Vol. 196, No. 1, 1993 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
October 15, 1993 Pages 440-446

# EXON STRUCTURE AT THE HUMAN ACP1 LOCUS SUPPORTS ALTERNATIVE SPLICING MODEL FOR **f** AND **s** ISOZYME GENERATION

K.D. Ayer Lazaruk*, J. Dissing**, and G.F. Sensabaugh*

*Forensic Science Group, School of Public Health, University of California, Berkeley, CA

**Institute of Forensic Genetics, University of Copenhagen, Copenhagen, Denmark

Received August 25, 1993

SUMMARY: The human ACP1 locus encodes a genetically polymorphic cytoplasmic low-molecular-weight acid phosphatase. Each of the common alleles encodes two isoforms, **f** and **s**. Both isozymes are of equal length (157 residues) but differ in sequence over an internal 34 residue segment. Substantial portions of the ACP1*A, *B and *C alleles common to Europeans have been sequenced. Six linearly positioned exons containing codons 14 to 157 were identified. Two exons of equal length (114bp) interspaced by a short (41bp), probably nonfunctional, intron encode the specific **f** and **s** segments, respectively. These findings strongly support an alternative RNA splicing hypothesis. In addition, three allele-specific base substitutions were encountered. 

A cytosolic low-molecular-weight acid phosphatase (cACP; E.C. 3.1.3.2) has been isolated from diverse vertebrate species(1-7); it is expressed in most, if not all, tissues. The human cACP is encoded at the ACP1 locus on the short arm of chromosome 2 (8) and is genetically polymorphic. Three common alleles, ACP1*A, ACP1*B, and ACP1*C, are found in caucasian populations (9,10). Each allele expresses a pair of electrophoretically distinct isozymes termed fast (f) and slow (s) according to their anodal mobility. Both the total amount of enzyme produced and the relative proportion of **f** and **s** isozymes differ for the various alleles; the **f:s** isozyme ratios for the *A, *B, and *C alleles are 2:1, 4:1, and 1:4 respectively (11). The **f** and **s** isozymes have been found to differ in significant catalytic properties, notably, in kinetic constants, level of activity modulation by purines and folates, and phosphotransferase activity (12-21); they also differ in immunological and stability properties (11,20-22). These differences coupled with the allelic variation in **f:s** ratio gives

0006-291X/93 $4.00

rise to significant genotype-dependent variation in enzymatic and molecular properties (19-21,23-27).

The molecular basis for the differences between **f** and **s** isozymes has been clarified by recent protein sequencing studies on the products of the *A, *B, and *C alleles (28,29). The **f** and **s** isozymes are identical over the protein sequence regions spanning residues 1-39 and 74-157, but differ significantly (59%) in sequence in the middle region, residues 40-73. The sequence differences in this "signature" region presumably account for the differences in the catalytic and molecular properties of the **f** and **s** isozymes. Subsequently determined cDNA sequences corresponding to the **Bf** and **Bs** isozyme sequences (30, Lazaruk and Sensabaugh, unpublished) show sequence identity for the regions spanning codons 0-39 and 76-157 and extending into the 3' untranslated region. The cDNA sequences of the **f** and **s** signature regions (codons 40-73) differ by almost 50%; the **f** and **s** cDNA sequences differ by translationally silent base substitutions at codons 74 and 75.

We have previously postulated that the **f** and **s** isozyme pairs are generated by an alternative splicing mechanism in which the **f** and **s** signature sequences are encoded in mutually exclusive exons (22,28). We demonstrate here that the exon structure of the ACP1 locus is consistent with this hypothesis.

## MATERIALS AND METHODS

Human reticulocyte cDNA libraries in λgt11 were kindly provided by Dr. J. Conboy (31) and Dr. Y. Edwards (32). Genomic DNA was extracted from donor blood samples of known ACP1 genotype using standard methods. A yeast artificial chromosome (YAC Y268C12) with a 450 Kb insert containing the ACP1 locus was identified from a YAC library at Lawrence Livermore National Laboratory using locus-specific primers (kindly performed by Dr. Harvey W. Mohrenweiser). AmpliTaq DNA polymerase was obtained from Perkin Elmer (Norwalk, CT), and NuSieve and Seakem agarose from FMC (Rockland, ME).

PCR reactions were carried out in a total volume of 25 or 50 µl with 50 ng of genomic DNA from ACP1*A, B, or C homozygote individuals or with 500 ng of reticulocyte cDNA. All reactions were performed in buffer containing, 50 mM KCl, 10 mM Tris·HCl (pH 8.3), 2.5 mM $MgCl_2$, 180 µM each dNTP, and 1-2.5 units AmpliTaq DNA polymerase. Ten to 50 pmol of each primer was used for initial amplification. PCR product gels (1.5% NuSieve/0.5% Seakem) were run in TBE buffer and relevant bands excised. DNA was eluted from excised bands into 50-100 µl of $H_2O$ and used as template for further amplification using the same or nested primers. When PCR amplification yielded multiple bands, the ACP-specific bands were

identified through the use of nested primers following the hot-blot procedure of Parker et al. (33). Primer sequences (Fig. 1) are available upon request.

Single-stranded DNA was generated by asymmetric PCR technique using a 50:1 ratio of excess:limiting primer (34). Asymmetric PCR products were purified on *Ultrafree MC 30,000 MW* spin filters (Millipore, Bedford, MA). The DNA was eluted off the membrane with 20-40 µl of $H_2O$ and 7 µl was used for sequencing by the method of Sanger (35) using a *Sequenase* 2.0 Sequencing kit (US Biochemical, Cleveland, OH). PCR products greater than 500 bp in length were sequenced directly using cycle sequencing (*fmol Cycle Sequencing*, Promega, Madison, WI) and a single primer.

## RESULTS

The structure of the ACP1 locus has been determined by genomic sequencing using overlapping genomic DNA fragments obtained by PCR. Initial segments of cDNA sequences were amplified from human reticulocyte cDNA libraries using degenerate primers based on the amino acid sequence of the **f** and **s** isozymes (28). This allowed the construction of specific primers facilitating the determination of the entire cDNA sequence and thereafter the genomic sequence. Two cDNA sequences were sequenced, one corresponding to the **Bf** isozyme and the other to the **Bs** isozyme. These sequences were identical to human cDNA sequences reported while this work was in progress (30).

Genomic sequence spanning approximately 6 Kb of the ACP1 locus has been determined (Fig. 1; GenBank Accession Nos. L06507 and L06508). Six linearly positioned exons have been identified; these account for all the coding sequence except that for amino acid residues 1-13 at the N-terminus. The **f** and **s** signature sequences are encoded in adjacent exons, E3F and E3S, each 114 bp in length; these exons contain codons 38-75 (Fig.2). These two exons are

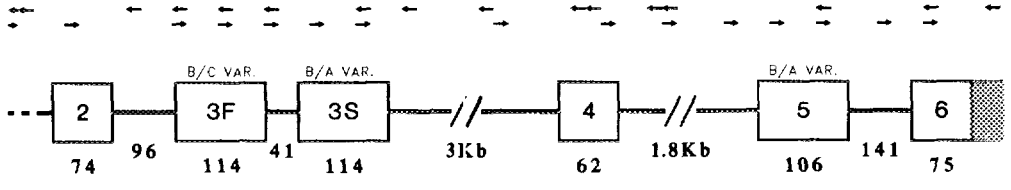


Fig. 1. Scheme of the ACP1 gene locus of human low-molecular-weight acid phosphatase. Boxes represent exons, lines represent introns, and the hatched box represents 3' non-coding sequence. The relative positions of the primers used for PCR amplification and sequencing are indicated by arrows. Exon and intron sizes (bp) are indicated below. Sites of allelic variation are indicated above.

```
 *B,*A,*C  ...ATC CAACCTAACC CTGTTTCCCC ACCCCTCCCT TTTTTTAAAG

Exon 2                    ------------------------10--->
 *B,*A,*C  GT AAC ATT TGT CGA TCA CCC ATT GCA GAA GCA GTT TTC

 *B,*A,*C  AGG AAA CTT GTA ACC GAT CAA AAC ATC TCA GAG AAT

Intron 2                                  <--25--------------------
 *B,*A,*C  GTAAGTACCA TTCATTATCT TAAAGAGGCC AACCTGAACT CCTCTGGG

 *B,*A,*C  CAGGAAATTGCA AAAAAAAAAA AAAATTCCAT GTTTCTTCCC CTGCAG

Exon 3F            <--11-------------19-->
 *B,*A     TGG AGG GTA GAC AGC GCG GCA ACT TCC GGG TAT GAG ATA
 *C        ... ... ... ... ..T ... ... ... ... ... ... ... ...

                                                <--8-------------7-
 *B,*A,*C  GGG AAC CCC CCT GAC TAC CGA GGG CAG AGC TGC ATG AAG

           -->
 *B,*A,*C  AGG CAC GGC ATT CCC ATG AGC CAC GTT GCC CGG CAG

Intron 3F       <--44-------------39-->
 *B,*A,*C  GTACCGTCCT TGGACTTGAA GTTGTGTGTT TTGTGTTTCA G

Exon 3S                                   ----------------------
 *B,*C     TGG GTC ATT GAC AGC GGT GCT GTT TCT GAC TGG AAC GTG
 *A        ... ... ... ..T ... ... ... ... ... ... ... ... ...

           20-->                                          <--16----
 *B,*A,*C  GGC CGG TCC CCA GAC CCA AGA GCT GTG AGC TGC CTA AGA

           ----------------------15-->
 *B,*A,*C  AAT CAT GGC ATT CAC ACA GCC CAT AAA GCA AGA CAG

Intron 3S
 *B  GTAGACAAGC TCTTGTTCAA TTTTCTAATA TATAGAGTCC AGTAACTTGA...
```

Fig. 2. The sequence of the ACP1 locus for the ACP1*B, *A, and *C alleles in the region encompassing the F and S alternatively spliced exons. Primers used for this region are indicated by numbered arrows. The length of the polyA tract in intron I2 is the consensus length from multiple sequencing experiments.

separated by a short (41bp) intervening sequence I3F. The E3F and E3S exons differ in 51 base positions, 10 of which are translationally silent. The genomic sequences have been determined for the intron and exon immediately 5' to E3F and for the three introns and exons 3' to E3S. The interior portions of the 3Kb and 1.8 Kb introns have yet to be determined.

Sequencing of genomic DNA from individuals of the three homozygous types A, B, and C, show an identical gene structure. The ACP1*B and *A allelic sequences were found to differ at two base positions, a silent C-T transition at codon 41 in the S exon and an A-G transition at codon 105. A single silent C-T transition at codon 43 in the F exon was found to distinguish the *C allele from *A and *B. The allelic sequence differences at codon 43 in the F exon and at codon 105 generate restriction sites that can be used for PCR-based genotyping at the ACP1 locus (36).

Amplification of genomic DNA with certain primer sets yielded products in addition to those comprising the sequence above. Analysis of these additional products showed sequence similar to, but not identical with the ACP1 gene sequence; some lacked intron sequence. These observations suggest the presence of one or more ACP1 pseudogenes.

## DISCUSSION

The existence of an **f** and **s** isozyme pair appears to be a characteristic feature of cACP in higher vertebrates (see ref. 29; Sensabaugh, unpublished); the mechanism for their generation is likely to be common across species. We had previously proposed alternative splicing of mutually exclusive exons to account for this phenomenon based on peptide mapping studies (22) and subsequently on protein sequence data (28). Although this model was recently questioned by Wo et al. (30), the present results provide evidence for the alternative splicing model.

The human ACP1 gene structure shows the **f** and **s** signature sequences to be encoded by two distinct exons, E3F and E3S; these exons are flanked on either side by exons encoding the N- and C-terminal sequences identical in the **f** and **s** isozymes. The 41 bp non-coding sequence segment between the exons E3F and E3S is too short to be spliced out on its own (37). This is consistent with a mechanistic picture for mutually exclusive alternative splicing in which a transcript segment containing either E3F-I3F or I3F-E3S is excised to form the mature transcripts for the **s** and **f** isozymes, respectively. There is no evidence of tissue-specific or developmental regulation of the **f**:**s** isozyme ratio for a given genotype (38,39); the mutually exclusive alternative splicing thus appears to be a constitutive trait.

The ACP1*A, *B, and *C allelic gene sequences are distinguished by base substitutions at three sites located in exons; there may be additional sequence variation in the as yet unsequenced regions of the ACP1 locus. An A-G transition in codon 105 of exon 5 accounts for the Gln-105 to Arg substitution that distinguishes the A enzyme types from the B and C types. The remaining two substitutions, a C-T transition 12 bases into the S exon that distiguishes the *A allele from the *B and *C alleles, and a C-T transition 15 bases into the F exon that distiguishes the

*C allele from the *A and *B alleles, are both translationally silent. Both, however, are of potential interest. Given their location in the alternatively spliced exons and the absence of any allelic sequence variation in surrounding exon and intron sequences, both may be involved in the genotype-specific regulation of **f:s** isozyme ratios. Exon based substitutions in the vicinity of splice sites have been shown to influence splice site recognition (40,41). This model for ACP1 expression predicts that rare genetic variants with altered **f:s** ratios will have base substitutions in the E3F-I3F-E3S sequence region.

## ACKNOWLEDGMENTS

Portions of this work were supported by grant 86-15-CX0044 from the National Institute of Justice to G.F.S. We thank Drs. J. Conboy, Y. Edwards, H. Mohrenweiser, and T. Laber for providing materials used in this study.

## REFERENCES

1. Baxter, J.H. and Suelter, C.H. (1985) Arch. Biochem. Biophys. 239, 29-37.
2. Fujimoto, S., Murakami, K., Ishikawa, A., Himi, K., and Ohara, A. (1988) Chem. Pharm. Bull. Tokyo 36, 3020-3026.
3. Dissing, J. and Svensmark, O. (1990) Biochim. Biophys. Acta 1041, 232-242.
4. Zhang, Z.Y. and Van Etten, R.L. (1990) Arch. Biochem. Biophys. 282, 39-49.
5. Saeed, A., Tremori, E., Manao, G., Camici, G., Cappugi, G., and Ramponi, G. (1990) Physiol. Chem. Phys. & Med. NMR 22, 81-94.
6. Panara, F., Angiolillo, A., Secca, T., Dirosa, I., Fagotti, A., and Pascolini, R. (1991) Int. J. Biochem. 23, 1115-1122.
7. Manao, G., Pazzagli, L., Cirri, P., Caselli, A., Camici, G., Cappugi, G., Saeed, A., and Ramponi, G. (1992) J. Protein. Chem. 11, 333-345.
8. Ferguson-Smith, M.A., Newman, B.F., Ellis, P.M., Thomson, D.M.G., and Riley, I.D. (1973) Nat. New Biol. 243, 271-274.
9. Hopkinson, D.A., Spencer, N., and Harris, H. (1963) Nature 199, 969-971.
10. Harris, H. (1980) The principles of human biochemical genetics, p.190. Elsevier/North-Holland Biomedical Press, Amsterdam.
11. Dissing, J. (1987) Biochem. Genet. 25, 901-917.
12. Luffman, J.E. and Harris, H. (1967) Am. J. Hum. Genet. 30, 387-400.
13. Bottini, E., Modiano, G., Santolamazza, C., Filippi, G., and Businco, L. (1971) Clin. Chim. Acta 31, 243-254.
14. Fisher, R.A. and Harris, H. (1971) Ann. Hum. Genet. 34, 439-448.
15. White, I.N.H. and Butterworth, P.J. (1971) Biochim. Biophys. Acta 229, 202-207.
16. White, I.N.H. and Butterworth, P.J. (1971) Biochim. Biophys. Acta 229, 193-201.

17. Fenton, M.R. and Richardson, K.E. (1971) Arch. Biochem. Biophys. 142, 13-21.
18. Dissing, J., Dahl, O., and Svensmark, O. (1979) Biochim. Biophys. Acta 569, 159-176.
19. Golden, V.L. and Sensabaugh, G.F. (1986) Hum. Genet. 72, 340-343.
20. Dissing, J. and Svensmark, O. (1990) Biochim. Biophys. Acta 1041, 232-242.
21. Dissing, J., Rangaard, B., and Christensen, U. (1993) Biochim. Biophys. Acta 1162, 275-282.
22. Dissing, J. and Sensabaugh, G.F. (1987) Biochem. Genet. 25, 919-927.
23. Mansfield, E. and Sensabaugh, G.F. (1977) Lancet 2, 201-202.
24. Mansfield, E. and Sensabaugh, G.F. (1978) In The Red Cell (G.J. Brewer, Ed.), pp. 233-247. Allan R. Liss, New York.
25. Sensabaugh, G.F. and Golden, V.L. (1978) Am. J. Hum. Genet. 30, 553-560.
26. Dissing, J. (1986) In Advances in Forensic Haemogenetics 1 (B. Brinkmann and K. Henningsen, Eds.), pp. 127-131. Springer, Berlin.
27. Dissing, J. (1992) In Advances in Forensic Haemogenetics 4 (C. Rittner and P.M. Schneider, Eds.), pp. 303-305. Springer, Berlin.
28. Dissing, J., Johnsen, A.H., and Sensabaugh, G.F. (1991) J. Biol. Chem. 266, 20619-20625.
29. Dissing, J. and Johnsen, A.H. (1992) Biochim. Biophys. Acta 1121, 261-268.
30. Wo, Y.Y.P., Mccormack, A.L., Shabanowitz, J., Hunt, D.F., Davis, J.P., Mitchell, G.L., and Van Etten, R.L. (1992) J. Biol. Chem. 267, 10856-10865.
31. Conboy, J., Kan, Y.W., Shohet, S.B., and Mohandas, N. (1986) Proc. Natl. Acad. Sci. USA 83, 9512-9516.
32. Barlow, J., Lowe, N., Edwards, Y., and Butterworth, P. (1987) Nucleic Acids Research 15, 2386.
33. Parker, J.D., Rabinovitch, P.S., and Burmer, G.C. (1991) Nucleic. Acids. Res. 19, 3055-3060.
34. Gyllensten, U.B. (1989) Biotechniques 7, 700-708.
35. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
36. Sensabaugh, G.F. and Lazaruk, K.Ayer (1993) Hum. Mol. Genet. 2, 1079.
37. Ruskin, B., Greene, J.M., and Green, M.R. (1985) Cell 41, 833-844.
38. Swallow, D.M., Povey, S., and Harris, H. (1973) Ann. Hum. Genet. 37, 31-38.
39. Blake, N.M., Kirk, R.L., Barnes, K.R., and Thompson, J.M. (1973) Jpn. J. Hum. Genet. 18, 10-23.
40. Cooper, Thomas A. (1992) J. Biol. Chem. 267, 5330-5338.
41. Libri, D., Goux-Pelletan, M., Brody, E., and Fiszman, M.Y. (1990) Mol. Cell. Biol. 10, 5036-5046.