Mammalian Genome 9, 327–330 (1998).

Mammalian Genome
Incorporating Mouse Genome

Structure of the human biotinidase gene

Heath Cole Knight,[1,*] Thomas R. Reynolds,[5] Gregory A. Meyers,[5] Robert J. Pomponio,[1,] Gregory A. Buck,[4,5] Barry Wolf[1,2,3]**

[1]Department of Human Genetics, Medical College of Virginia at Virginia Commonwealth University, P.O. Box 980033, Richmond, Virginia 23298, USA
[2]Department of Pediatrics, Medical College of Virginia at Virginia Commonwealth University, Richmond, Virginia, USA
[3]Department of Biochemistry and Molecular Biophysics, Medical College of Virginia at Virginia Commonwealth University, Richmond, Virginia, USA
[4]Department of Microbiology and Immunology, Medical College of Virginia at Virginia Commonwealth University, Richmond, Virginia, USA
[5]Commonwealth Biotechnologies, Inc., Richmond, Virginia, USA

Received: 30 September 1997 / Accepted: 5 December 1997

Abstract. Biotinidase cleaves biotin from biocytin, thereby recycling the vitamin. We have determined the structure of the human biotinidase gene. A genomic clone, containing three exons that code for the mature enzyme, was obtained by screening a human genomic bacteriophage library with the biotinidase cDNA by plaque hybridization. To obtain a clone containing the most 5′ exon of the biotinidase cDNA, a human PAC library by PCR was screened. The human biotinidase gene is organized into four exons and spans at least 23 kb. The 5′-flanking region of exon 1 contains a CCAAT element, three initiator sequences, an octamer sequence, three methylation consensus sites, two GC boxes, and one HNF-5 site, but has no TATA element. The region from nt −600 to +400 has features of a CpG island and resembles a housekeeping gene promoter. The structure and sequence of this gene are useful for identifying and characterizing mutations that cause biotinidase deficiency.

Biotinidase (EC 3.5.1.12) catalyzes the release of biotin, an essential water-soluble B vitamin, from biocytin, the degradative product of four biotin-dependent holocarboxylases (Achuta Murthy and Mistry 1977; Wolf and Feldman 1982). Mammals cannot synthesize biotin and, therefore, must obtain the vitamin by processing protein-bound dietary biotin and from recycling endogenous biotin. Biotinidase deficiency, inherited as an autosomal recessive trait, results in a secondary biotin deficiency, which leads to multiple carboxylase deficiency (Wolf 1995; Wolf et al. 1983). This disorder is characterized by neurological and cutaneous symptoms that can be effectively treated with biotin supplementation (Wolf 1995). We have isolated and characterized the cDNA for normal human serum biotinidase (Cole et al. 1994a) and localized the gene to human Chromosome (Chr) 3p25 (Cole et al. 1994b). We now describe the structure of the human biotinidase gene.

To isolate a genomic clone containing the biotinidase gene, a human genomic bacteriophage library (Stratagene, LaJolla, Calif.) was plated at a density of 30,000 pfu/plate and approximately 7.2 × 10^5 recombinant bacteriophage were screened by plaque hybridization (Sambrook et al. 1989) with the cloned full-length biotinidase cDNA (BTD2000; Cole et al. 1994a). Three clones, with different inserts based on restriction mapping and Southern blotting (Sambrook et al. 1989), hybridized with BTD2000. Regions of the biotinidase gene within one of these lambda clones, #19, which has an insert of approximately 17 kb (Cole et al. 1994b), were sequenced as described previously (Cole et al. 1994a) on both strands, by automated DNA sequencing technology and 1 μg of purified bacteriophage lambda DNA (Sambrook et al. 1989). Gene-specific oligonucleotides, derived from the cDNA sequence and from newly determined intron sequences, served as primers in the sequencing reactions. Sequence analysis of this clone confirmed the sequence of the biotinidase cDNA and revealed that the entire coding region for the mature enzyme is contained within three exons: exon 2, 265 bp; exon 3, 150 bp; and exon 4, 1502 bp (Fig. 1A and B). Introns 2 and 3 are 6220 bp and 712 bp, respectively. This clone also contains 3562 bp 5′ of exon 2, but did not contain the first 79 bases of the biotinidase cDNA in which the first of two in-frame ATG codons is located (Cole et al. 1994a). Because these two ATG codons are in-frame and separated by an intron, it is possible that either or both codons initiate translation. Exon 2 contains the second putative translation initiation ATG codon. Translation from this second ATG is predicted to produce a leader peptide that is similar to the motif of other secretory signal peptides (Perlman and Halvorson 1983). The stop codon and polyadenylation signal are located in exon 4.

To determine whether the most 5′ 79 bases of the biotinidase cDNA are contained within one exon, we performed a step-down PCR (Hecker and Roux, 1996), using primers synthesized from the 5′ (1.S, 5′-GCCAGCTGGAGCGTTTTC-3′) and the 3′ (69.A, 5′-TCTTAGCGCGCCTTCC-3′) ends of the 79 bases (Fig. 2). Ten ng of the cloned biotinidase cDNA or 1 μg of genomic DNA served as templates in 50-μl reactions of 1 × PCR buffer, 2.5 units of AmpliTaq DNA polymerase (Perkin-Elmer, Foster City, Calif.), 2 mM each dNTP, and 2.5 mM $MgCl_2$. The temperature gradient for the step-down PCR was 70°C to 55°C at increments of 3°C with 6 cycles per temperature and 10 cycles at the final temperature. Analysis on a 2% agarose gel revealed that a 79-bp product was amplified from both the cDNA and genomic DNA templates (data not shown). The sequence analyses of the two PCR products were identical to the sequence reported for the 5′ end of the biotinidase cDNA and did include the first putative translation initiation ATG codon. The 79 bases, therefore, constitute a single exon (designated exon 1) (Fig. 1A and B).

To obtain a genomic clone containing the 5′ end of the biotinidase gene, a human PAC library (Genome Systems, Inc., St. Louis, Mo.) containing approximately 125,000 recombinant clones was screened by Genome Systems, Inc., using PCR with oligonucleotides 161.S (5′-GCAGGATTCTTTATTCAGCTG-3′) and 156.A (5′-GCAATCTGCTCTGTATGAGAG-3′) to amplify exon 2 of the biotinidase gene. A single clone, #14541, with an insert of approximately 100–120 kb, was obtained. Exon 1 and exon 2 were amplified with PCR as described above and in reference

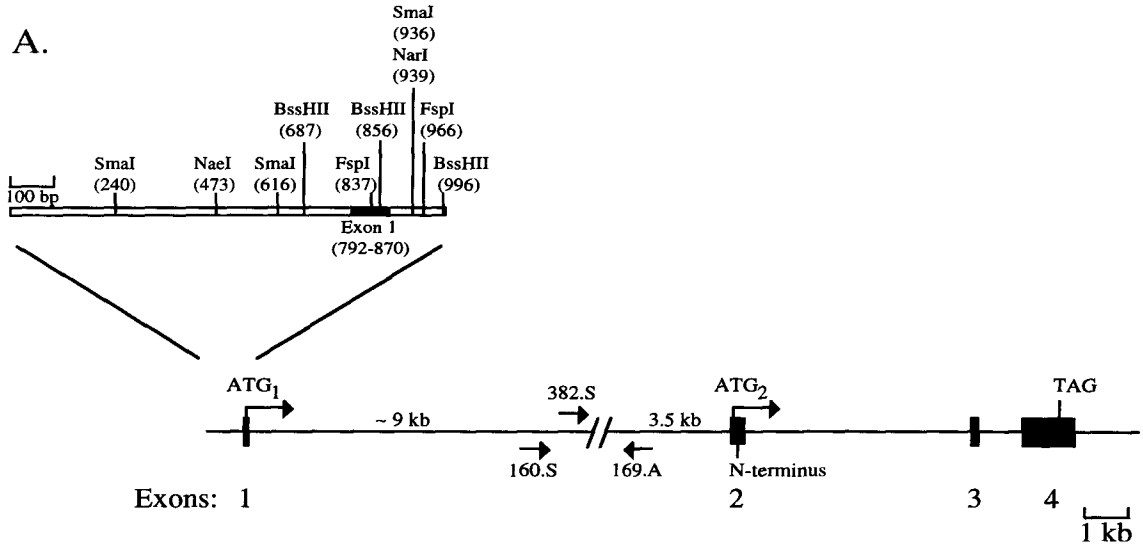
The nucleotide sequence data reported in this paper have been submitted to GenBank and have been assigned the accession numbers AF018630 and AF018631.

* *Present address:* Mayo Clinic at Jacksonville, Jacksonville, FL, USA

** *Present address:* Genzyme Genetics, Genzyme Corp., Framingham, MA, USA.

Correspondence to: B. Wolf

A.



B.

Exon number (letter)[a]	Size (bp)	cDNA position of exon	5′ Splice site	Intron number	Size (bp)	3′ Splice site
1 (A)	79	−35 to 44	AGAGCAG gtacgg	1	12,500	ccccattacattccag ATTTGT
2 (B)	265	45 to 309	CCAAAAG gcaaga	2	6220	tgttttcattttctag GATGTA
3 (C)	150	310 to 459	CACAGAG gtgatt	3	712	cctttttttcctctctag GTGCTC
4 (D)	1502	460 to 1961				

[a] Exons were previously designated by the letters A–D because the complete gene structure was unknown. The exons have now been reassigned numerical designations according to standard convention.

Fig. 1. Organization and exon-intron boundary sequences of the human biotinidase gene. **(A)** The four exons are shown as black boxes with two putative translation initiation ATG codons shown. Location of the N-terminus of the mature serum enzyme and TAG stop codon are indicated. The three introns are shown as horizontal lines between the exons. The // represents a gap in the sequence. The approximate locations of oligonucleotides 160.S, 169.A, and 382.S used in PCR and hybridization experiments are indicated. The restriction sites for rare-cutting enzymes near exon 1 are shown. Numbering for the restriction sites corresponds to the first base obtained in clone 5 (see text). **(B)** The 5′ and 3′ splice site sequences and the sizes of the four exons and introns 2 and 3 are noted. The total size of intron 1 is unknown at present, but is greater than 12.5 kb. Uppercase letters indicate exon sequence, and lowercase letters are sequences from introns. Polypyrimidine tracts were located in the introns just 5′ of the 3′ splice acceptor site.

Pomponio et al. (1995), and were found to hybridize with DNA of PAC clone 14541. Hybridization with the exon 1 probe indicates that this PAC clone contains the 5′ end of the biotinidase cDNA (Fig. 1A and B). Extensive restriction mapping of PAC DNA from clone 14541 with 21 enzymes, many of which are infrequent cutters of human DNA, failed to identify a single, discernible restriction fragment that hybridized with both exon 1 and exon 2 probes or with both oligonucleotides 160.S (5′-GTCAGGGTCATGGTTACTTG-3′) and 169.A (5′-AGAAACAGCAAGGACAGCAG-3′) (Fig. 1A); this indicates that these exons are far apart. Long-distance PCR, performed with primers 160.S and 169.A, 1 μg of genomic DNA, and the Gene Amp XL PCR kit (Perkin-Elmer) for 35 cycles at 95°C for 15 s and 60°C for 12 min, failed to produce a product that hybridized with oligonucleotide 382.S (5′-GTTGGTAG GACAAGTAGTCC-3′), which is 201 bases 3′ of oligonucleotide 160.S (Fig. 1A). This result also suggests exons 1 and 2 are far apart.

In order to obtain the sequences adjacent to exon 1, 1 μg of PAC clone 14541 DNA was digested with *NsiI* and "shotgun" cloned into the *PstI* site of pGem-3ZF+ (Promega, Madison, WI). Colony hybridization with the exon 1 probe identified a clone (#5) which contains exon 1, 791 bp of 5′-flanking sequence, and approximately 9 kb of sequence from intron 1 (Fig. 1A and B). The total size of intron 1 is unknown at present, but is greater than 12.5 kb, which consists of approximately 9 kb in the *NsiI* clone #5 (the most 5′ region of intron 1) and 3462 bp in lambda clone #19 (the most 3′ region of intron 1).

The human biotinidase gene is organized into four exons and spans at least 23 kb. The exon-intron boundary sequences conformed to the gt/ag rule (Jacob and Gallinaro 1989; Fig. 1B), the only exception being that the 5′ splice donor site of intron 2 is gc instead of gt, which occurs in approximately 1% of donor sites (Jackson, 1991).

The presence of an intron between the two putative translation initiation ATG codons may permit alternate splicing to occur. If exon 1, which contains the first ATG codon, is not included in the biotinidase transcript, only the putative secretory signal peptide (Cole 1994) would be present, facilitating secretion of the enzyme. However, if alternate splicing occurs and the larger putative signal peptide is present in the transcript, the enzyme may be targeted elsewhere. The presence of two promoters may also result in the production of two biotinidase transcripts that differ only in the signal peptide. Therefore, the regions 5′ of exon 1 and of exon 2 were examined for putative promoter elements.

The nucleotide sequences of exon 1, its 5′-flanking region, and the 5′ region of intron 1 of the human biotinidase gene are shown in Fig. 2. The 5′ end of the biotinidase cDNA is shown as the 5′ end of exon 1. The transcription initiation site has not yet been determined. The 5′-flanking region has features of a housekeeping gene promoter, including a high G + C content and lack of a consensus TATA sequence. The region from nt −600 to +400, including exon 1 and part of intron 1, shows common features of a CpG island (Bird 1986): an increase in the number of CpG dinucleotides [an average of 7.5/100 bp as compared with the typical density in bulk DNA of approximately 1/100 bp (Lewin 1977)], a clustering of CpG dinucleotides, and a high G + C

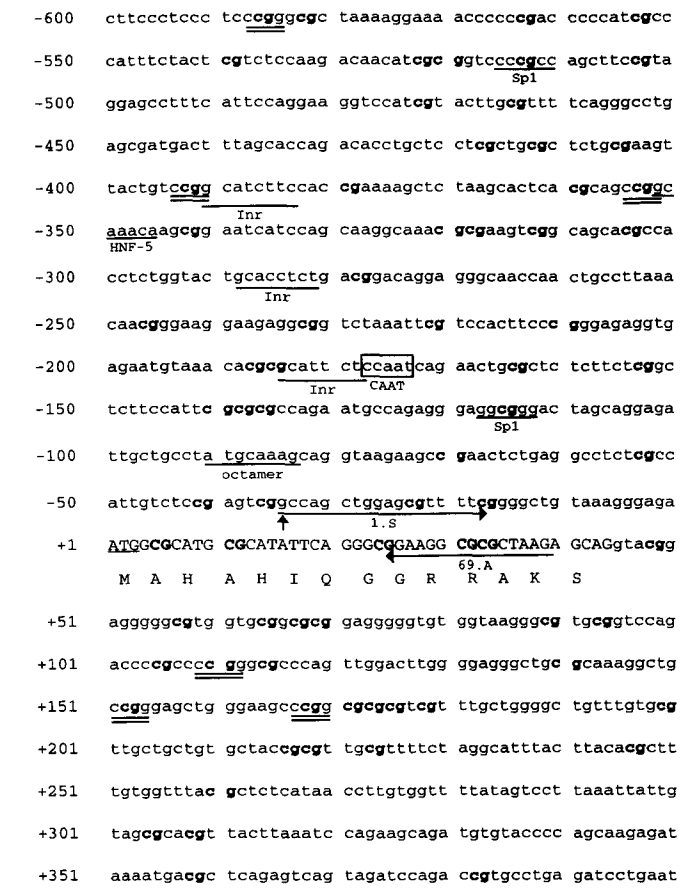


Fig. 2. Nucleotide sequence of exon 1, its 5′-flanking region, and the 5′ region of intron 1 of the human biotinidase gene derived from PAC clone #14541. The sequence is numbered starting from adenine of the first ATG (underlined), one of two putative translation initiation codons (see Fig. 1A and text). Uppercase letters indicate the coding region of exon 1, and lowercase letters are non-coding sequences. The amino acid sequence coded by exon 1 is shown below the nucleotide sequence. The 5′ end of the biotinidase cDNA is shown as the 5′ end of exon 1 (arrowhead). The transcription initiation site has not yet been determined. Sequences of primers 1.S and 69.A, used to amplify exon 1 by PCR (see text) are indicated. A consensus sequence for a CCAAT element is boxed. Initiator (Inr) sequences, an octamer sequence, and potential binding sites for transcription factors (Sp1 and HNF-5) are underlined. CpG dinucleotides are bolded and consensus methylation sequences are double underlined.

content (58%). In the region surrounding exon 1, restriction sites for enzymes that are rare cutters because their recognition sites contain one or more CpG dinucleotides, are present (*Nae*I and *Nar*I) or clustered (*Bss*HII, *Fsp*I, and *Sma*I; Fig. 1A). A CCAAT element (−178 to −174), an octamer sequence (−91 to −84), six consensus CCGG methylation sites (beginning at 163, 147, 104, −355, −394, and −588), and three initiator (Inr) sequences (beginning at −185, −289, and −391), are present in the 5′-flanking region of the biotinidase gene. Inr sequences are reported to be important in transcription initiation in promoters without a TATA element (Smale and Baltimore 1989). Consensus sequences for Sp1 (GC boxes, at −118 to −113 and −516 to −511) and the liver-specific transcription factor HNF-5 (GCAAACA, −352 to −346; Grange et al., 1990) are also present in this region. These elements in the putative promoter region of the biotinidase gene are consistent with the ubiquitous expression of biotinidase, with liver being the major site of enzyme production (Cole et al. 1994a; Pispa 1965).

The nucleotide sequence of the 3′ end of intron 1, which is 5′ of exon 2 and contains the second putative translation initiation ATG codon, was examined for potential promoter elements. No consensus TATA elements, CCAAT sequences, or higher G + C content were found. Consensus sequences for Ap1 and the liver-specific transcription factor C/EBP were present within 300 bases of the 5′ end of exon 2.

The structure and sequence of the human biotinidase gene are useful for identifying and characterizing mutations causing biotinidase deficiency. We have identified the two most common mutations that cause profound biotinidase deficiency (less than 10% of mean normal serum enzyme activity) in symptomatic children (Pomponio et al., 1995, 1997a). We have also identified other mutations that cause biotinidase deficiency (Norrgard et al. 1997a, 1997b; Pomponio et al. 1996; Pomponio et al. 1997b).

Acknowledgments. This work was partially supported by grant DK33022 from the National Institutes of Health (B. Wolf). The authors thank Dr. Joyce Lloyd for helpful advice during this project and Dr. Eric Westin and Allie Anderson of the Massey Cancer Center, Molecular Biology Core Facility, MCV/VCU (Richmond, Va.) for their assistance in the bacteriophage library screening.

References

Achuta Murthy PN, Mistry SP (1977) Biotin Prog Food Nutr Sci 2, 405–455

Bird AP (1986) CpG-rich islands and the function of DNA methylation. Nature 321, 209–213

Cole H (1994) Cloning and characterization of the human biotinidase gene (Richmond, Va.: Doctoral Thesis, Medical College of Virginia/Virginia Commonwealth University)

Cole H, Reynolds TR, Buck GB, Lockyer JM, Denson T, Spence JE, Hymes J, Wolf B (1994a) Human serum biotinidase: cDNA cloning, sequence and characterization. J Biol Chem 269, 6566–6570

Cole H, Weremowicz H, Morton CC, Wolf B (1994b) Localization of serum biotinidase (BTD) to human chromosome 3 in band p25. Genomics 22, 662–663

Grange T, Roux J, Rigaud G, Pictet R (1990) Cell-type specific activity of two glucocorticoid responsive units of rat tyrosine aminotransferase gene is associated with multiple binding sites for C/EBP and a novel liver-specific nuclear factor. Nucleic Acids Res 19, 131–139

Hecker KH, Roux KH (1996) High and low annealing temperatures increase both specificity and yield in touchdown and stepdown PCR. BioTech 20, 478–484

Jackson IJ (1991) A reappraisal of non-consensus mRNA splice sites. Nucleic Acids Res 19, 3795–3798

Jacob M, Gallinaro H (1989) The 5′-splice site: phylogenetic evolution and variable geometry of the association with U1RNA. Nucleic Acids Res 17, 2159–2180

Lewin B (1997) Genes VI (Oxford: Oxford University Press)

Norrgard KJ, Pomponio RJ, Swango KL, Hymes J, Reynolds TR, Buck GA, Wolf B (1997a) Double mutation (A171T and D444H) is a common cause of profound biotinidase deficiency in children ascertained by newborn screening in United States. Hum Mutat

Norrgard KJ, Pomponio RJ, Swango KL, Hymes J, Reynolds TR, Buck GA, Wolf B (1997b) Mutation (Q456H) is the most common cause of profound biotinidase deficiency in children ascertained by newborn screening in the United States. Biochem Mol Med 61, 22–27

Perlman D, Halvorson HO (1983) A putative signal recognition site and sequence in eukaryotic and prokaryotic signal peptides. J Mol Biol 167, 391–409

Pispa J (1965) Animal biotinidase. Ann Med Exp Biol Fenn 43 (Suppl 5), 1–39

Pomponio RJ, Reynolds TR, Cole H, Buck GA, Wolf B (1995) Mutational "hotspot" in the human biotinidase gene as a cause of biotinidase deficiency. Nat Genet 11, 96–98

Pomponio RJ, Narasimhan V, Reynolds TR, Buck GA, Povirk LF, Wolf B (1996) Deletion/insertion mutation that causes biotinidase deficiency may result from the formation of a quasipalindromic structure. Hum Mol Genet 5, 1657–1661

Pomponio RJ, Norrgard KJ, Reynolds TR, Hymes J, Buck GA, Wolf B (1997a) Arg538 to Cys mutation in a CpG dinucleotide of the human biotinidase gene is the second most common cause of biotinidase defi-

ciency in symptomatic children with biotinidase deficiency. Hum Genet 99, 506–512

Pomponio RJ, Reynolds TR, Mandel H, Admoni O, Melone PD, Buck GA, Wolf B (1997b) Profound biotinidase deficiency caused by a point mutation that creates a downstream cryptic 3′ splice acceptor site within an exon of the human biotinidase gene. Hum Mol Genet 6, 739–745

Sambrook J, Fritsch EF, Maniatis T (1989) Analysis and cloning of eukaryotic genomic DNA. In *Molecular Cloning: A Laboratory Manual.* (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press), pp 9.44–9.5

Smale ST, Baltimore D (1989) The "initiator" as a transcription control element. Cell 57, 103–113

Wolf B (1995) Disorders of biotin metabolism. In *The Metabolic and Molecular Bases of Inherited Disease,* CR Scriver, AL Beaudet, WS Sly, D Valle, eds. (New York: McGraw-Hill, Inc.), pp 3151–3180

Wolf B, Feldman GL (1982) The biotin-dependent carboxylase deficiencies. Am J Hum Genet 34, 699–716

Wolf B, Grier RE, Allen RJ, Goodman SI, Kien CL (1983). Biotinidase deficiency: the enzymatic defect in late-onset multiple carboxylase deficiency. Clin Chim Acta 131, 273–281