length poly(A) tails (18-70 bp), but one clone (Fig. 1, IIa) had a shorter 3' non-translated sequence, i.e., ca 220 instead of 320 bp. Analysis of this 3' sequence indicated the use of an alternative poly(A) splicing site AACAAA (17) instead of AATATA (18). Both sites plus another AATATA are present in the longer 3' non-translated sequences of the Type I clones and the full-sized Hb_5 cDNA (Fig. 1) (9). There is no consensus AATAAA poly(A) signal in the Hb_5 sequence (9,19). Overall, there are three possible poly(A) splicing signals, of which two appear to be used. In addition, the sequence at the Type II intron-exon junction was different to that of the Type I intron-exon junction (Fig. 1). Using a combination of restriction enzyme mapping and DNA sequencing, we determined that the Type II sequence was

actually part of the Type I sequence occurring ca 1000 bp 5' of the Type I intron-exon junction (Fig. 1). The 3' end of the Type I intron, and both the 5' and 3' ends of the spliced sequence leading to the Type II intron agree with the GT-AG consensus sequence (20), which might explain the alternative splicing leading to the Type II clones.

The analysis of the Hb_5 gene system is further complicated by the presence of pseudogenes (Fig. 2). Neither of the pseudogenes contained introns in the coding sequence, and both are probably due to the retroviral insertion of Hb_5 mRNA. $Psgb_5 1$ was isolated from the Maniatis library and $Psgb_5 2$ from the Clontech library.

$Psgb_5 1$ contains the complete nucleotide sequence for liver b_5 plus 39bp of flanking 5' and 118bp of flanking 3' sequence. It also contains five deletions and 88 base changes, 43 of these changes occur in the coding region of the b_5 sequence (Fig. 2). In contrast $Psgb_5 2$ is a truncated b_5 sequence (Fig. 2), with the coding region equivalent starting at amino acid 43. There are 83bp 5' of this junction and 207 bp in the 3' flanking sequence (Fig. 2). Overall there are 59 bp changes, 31 in the coding region plus two deletions and one 15bp insertion. We do not yet know whether there are other pseudogenes present in the two libraries. Possibly different populations contain different numbers or types of pseudogenes. $Psgb_5 1$ has two small (6bp, 7bp) repeat sequences in the 5' and 3' flanking regions (Fig. 2 underlined sequences) whereas $Psgb_5 2$ has after allowing for minor insertions, 17bp repeat sequences in the 5' and 3' flanking regions (Fig. 2 underlined sequences). This is in agreement with other pseudogenes (21), and is indicative of the retroviral insertion of mRNA sequences (21).

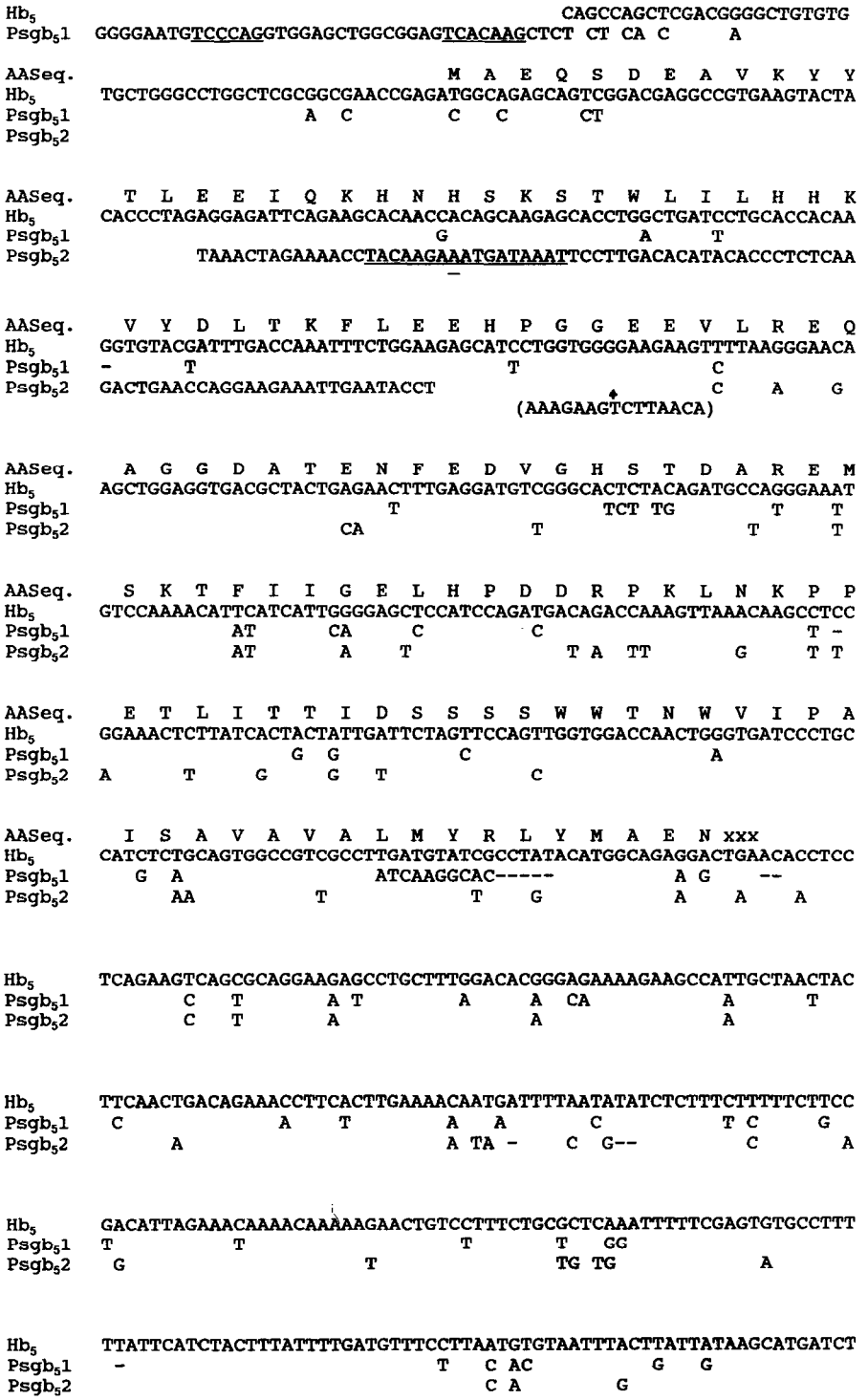

Figure 2. A comparison of the nucleotide sequences of the Hb_5 pseudogenes, $Psgb_51$, $Psgb_52$, with Hb_5. AASeq. indicates the amino acid sequence of Hb_5. For the two pseudogenes, only nucleotide changes are indicated. - indicates a deletion. Insertions into the coding sequence are bracketed. Insertions into sequences outside of the coding regions are double underlined. The repeat sequences are single underlined.

```
Hb5      TTTAAAAATATATTTGGCTTTTAAAGTAAAAAAAAAAAAAAAAAA
Psgb5 1      C        CC          A  C     T T GGCTGGGCGCCGTGGCTCATGCCTA
Psgb5 2              C  A       G              TAACAAGTTCTGAAATGATAAATAG

Psgb5 1  TAATCCCAGCACTTTGGGAGGCTGAAGGGGTTGGATCACAAGGTCAGGAGTTCGAAACCAACC
Psgb5 2  CCTACCAACCAAAAAAAAAAAAAAAAAAAAAAGGCCAGGACCAGATGGTTTCACAGCTGAATTC

Psgb5 1  TGGCCAATATGGTGAAACCCCGTGTCCACT
Psgb5 2  TACCAGATGTATAAAGAAAAGCTGGTACTATTTCTACTGAAACTATTCCAAAAAAATTGAAAA

Psgb5 2  AGAGGGACTCCTCCTTAACTCATTCTATAAAGCCAGCATCATCCTGATACCAAAGC
```

Fig. 2 - Continued

Overall, our data from the partially processed mRNA's would indicate that the Hb_5 gene(s) contains at least two exons with a >2kb intron separating them. Southern blot analysis of Hind III digested genomic DNA showed 5 hybridizing bands of 3-9 kb in size (data not shown). Other estimates based on the chicken and rat genomes are $\geq$2 and >5 exons, respectively (8,11). All the data could be explained by two genes, each with two exons, with possibly one or more pseudogenes. If, however, $Psgb_5 2$ was derived from partially processed Hb_5 mRNA then the truncation point in $Psgb_5 2$ might represent an intron-exon junction. In this case, the Hb_5 gene would have at least 3 exons, e.g., 5'-42, 43-86, 87-3', which is similar to one prediction (11). It is somewhat odd, that all the partially processed mRNA's we found contained coding sequences 87-3', whereas $Psgb_5 2$ contained coding sequences 43-3'. This would suggest the possibility of finding, partially processed mRNA's with coding sequences 43-3' and psuedogenes with coding sequences 87-3'. Clearly the analysis of the Hb_5 gene family is going to be much more complicated than originally anticipated.

ACKNOWLEDGMENTS. This work was supported by a grant from the American Heart Association, Ohio Affiliate.

The nucleotide sequences of Hb_5, $Psgb_5 1$ and $Psgb_5 2$ have been filed with Genbank under accession numbers M22865, M25765, and M24781 respectively.

REFERENCES

1. Oshino N (1982), in Int. Encyclopedia of Pharm. and Ther., eds. Schenkman T and Kupffer D. Pergamon Press, pp. 407-487.
2. Hultquist DE, Slaughter SR, Douglas RH, Sommer LJ, and Sahagian GG (1978), The Red Cell, ed. G Brewer, Alan R. Liss, pp. 199-211.
3. Hegesh E, Hegesh J and Kaftory A (1986), New Engl. J. Med., 314, 757-761.
4. Cristiano RJ, Yoo M, Dariush N and Steggles AW (1989), The Red Cell, ed. G Brewer, in press.
5. Mathews FS (1985), Prog. Biophys. Mol. Biol., 45, 1-56.
6. Douglas RH and Hultquist DE (1978), Proc. Natl. Acad. Sci. USA, 75, 3118-3122.
7. Cristiano RJ and Steggles AW (1989), Nucl. Acids Res., 17, 799.
8. Zhang H and Somerville C (1988), Arch. Biochem. Biophys., 264, 343-347.
9. Yoo M and Steggles AW (1988), Biochem. Biophys. Res. Commun., 156, 576-580.
10. Dariush N, Fisher CW and Steggles, AW (1988), Prot. Seq. Data Analysis, 1, 351-353.
11. Rigby JS, Bull PC, Ashworth A, Shephard EA, Santisteban I, and Phillips IR (1989), Biochem. Soc. Trans., 17, 194-195.
12. Lawn, RM, Fritsch EF, Parker RC, Blake G and Maniatis T (1978), Cell, 15, 1157-1174.
13. Maniatis T, Fritsch EF, and Sambrook J (1982), Molecular Cloning, A Laboratory Manual, Cold Spring Harbor.
14. Feinberg A and Vogelstein B (1984), Anal. Biochem., 137, 266-267.
15. Sanger F and Coulson AR (1975), J. Mol. Biol., 94, 441-448.
16. Steggles AW (1989), BioTechniques, 7, 241-242.
17. Spies M and Lodish HF (1985), Proc. Natl. Acad. Sci. USA, 82, 6465-6469.
18. Maniatis T and Reed R (1987), Nature, 325, 673-678.
19. Proudfoot N (1982), Nature, 298, 516-517.
20. Breathnach R and Chambon P (1981), Ann. Rev. Biochem., 50, 349-383.
21. Takenaka M, Noguchi T, Inoue H, Yamada K, Matsuda T and Tanaka T (1989), J. Biol. Chem., 264, 2363-2367.