Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 1234–1238, March 1987
Biochemistry

Nucleotide and amino acid sequences of human intestinal alkaline phosphatase: Close homology to placental alkaline phosphatase

(cDNA libraries/isozymes/multigene enzyme family/gene evolution)

PAULA S. HENTHORN*†, MICHAEL RADUCHA*, YVONNE H. EDWARDS‡, MITCHELL J. WEISS*, CLIVE SLAUGHTER§, MARY ANN LAFFERTY*, AND HARRY HARRIS*¶

*Department of Human Genetics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104; †Division of Biochemical Development and Molecular Diseases, Children's Hospital of Philadelphia, Philadelphia, PA 19104; ‡Medical Research Council, Human Biochemical Genetics Unit, University College London, London NW1, England; and §Howard Hughes Medical Institute, University of Texas Health Science Center at Dallas, Dallas, TX 75235

Contributed by Harry Harris, November 17, 1986

ABSTRACT **A cDNA clone for human adult intestinal alkaline phosphatase (ALP) [orthophosphoric-monoester phosphohydrolase (alkaline optimum); EC 3.1.3.1] was isolated from a λgt11 expression library. The cDNA insert of this clone is 2513 base pairs in length and contains an open reading frame that encodes a 528-amino acid polypeptide. This deduced polypeptide contains the first 40 amino acids of human intestinal ALP, as determined by direct protein sequencing. Intestinal ALP shows 86.5% amino acid identity to placental (type 1) ALP and 56.6% amino acid identity to liver/bone/kidney ALP. In the 3′-untranslated regions, intestinal and placental ALP cDNAs are 73.5% identical (excluding gaps). The evolution of this multigene enzyme family is discussed.**

The enzymes termed alkaline phosphatases (ALPs) [orthophosphoric-monoester phosphohydrolase (alkaline optimum); EC 3.1.3.1] hydrolyze a wide variety of monophosphate esters at high pH optima and are widely distributed in nature (1). In man and other higher primates, these membrane-bound glycoproteins appear to be the products of at least three related gene loci, named after the tissues in which they are characteristically expressed [placental, intestinal, and liver/bone/kidney (L/B/K) ALPs] (2–5). These conclusions are based on an abundance of biochemical, immunological, and direct protein sequence data, which also indicate that the placental and intestinal ALPs are more closely related to each other than either is to L/B/K ALP (2, 6–8). More recently, cDNA clones of placental and L/B/K ALPs have been isolated (9–12). To further the study of this ALP multigene enzyme family, we have isolated and sequenced a cDNA that encodes the ALP expressed in adult human intestine and compared this sequence with those previously determined for the placental and L/B/K ALP cDNAs.

MATERIALS AND METHODS

Purification of Adult and Fetal Intestinal ALPs. Human adult and fetal intestinal ALPs were purified from adult small intestinal mucosal scrapings and newborn meconium samples, respectively, as described (13).

Antisera. Antisera to human placental ALP and to human fetal intestinal ALP were raised in rabbits as described (11, 14). Antiserum against placental ALP has been shown to cross-react with intestinal ALP (15).

Protein Sequence Determination. Amino-terminal amino acid sequence analysis of purified adult intestinal ALP was performed as described (11).

Cell Lines. D98/AH-2 is a cell line, originally derived from HeLa cells, that expresses considerable amounts of an ALP that closely resembles the adult intestinal ALP (14, 16, 17).

Construction of a D98/AH-2 cDNA Library. Poly(A)-RNA was isolated from D98/AH-2 cells and used to construct a cDNA library in the bacteriophage expression vector λgt11 as described (12). The resulting library of 10^6 recombinant phages, as judged by the number of clear plaques on appropriate indicator medium (18), was amplified on *Escherichia coli* strain Y1090r$^-$ (Promega Biotec, Madison, WI) prior to screening.

Construction of a Human Intestinal cDNA Library. The human intestinal cDNA library was constructed in λgt11. The cDNA was prepared from poly(A)-RNA from full thickness lower jejunum from a male European cadaver kidney donor and was used in the construction of the cDNA library as described (19). The resulting library contained 3×10^5 independent recombinant phages (92% clear plaques) before amplification.

Screening the D98/AH-2 cDNA Library. The D98/AH-2 cDNA library was screened with a 1:1 mixture of the antisera to human placental and fetal intestinal ALPs as described (11, 18). The fragment probe used for hybridization rescreening, a 1.4-kilobase (kb) *Sst* I fragment containing both protein-coding and 3′-untranslated sequences of a placental ALP cDNA (11), was labeled by calf thymus DNA primer labeling (20). DNA from phage candidates was digested with *Eco*RI, separated by electrophoresis in agarose gels, transferred to a Zetabind membrane (AMF Specialty Materials Group, Meriden, CT), and hybridized to a ^{32}P-radiolabeled probe according to the manufacturer's instructions. The membrane was washed in 0.45 M NaCl/45 mM sodium citrate at 68°C and was autoradiographed. The cDNA inserts of hybridizing phage clones were isolated and inserted into plasmid vectors for further characterization.

Phage candidates that contained the amino-terminal protein coding region were identified by hybridization to a mixture of oligonucleotides of sequence 5′ GARGARGAR-AAYCC 3′ (Biosearch, San Rafael, CA) (where R = unspecified purine nucleoside and Y = unspecified pyrimidine nucleoside) corresponding to amino acid positions 5 through 9 of the intestinal ALP (11, 21) as described (11).

Screening the Human Gut cDNA Library. The 2.5-kb insert from a cDNA clone isolated from the D98/AH-2 cDNA library (called cD98#7, see *Results*) was used as a hybridization probe for screening the adult gut cDNA library. Transfer of bacteriophage plaques to nitrocellulose membranes and hybridization of membranes to labeled probe were performed under standard conditions (22).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: ALP, alkaline phosphatase; L/B/K, liver/bone/kidney.

¶To whom reprint requests should be addressed.

DNA Sequence Analysis. All DNA sequences were determined by the Sanger dideoxy chain-termination method (23). M13 phage subclones were generated by forced cloning and by the T4 polymerase deletion method of Dale *et al.* (24). The sequencing strategy is illustrated in Fig. 1. Note that not all base sequences were determined on both strands. A single base ambiguity (adenosine and guanosine were seen in different subclones) at position 207 was not resolved. A guanosine residue at this position has been used in all analyses. Computer analyses of DNA and protein sequences were performed using the IBI/Pustell DNA and Protein Sequence Analysis System (International Biotechnologies, New Haven, CT), NUCALN and PRTALN (25).

RESULTS

Isolation of Intestinal ALP cDNA Clones from D98/AH-2 cDNA. The D98/AH-2 cDNA library was screened with a mixture of anti-placental and anti-fetal intestinal ALP antisera. We chose to screen with this mixture because our previous experience in isolating cDNA clones for placental ALP indicated that the anti-fetal intestinal antiserum may not recognize ALP fusion proteins produced in *E. coli* (P.S.H., unpublished observation). A similar observation concerning an anti-placental ALP antiserum has been reported by Millan (10). Of the $>10^6$ phages screened, 25 were found to react with the antiserum mixture. Of these 25, 7 phages hybridized to a DNA fragment of a placental ALP cDNA clone. Only one phage clone, designated cD98#7, hybridized to an oligonucleotide mixture predicted from the amino-terminal protein sequence of intestinal ALP. The DNA sequences of the distal ends of this cDNA were determined. One of these sequences encodes a polypeptide that contains the first 40 amino acids of fetal and adult intestinal ALPs, which are identical as determined by amino acid sequencing of fetal and adult intestinal ALP amino termini. This amino acid sequence is identical to that determined by Hua *et al.* (8) except for a lysine for threonine substitution at amino acid position 31. Based on this DNA and amino acid sequence data, we conclude that clone cD98#7 contains a cDNA for the intestinal ALP expressed in D98/AH-2.

Isolation of Intestinal ALP cDNA Clones from Human Intestinal cDNA. The cDNA insert of clone cD98#7 was used as a probe to screen a human intestinal cDNA library for intestinal ALP cDNA sequences. Of the $>5.0 \times 10^5$ recombinant phages screened, 12 phage candidates that hybridized to the cD98#7 insert fragment were isolated. The phage clone containing the longest cDNA insert, clone gut5, was chosen for further analysis. A restriction enzyme map of this cDNA insert is shown in Fig. 1. Note that the insert is not flanked by *Eco*RI enzyme sites as would be expected from the cloning strategy used to construct cDNA libraries in phage λgt11.

DNA Sequence Analysis. The nucleotide sequence of the cDNA insert of clone gut5 was determined by the sequencing strategy indicated in Fig. 1. The cDNA is 2513 base pairs (bp)

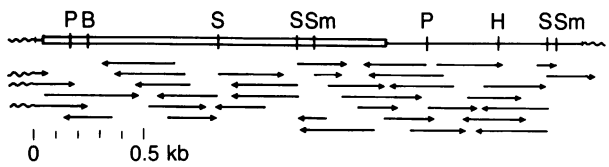

FIG. 1. Restriction enzyme map of the gut5 intestinal ALP cDNA insert. Locations of restriction enzyme-cut sites are shown above the line. The straight portions of the line represent 5'- and 3'-untranslated sequences. The open box represents the location of the 1587-bp open reading frame. The flanking wavy lines represent phage λgt11 sequences. The strategy used to obtain the sequence of the insert is illustrated below the line. Abbreviations: B, *Bam*HI; H, *Hin*dIII; P, *Pst* I; S, *Sst* I; Sm, *Sma* I.

in length and contains an open reading frame, beginning at base position 24, that encodes a 528-amino acid polypeptide. The complete nucleotide sequence of the gut5 cDNA open reading frame and the 528-amino acid polypeptide that it encodes are shown in Fig. 2. The termination codon at nucleotide positions 1608–1610 is followed by 903 nucleotides of 3'-untranslated sequence (shown in Fig. 3). A single poly(A) addition signal (AATAAA) is located at positions 2473–2478 and precedes the poly(A) sequence by 16 nucleotides.

The protein encoded by the gut5 cDNA clone contains the first 40 amino acids of human intestinal ALP as determined by protein sequencing. These 40 amino acids are preceded by 19 residues that presumably represent a signal peptide, as seen in other ALP cDNA clones (10, 12). Also seen in this cDNA, as in other ALP cDNA clones, is a stretch of hydrophobic amino acids at the carboxyl terminus that could potentially participate in membrane localization. Three potential N-linked glycosylation sites (Asn-Xaa-Thr/Ser) are present, beginning at amino acid positions 122, 249, and 410.

DISCUSSION

Comparisons of the coding regions of the intestinal and placental ALP cDNAs and their deduced amino acid sequences are shown in Fig. 2. Below the intestinal ALP nucleotide and protein sequences are written those amino acid and nucleotide residues that differ in sequence from the placental ALP cDNA. Two gaps, one within the signal peptide and the other near the carboxyl terminus, have been introduced to optimize the alignment. The gap in the intestinal ALP signal sequence results in a shorter hydrophobic region as compared to the placental ALP signal peptide, but the overall structure of the intestinal ALP signal peptide is well within established expectations of a signal sequence (26). The proteins differ at 71 amino acid residues (excluding the gaps) to give 86.6% identity at the amino acid level. At the nucleotide level, these sequences show 89.5% identity. The histograms shown in Fig. 4 plot the silent and replacement substitutions between intestinal and placental ALP cDNA sequences as a function of position. Silent substitutions appear to be randomly distributed along the protein length, except perhaps between amino acid residues 391 and 410. Two features of the distribution of replacement substitutions are notable. Although the ratio of silent to replacement substitutions is 1:1 over the whole protein, this ratio is 15:2 between amino acid residues 111 and 200, implying that the amino acid sequence in this region is highly conserved. The other unusual feature is seen between amino acid residues 391 and 430. Here there are 32 substitutions (17-replacement and 15-silent) representing 19% of the total substitutions in a region containing only 7.5% of the open reading frame. The significance of these observations awaits further studies relating ALP structure and function as well as analyses of ALP genes in other species.

DNA sequence homology is also seen between intestinal and placental ALP cDNAs in their 5'- and 3'-untranslated regions. An alignment of these sequences is shown in Fig. 3. In the 3'-untranslated region, 73.9% of the aligned bases are identical if gaps are not counted. The 26.1% divergence seen in the 3'-untranslated region closely approximates the estimated 26% divergence at silent sites in the protein coding region (27). This may indicate that the rates of evolutionary divergence in these two regions have been essentially the same.

This relatively long 3'-untranslated region may be useful in estimating the time in evolution when the gene duplication giving rise to placental and intestinal ALPs occurred. Assuming that this region is not under selection, we can compare the percentage of divergence (26.1%) to the divergence between presumably unselected DNA sequences in

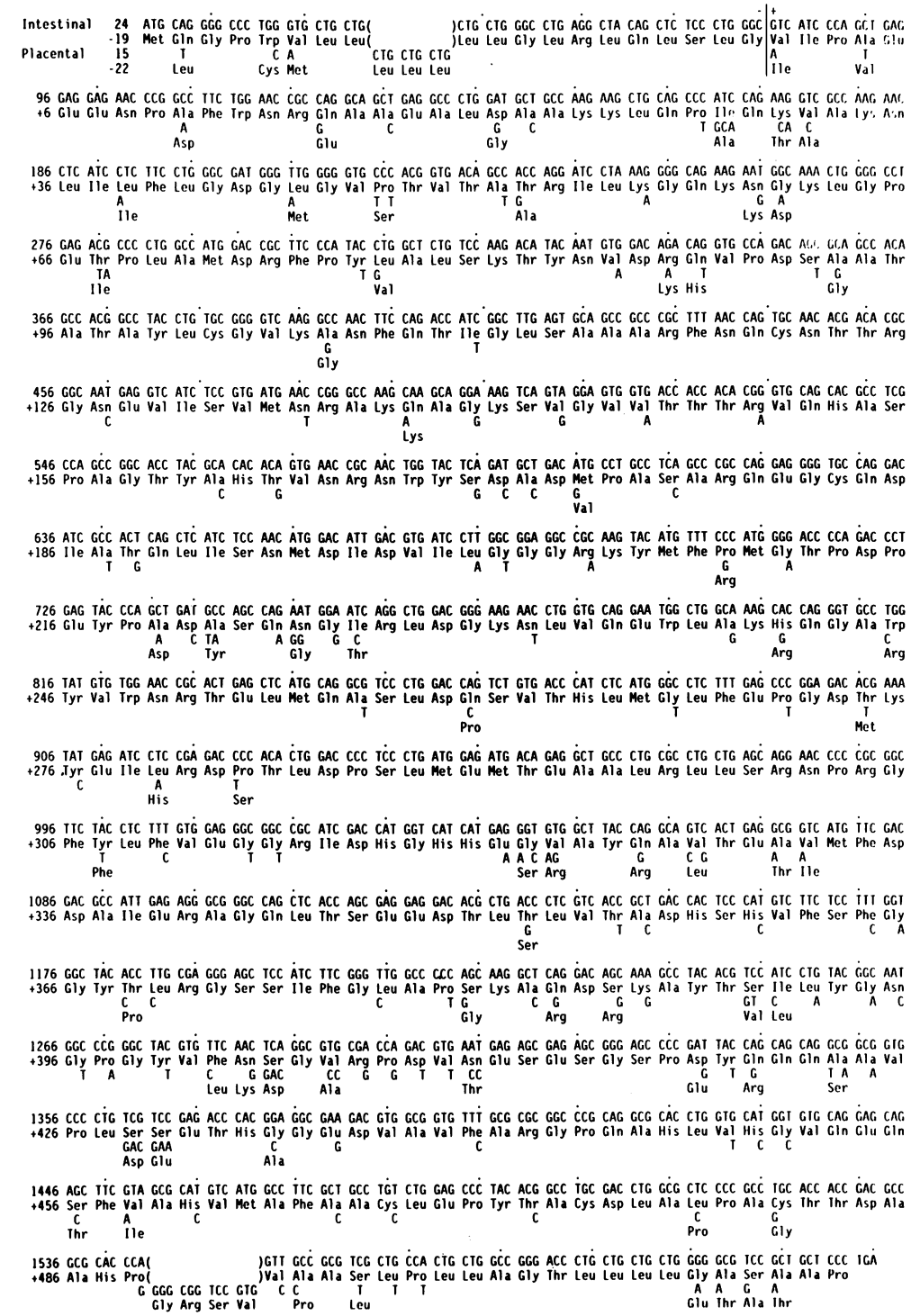

FIG. 2. Comparison of the coding regions of intestinal and placental ALP cDNAs. The top line shows the nucleotide sequence of the gut5 intestinal ALP cDNA open reading frame, beginning at position 24 of the entire cDNA insert. Directly below it is the translation of the open reading frame. The first amino acid corresponds to position −19 of the putative signal peptide sequence. The boundary between the signal peptide and mature protein is indicated by a vertical line. Below these sequences are shown the nucleotide and the deduced amino acids of a type 1 placental ALP cDNA that differs from the intestinal sequences. The placental ALP sequence (type 1) is from ref. 11. Gaps introduced to optimize the alignment are enclosed by parentheses and dots are placed above every 10th nucleotide residue. Amino acid position numbers (preceded by + or −) and nucleotide position numbers are at the beginning of each row. Position numbers are identical between the placental and intestinal ALP sequences between the two gaps.

other pairs of species. For example, the η-globin pseudogene is 12.6% divergent between apes and a New World monkey, the spider monkeys (28). DNA·DNA hybridization analyses of single copy DNAs show 13% divergence between apes and spider monkeys (29). On the other hand, the Galago (a prosimian) shows a 25% divergence from human in sequences flanking the ε-globin gene (30). Thus, the 26.1% divergence seen between placental and intestinal ALP 3′-untranslated regions places the gene duplication before the divergence of New World monkeys and the lineage that gave rise to both the Old World monkeys and the hominoids, perhaps as early as the time of the divergence of the prosimians from the anthropoids (monkeys and apes). This estimate of the time of duplication is consistent with an unexpected feature of

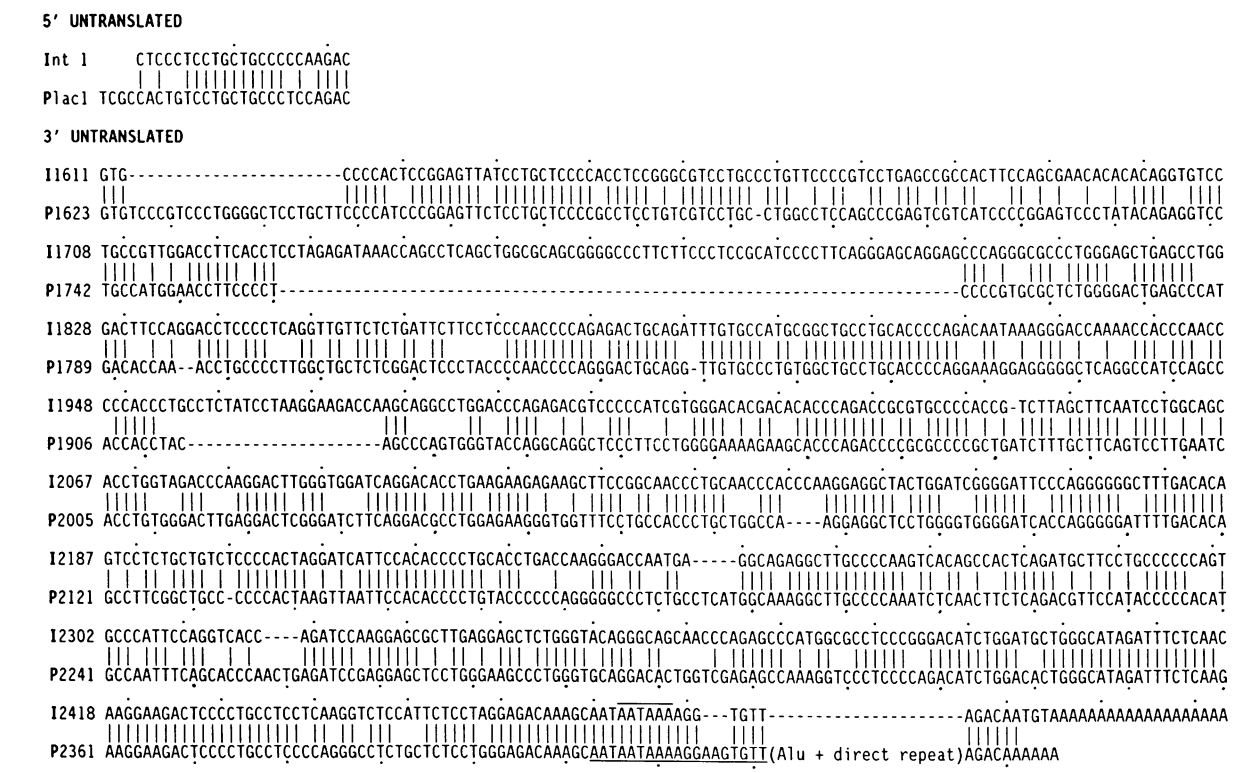

FIG. 3. Nucleotide alignment of the 5′- and 3′-untranslated regions of the gut5 intestinal ALP cDNA insert (Int and I) and placental ALP cDNA insert (Plac and P). The placental ALP cDNA 5′-untranslated sequence is from ref. 10. The placental ALP cDNA 3′-untranslated sequence is from the type 3 placental ALP cDNA (11). Nucleotide position numbers are shown at the left. Dashes show gaps that are introduced to optimize the total number of matches between the two sequences. Vertical lines connect identical bases. The single poly(A) addition signal in the intestinal ALP cDNA is overlined. In the placental cDNA, the sequences of an *Alu* family repeat element and one copy of a direct repeat that flanks it are not given, but their position is indicated. A single copy of that *Alu*-flanking direct repeat is underlined. Dots are placed at 10-nucleotide intervals.

primate ALPs. The placental ALP locus is expressed in the placentas of the great apes but not in the placentas of many other primate species (31). However, a very similar enzyme has been found in lung though not in placentas or other tissues of Old World monkeys (31). The gene expressed in Old World monkey lungs may be orthologous to the human placental ALP gene.

The alignment of 3′-untranslated sequences in Fig. 3 also shows that an *Alu* family repetitive sequence found in the placental ALP cDNAs (11) is not seen in the intestinal ALP cDNA 3′-untranslated region. This suggests that the *Alu* family sequence was inserted after the gene duplication that gave rise to the placental ALP gene.

The location of this *Alu* sequence is of interest with respect to the locations of poly(A) addition signals. The only poly(A) addition signal in the intestinal ALP cDNA sequence corresponds to the poly(A) addition signal found on the 5′ side of the *Alu* family sequence in placental ALP cDNAs (see Fig. 3). This 5′-most signal does not appear to be utilized in placental ALP mRNAs. Evidence for this statement is 2-fold. First, of four independent placental ALP cDNA clones reported (9, 11), all contain the *Alu* family sequence and, therefore, have not utilized the 5′-most poly(A) addition signal. In addition, only a single mRNA size class is seen in placental mRNA when probing with placental ALP cDNA sequences (ref. 9; P.S.H., unpublished data). We presume that the insertion of

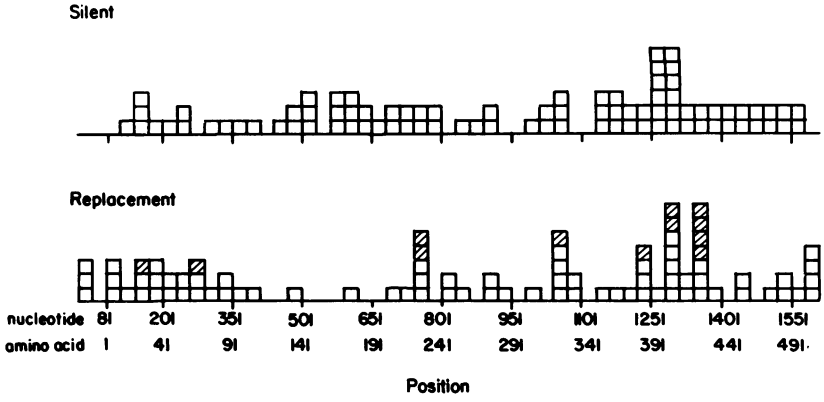

FIG. 4. Histograms plotting silent and replacement nucleotide substitutions between the protein coding regions of placental and intestinal ALP cDNAs as a function of position. Each square represents a single nucleotide substitution in an interval of 30 bp (10 amino acids). Amino acid and nucleotide positions are indicated below. Nucleotide substitutions that could potentially cause an amino acid replacement, but are the second or third replacement substitutions in a single codon, are indicated as cross-hatched boxes.

the *Alu* sequence effectively removes a sequence that is immediately 3′ to the 5′-most poly(A) addition signal and is necessary for some aspect of poly(A) addition. Evidence that downstream sequences affect the utilization of poly(A) addition signals has been found in other systems (32, 33).

Immunological and biochemical studies of these proteins indicate that they are members of a multigene enzyme family, presumably derived from a common ancestor by a series of gene duplications. Immunological techniques have also shown that the ALPs isolated from intestine and placenta are much more closely related to one another than either is to the L/B/K ALP (15). These relative relationships are also seen in comparisons of the deduced amino acid sequences of the intestinal, placental (9–11), and L/B/K proteins (12). An alignment of the deduced amino acid sequences of intestinal and L/B/K ALPs (not shown) indicates that these proteins are 56.6% identical over aligned residues (gaps introduced to optimize the alignment were not counted). As expected, this is similar to the 52% identity seen between placental and L/B/K ALP-deduced amino acid sequences (12).

The placental and intestinal ALP genes both map to the same region of chromosome 2, while the L/B/K ALP gene maps to chromosome 1 (refs. 9 and 34; P.S.H., M.J.W., H.H., Constance Griffin, Beverly Emanuel, and Moyra Smith, unpublished data). The fact that the L/B/K ALP gene maps to a different chromosome than the placental and intestinal ALP genes, which are evidently closely-linked, is consistent with the immunological and sequence data indicating that the separation of the L/B/K ALP gene from the placental and intestinal ALP genes was, in evolutionary terms, a more ancient event than the separation of the placental and intestinal ALP genes. Chromosomal separation of members of a multigene family explicable in the same terms is well established (e.g., the globin genes).

Many features of this multigene enzyme family (see ref. 2 for review) are yet to be explored at the RNA and DNA levels. For example, it is known that human testis and thymus contain small amounts of an ALP closely resembling, but not identical to, placental ALP. This ALP is usually referred to as placental-like ALP or the Nagao isozyme. Its relationship to placental ALP is not yet clear, but it seems probable that it is the product of a separate locus (35). There is also evidence that the adult and fetal forms of intestinal ALP may be determined by separate loci (7, 14, 36). Thus it is possible that at least five different loci are involved in coding for the various ALPs. Analyses at the genomic level will no doubt resolve some of these questions and also provide an approach for understanding the remarkable differences in tissue expression between the various ALP gene loci (2).

We thank Ms. Joanne L. Borthwell for her technical assistance in the preparation of this manuscript and Drs. Nobuyo Maeda and Morris Goodman for helpful discussions. We are indebted to Mr. M. Bewick of Guys Hospital, London, for the sample of jejunum. This work was supported by National Institutes of Health Grant GM 27018 and March of Dimes Grant 858. P.S.H. is supported by National Institutes of Health Training Grant HD 07107 to the Division of Biochemical Development and Molecular Diseases, Children's Hospital of Philadelphia. M.J.W. is supported by National Institutes of Health Scientist Training Program Grant GM 07170.

1. McComb, R. B., Bowers, G. N., Jr., & Posen, S. (1979) *Alkaline Phosphatase* (Plenum, New York).
2. Harris, H. (1982) *The Harvey Lectures: Series 76* (Academic, New York), pp. 95–124.
3. Mulivor, R. A., Plotkin, L. I. & Harris, H. (1978) *Ann. Hum. Genet.* **42,** 1–13.
4. Seargeant, L. E. & Stinson, R. A. (1979) *Nature (London)* **281,** 152–154.
5. McKenna, M. J., Hamilton, T. A. & Sussman, H. H. (1979) *Biochem. J.* **181,** 67–73.
6. Lehmann, F.-G. (1980) *Biochim. Biophys. Acta* **616,** 41–59.
7. Vockley, J., D'Souza, M. P., Foster, C. J. & Harris, H. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 6120–6123.
8. Hua, J.-C., Berger, J., Pan, Y.-C. E., Hulmes, J. D. & Udenfriend, S. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 2368–2372.
9. Kam, W., Clauser, E., Kim, Y. S., Kan, Y. W. & Rutter, W. J. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 8715–8719.
10. Millan, J. L. (1986) *J. Biol. Chem.* **261,** 3112–3115.
11. Henthorn, P. S., Knoll, B. J., Raducha, M., Rothblum, K. N., Slaughter, C., Weiss, M., Lafferty, M. A., Fischer, T. & Harris, H. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 5597–5601.
12. Weiss, M. J., Henthorn, P. S., Lafferty, M. A., Slaughter, C., Raducha, M. & Harris, H. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 7182–7186.
13. Vockley, J. & Harris, H. (1984) *Biochem. J.* **217,** 535–541.
14. Vockley, J., Meyer, L. J. & Harris, H. (1984) *Am. J. Hum. Genet.* **36,** 987–1000.
15. Lehmann, F.-G. (1975) *Clin. Chim. Acta* **65,** 257–269.
16. Benham, F. J. & Harris, H. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 4016–4019.
17. Wray, L. K. & Harris, H. (1983) *Cancer Res.* **43,** 758–762.
18. Young, R. A. & Davis, R. W. (1985) in *Genetic Engineering,* eds. Setlow, J. & Hollaender, A. (Plenum, New York), Vol. 7, pp. 29–41.
19. Lloyd, J. C., Isenberg, H., Hopkinson, D. A. & Edwards, Y. H. (1985) *Ann. Hum. Genet.* **49,** 241–251.
20. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132,** 6–13.
21. Ezra, E., Blacher, R. & Udenfriend, S. (1983) *Biochem. Biophys. Res. Commun.* **116,** 1076–1083.
22. Mager, D. L. & Henthorn, P. S. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 7510–7514.
23. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
24. Dale, R. M. K., McClure, B. A. & Houchins, J. P. (1985) *Plasmid* **13,** 31–40.
25. Wilbur, W. J. & Lipman, D. J. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 726–730.
26. von Heijne, G. (1985) *J. Mol. Biol.* **184,** 99–105.
27. Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R. & Dodgson, J. (1980) *Cell* **20,** 555–566.
28. Koop, B. F., Goodman, M., Xu, P., Chan, K. & Slightom, J. L. (1986) *Nature (London)* **319,** 234–238.
29. Sibley, C. G. & Ahlquist, J. E. (1984) *J. Mol. Evol.* **20,** 2–15.
30. Koop, B. F., Miyamoto, M. M., Embury, J. E., Goodman, M., Czelusniak, J. & Slightom, J. L. (1987) *J. Mol. Evol.,* in press.
31. Goldstein, D. J., Rogers, C. & Harris, H. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 879–883.
32. Birnstiel, M. L., Busslinger, M. & Strub, K. (1985) *Cell* **41,** 349–359.
33. Sadofsky, M., Connelly, S., Manley, J. L. & Alwine, J. C. (1985) *Mol. Cell. Biol.* **5,** 2713–2719.
34. Swallow, D. M., Povey, S., Parkar, M., Andrews, P. W., Harris, H., Pym, B. & Goodfellow, P. (1986) *Ann. Hum. Genet.* **50,** 229–234.
35. Millan, J. L. & Stigbrand, T. (1983) *Eur. J. Biochem.* **136,** 1–7.
36. Mueller, H. D., Leung, H. & Stinson, R. A. (1985) *Biochem. Biophys. Res. Commun.* **126,** 427–433.