Am. J. Hum. Genet. 63:380–389, 1998

Identification of a New Splice Form of the *EDA1* Gene Permits Detection of Nearly All X-Linked Hypohidrotic Ectodermal Dysplasia Mutations

Alex W. Monreal, Jonathan Zonana, and Betsy Ferguson

Department of Molecular and Medical Genetics, Oregon Health Sciences University, Portland

Summary

X-linked hypohidrotic ectodermal dysplasia (XLHED), the most common of the ectodermal dysplasias, results in the abnormal development of teeth, hair, and eccrine sweat glands. The gene responsible for this disorder, *EDA1,* was identified by isolation of a single cDNA that was predicted to encode a 135–amino-acid protein. Mutations in this splice form were detected in <10% of families with XLHED. The subsequent cloning of the murine homologue of the *EDA1* gene (*Tabby* [*Ta*]) allowed us to identify a second putative isoform of the EDA1 protein (isoform II) in humans. This *EDA1* cDNA is predicted to encode a 391-residue protein, of which 256 amino acids are encoded by the new exons. The putative protein is 94% identical to the Ta protein and includes a collagen-like domain with 19 repeats of a Gly-X-Y motif in the presumptive extracellular domain. The genomic structure of the *EDA1* gene was established, and the complete sequence of the seven new exons was determined in 18 XLHED-affected males. Putative mutations, including 12 missense, one nonsense, and four deletion mutations, were identified in ~95% of the families. The results suggest that EDA1 isoform II plays a critical role in tooth, hair, and sweat gland morphogenesis, whereas the biological significance of isoform I remains unclear. Identification of mutations in nearly all of the XLHED families studied suggests that direct molecular diagnosis of the disorder is feasible. Direct diagnosis will allow carrier detection in families with a single affected male and will assist in distinguishing XLHED from the rarer, clinically indistinguishable, autosomal recessive form of the disorder.

Received April 21, 1998; accepted for publication June 8, 1998; electronically published July 10, 1998.

Address for correspondence and reprints: Dr. Jonathan Zonana, Department of Molecular and Medical Genetics, L-103, Oregon Health Sciences University, 3181 S.W. Sam Jackson Park Road, Portland, OR 97201-3098. E-mail: zonanaj@OHSU.edu

0002-9297/98/6302-0013$02.00

Introduction

X-linked hypohidrotic ectodermal dysplasia (XLHED; MIM 305100), the most common form of the ectodermal dysplasias, results in the abnormal development of teeth, hair, and eccrine sweat glands (Clarke 1987; Zonana 1993). It has been suggested that the associated morphological defects result from a failure of cell signaling or cell migration during the epithelial-mesenchymal inductive process (Holbrook 1988). Identification of the *EDA1* gene is of importance in the understanding of the molecular basis of XLHED, as well as the molecular mechanisms involved in normal tooth, hair, and eccrine–sweat gland development. It also should permit direct mutation testing for XLHED in potentially affected males and carrier females.

Heterozygous carriers of XLHED may have minor or moderate degrees of hypodontia, hypotrichosis, and hypohidrosis, although many show no obvious clinical manifestations (Pinheiro and Freire-Maia 1979). Because of this clinical variation, presumably caused by random X inactivation (Lyonization), accurate diagnosis of carrier females is difficult. Although indirect testing for carrier status by linkage analysis is possible in informative families (Zonana 1993), carrier detection by this method is impossible in families with single affected individuals, male or female, whose disorder may be the result of a de novo mutation. Detection of mutations within the *EDA1* gene is also of importance for families with only a single affected sibship, since a rarer, autosomal recessive form of the disorder is clinically indistinguishable from XLHED, in affected males (Munoz et al. 1997).

The *EDA1* gene recently was isolated by positional cloning (Kere et al. 1996). A single 858-bp cDNA, representing a full-length transcript composed of two exons, was identified from an adult–sweat gland cDNA library. In situ analysis showed that the *EDA1* gene is expressed in hair follicles and in the epidermis of adult skin. The putative gene product is a 135–amino-acid protein (isoform I) that has no clear homology to other proteins. The protein is predicted to contain a single transmembrane domain, and fractionation studies of transfected cell lines showed that the protein product is localized to

the plasma membrane (Ezer et al. 1997). Three lines of evidence suggested that there are unidentified alternative transcripts of the *EDA1* gene. First, although the full length of the *EDA1* cDNA was 858 bp, a predominant 5–6 kb transcript was detected, by northern analysis, in several human tissues (Kere et al. 1996). Second, mutation screening of 173 unrelated families with XLHED showed that only 7% of the families were likely to have mutations within exon 1, and none had variants within exon 2 (Ferguson et al. 1998). Finally, cDNAs from the homologous murine gene, *Tabby* (*Ta*), recently were isolated by our group and by others and were found to include alternative exons (Ferguson et al. 1997; Srivastava et al. 1997). The *Ta* cDNA was 87% identical to exon 1 of *EDA1* but was followed by an alternative sequence, lengthening the predicted open reading frame (ORF) by 768 bp. As seen with the *EDA1* cDNA, the *Ta* cDNA was found, by northern analysis, to hybridize to a 5–6 kb transcript. The anticipated *Ta* protein is 391 amino acids and, as with the EDA1 protein, contains a likely single transmembrane domain. In addition, the predicted Ta protein includes a short collagen-like domain, $(Gly\text{-}X\text{-}Y)_{19}$.

In this article, we report the isolation of an *EDA1* cDNA splice form that is homologous to the *Ta* cDNA and that codes for a putative, second isoform of the EDA1 protein (isoform II). Nearly all of the mutations associated with XLHED are located within the exons identified in this new splice form. These results provide evidence that EDA1 isoform II is essential for hair, tooth, and eccrine–sweat gland morphogenesis. In addition, the identification of the additional exons permits direct molecular diagnostic testing for XLHED, by mutation analysis.

Patients and Methods

Patients Studied

Affected males from 18 unrelated families with presumed XLHED were selected for mutation analysis of the *EDA1* gene. Genomic DNA had been extracted previously and had been analyzed for mutations within exons 1 and 2 of the *EDA1* gene, but no mutations had been found (Ferguson et al. 1998). All affected males had the classic XLHED phenotype, including tooth, hair, and sweat gland abnormalities. Ten of the families showed vertical transmission of the trait, with two or more affected generations. Eight families had a single affected generation, with two affected brothers in three of the families and a single affected male in five of the families. None of the families was consanguineous or had severely affected females. When applicable, DNA samples from other family members were studied to determine if the variants detected represented de novo mutations. Appropriate informed consent was obtained from all subjects included in this study, in accordance with the institutional review board guidelines of Oregon Health Sciences University.

Isolation and Analysis of EDA1 *cDNAs*

A human cDNA library constructed by means of mRNA isolated from 20–22-wk fetal liver tissue was obtained from Clontech. The DNA primers used for PCR amplifications are listed in table 1. PCR reactions for the cDNA amplifications included 16.6 mM $(NH_4)SO_4$, 67 mM Tris (pH 8.8), 6.7 mM $MgCl_2$, 170 μg BSA/ml, 6.7 μM EDTA, 10 mM β-mercaptoethanol, 10 ng cDNA, 12.5 pmol of each primer, 25 nM dNTP, and 0.75 U AmpliTaq polymerase (Perkin-Elmer). Reactions were cycled at 92°C for 30 s, 58°C for 30 s, and 72°C for 1 min, for 32 cycles. PCR products were electrophoresed on agarose gels, purified by use of GeneClean (Bio101), and sequenced by use of *Taq* DyeDeoxy Terminators (Applied Biosystems). Sequencing reactions were run on a Model 373A DNA sequencer (Applied Biosystems). Alignments of nucleotide and protein sequences were performed by use of the FASTA program (Pearson 1994). Nucleic acid sequence-similarity searches, which were compared against the nonredundant sequence and expressed-sequence-tag databases of the National Center for Biotechnology Information, were performed by use of the BLAST program (Altschul et al. 1990).

Analysis of EDA1 *Gene Structure*

With the exception of exons 1–3, PCR amplification between all exons was achieved by use of 1 U of rT^{th} DNA polymerase XL (Perkin-Elmer) in a reaction mixture with 1 × Buffer 2, 1 mM $Mg(OAc)_2$, 800 μM dNTP, 30 pmol of each primer, and 10 ng of human YAC DNA template yWXD1341. Intron sizes were estimated by agarose-gel electrophoresis of the PCR products. DNA primer sequences are listed in table 1. The reactions were heated to 94°C for 1 min and to 60°C for 5 min, for a total of 30 cycles, followed by incubation at 72°C for 10 min. Vectorette PCR was used to obtain the intron sequence flanking exon 3 (Arnold and Hodgson 1991). PCR products containing the intron/exon boundaries were sequenced as described in the preceding section. Exons 2, 3, and 9 were physically mapped by use of human YACs yWXD1851, yWXD3583, and yWXD1341, which were obtained from the American Type Culture Collection. Amplification of the exons from these YACs was performed by use of the *Taq* PCR conditions described in the preceding section.

Table 1

DNA Primers Used in This Study

Amplification Type and Exon(s)	Forward Primer (5′→3′)	Reverse Primer (5′→3′)	Sequencing Primer(s) (5′→3′)[a]	Method[b]
cDNA:				
1–3	ACCTCTGGCACCCTAAGCAGC	GCCATCTGCTCCTTCAT	…	…
1–5	Same as for exons 1–3	CCAGGGGGTCCTTGAGG	…	…
1–7	Same as for exons 1–3	TTGAATTGCTGACCCTTGGC	…	…
1–9	Same as for exons 1–3	CTGGGAAGTCACATAGGCC	…	…
Interexon, for genomic structure:				
3–4[c]	TTGTAATTTTTACAGATGGCCC	ACAGACAGACAATGCTGAAAG	…	…
4–5	GCCCAGTTAAAAACAAGAAAAAG	GGTCCTGGAGGGCCATTG	…	…
5–6	GACCACCTGGTCCTCCA	TGGTTTTCTCGAGTTCCAGC	…	…
6–7	GCTGGAACTCGAGAAAACCA	TTGAATTGCTGACCCTTGGC	…	…
7–8	GCCAAGGGTCAGCAATTCAA	CACCTTGGGGTTCATAGTGA	…	…
8–9	TCACTATGAACCCCAAGGTG	GGCAAAGTCAGTGAAGTTGATG	…	…
Exon, for mutation screening:				
1[d]	GTCGGCCGGGACCTCCTC	GCCGCCGCCCCTACTAGG	F: GGTGACTGGTGATGGGGCTGTC R: AGTTGCGCTCGGAGTTG	S
2[e]	TGGCTTCTCTAGTTAGGTTGGG	CATCTCAAATTTTCCTTCTGGG	F: TGGCTTCTCTAGTTAGGTTGGG R: CATCTCAAATTTTCCTTCTGGG	S
3	GGGCTCAGGGTTTAGACACA	CTCTATAAGGGCCTCATCTC	F: TTGTAATTTTTACAGATGGCCC R: GCTCACGCCTGTAATCCTA	S
4	GTGGCCTCAGGAGTCAGAAG	GAAAGAGGTGAATCATCACTGAA	R: ACAGACAGACAATGCTGAAAG	S
5[f]	CGGGAGGTGGAGGTTTCA AGTGAACCGAGATTGTGCCA	CACCTTAGGGTTCATAGTGA TGGTTTTCTCGAGTTCCAGC	F: GAACTCCAGCCTGGGCAA R: CTCTCAGGATCACCCACTC	XL
6[g]	AGAAAGCAGGACCTCCTGG GACCACCTGGTCCTCCA	CACCTTAGGGTTCATAGTGA CACCTTAGGGTTCATAGTGA	F: GGGGTGCACTCTGACTCTTC R: GAAAACCGTCAGAATCTCCG	XL
7[g]	Same as for exon 6	Same as for exon 6	F: CAGGGGAGAGGGATCAGAAT R: GGGGAGAAGCTCCTCTTTG	XL
8	GCCAAGGGTCAGCAATTCAA	GCACCGGATCTGCATTCTGG	F: GGGTTGTGAACTCCTTGGTA R: GAAGAGTTAGGCCTAAGACC	XL
9	TCACGTGTCCTTTCCTGTTG	TTGTCACCCTGGAGTCACT	F: TCACGTGTCCTTTCCTGTTG R: CACAGCAGCACTTAGAGG	S

[a] F = forward; R = reverse.

[b] S = standard PCR amplification using *Taq* DNA Polymerase or *rTth* DNA Polymerase; XL = extra-long amplification.

[c] The primers are located in the intronic sequence 5′ of exon 3 (forward) and in the intronic sequence 3′ of exon 4 (reverse).

[d] Primers previously published by Kere et al. (1996).

[e] Primers previously published by Ferguson et al. (1997).

[f] Product was obtained by nested PCR amplification.

[g] Product was obtained by hemi-nested PCR amplification.

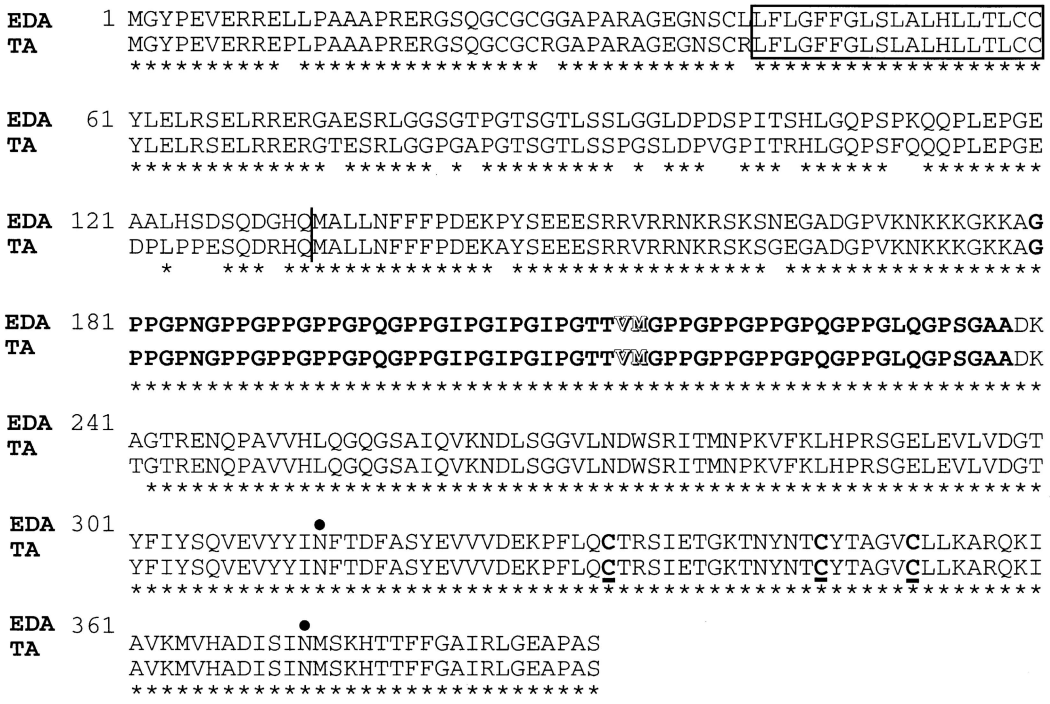

Figure 1 Comparison of the presumptive EDA1 isoform II and Ta proteins. Amino acid identities are indicated by an asterisk (*). The putative transmembrane domain is boxed. A vertical line designates the start of the protein sequence that is unique to isoform II. The G1y-X-Y domain is indicated by boldface type, with the 2–amino-acid interruption indicated by shadowed lettering. A blackened circle is shown above two potential N-linked glycosylation sites, and both C-terminal cysteines are indicated by underlining and boldface type.

Northern Blot Analysis

Human Multi-Tissue northern blots I and IV (Clontech) were hybridized with a ^{32}P-labeled PCR product that included exons 1–9 (the PCR primers shown in table 1). The filters were hybridized and washed in accordance with the manufacturer's recommendations. The signals were visualized by use of Kodak X-Omat film, with an exposure time of 3 d.

Mutation Analysis

Exons were amplified from genomic DNA by use of the primers listed in table 1 and the reaction conditions with AmpliTaq polymerase, described above. Amplicons were designed to include $\geqslant$28 bp of the 3′ and 5′ flanking intronic regions. PCR products were treated with Shrimp Alkaline Phosphatase and Exonuclease I (Amersham), to remove primers and excess nucleotides, and DNAs were sequenced by use of the RediVu Thermosequenase kit (Amersham). Sequencing reactions were visualized by electrophoretic separation on an 8% polyacrylamide gel and subsequent exposure on Kodak X-Omat film for 1–3 d. DNA-sequence variants were verified by automated sequencing using an ABI Model 373A sequencer (Applied Biosystems). Allele-specific oligonucleotide (ASO) analysis was performed, to detect polymorphisms in an unaffected population and to identify de novo mutations in a subset of families, as described in previous studies (Ferguson et al. 1998). One exception, the G1136A mutation, was analyzed by restriction digestion of PCR-amplified exon 8 DNAs, using *Ban*I (New England Biolabs). In all cases, genomic DNA samples from 40 control individuals, representing 60 X chromosomes, were analyzed. The de novo status of Del794–829 in families ED1050 and ED1204 was established by PCR amplification of exon 3 sequences, using primers 5′-AGA AAG CAG GAC CTC CTG G-3′ and 5′-CTC TCA GGA TCA CCC ACT C-3′, and by separation of the products on a 3% agarose gel (SeaKem).

Results

Identification of a New EDA1 *Splice Form*

A novel *EDA1* cDNA was PCR amplified from a human fetal liver library, by use of primers derived from the sequence of the homologous murine gene, *Ta* (GenBank accession number AF004435). The 1.5-kb human cDNA included 610 bp of sequence that was identical to the first exon of a previously identified 0.8-kb *EDA1* cDNA (Kere et al. 1996), followed by a 930-bp unique DNA sequence. This cDNA included a

Table 2

Intron-Exon Boundaries of Human *EDA1* Splice Form II

Exon	Exon Size (bp)	cDNA (nt)	5′ Intron/Exon…Exon/Intron 3′	Intron	Intron Size (kb)
1	637	1–637	ATTCCC…CACCAG/gtgagt	1b	>300
3	106	638–743	ttacag/ATGGCC…CAGATG/gtaagt	2	ND[a]
4	23	744–766	ttatag/GCCCAG…AAAAAG/ggtaag	3	5.0
5	181	767–947	ttttca/GGAAAG…CTTCTG/gtgagt	4	1.6
6	35	948–982	ttgcag/GTGCTG…AACCAG/gttggc	5	1.0
7	52	983–1034	tgccag/CCAGCT…AGAATG/gtaaga	6	2.8
8	131	1035–1165	actgag/ATCTTT…GTAGAA/gtgagt	7	2.0
9	252	1166–?	ttcgag/GTATAC…	…	…

[a] ND = not determined.

1,173-bp ORF, followed by a 160-bp 3′ UTR (Genbank AF060999). No polyadenylation signal sequence was identified. The cDNA is predicted to encode a 391-residue protein, of which 256 amino acids are encoded by new exons. The predicted protein is 94% identical to the Ta protein and includes a collagen-like domain with 19 repeats of a Gly-X-Y motif, interrupted by two amino acids between repeats 11 and 12 (fig. 1).

Gene Structure

The exon boundaries and the flanking sequences of all introns were established by use of either interexon or vectorette PCR amplification, followed by direct sequencing of the intron/exon junctions (table 2; GenBank AF060992–AF060998). A total of seven new exons in the *EDA1* gene were identified (fig. 2 and table 2). Exons 3–9 are within relatively close proximity to one another, with introns of 1–5 kb in length. The intron between exon 1 and exon 3 is estimated to be $\geqslant$300 kb, as was deduced by analysis of YACs that had been mapped previously to the *EDA1* region (YACs yWXD1850, yWXD3583, and yWXD1341 map from centromere to telomere and from 5′ to 3′ of *EDA1*) (Srivastava et al. 1996). By means of both Southern (Ferguson et al. 1997) and PCR analysis, exons 3–9 were localized to yWXD1341 but not to yWXD1850 or yWXD3583 (data not shown). The previously identified exon 2 (Kere et al. 1996), which is not present in this splice form, is present on all three YAC clones. When considered together, these data suggest that the newly identified exons are 3′ and telomeric to exon 2 and that they map $\geqslant$300 kb from exon 1. Analysis of the murine (*Ta*) gene structure revealed that the exon boundaries of the *EDA1* and *Ta* genes are completely conserved (data not shown).

Expression Studies

Northern blot analysis of RNA isolated from several human tissues, hybridized with exons 1–9 of the cDNA, detected a 5–6 kb transcript (fig. 3). No additional hybridization was observed with longer exposures. The most abundant signal was detected in RNA isolated from adult heart, pancreas, prostate, and uterus tissues. A weaker signal was seen in RNA isolated from muscle, spleen, thymus, testis, and small intestine tissues. To identify any rare, alternative *EDA1* splice products in the fetal liver library, the cDNA was amplified with DNA primers between exon 1 and exon 3, 5, 7, or 9. All PCR products analyzed had identical sequences, with no differences in exon usage. To identify splice products that contain exon 1 but not the novel exons, the library was PCR amplified by use of primers from either exon 1 or exon 3, in the 3′ direction, in combination with library vector primers. All the PCR products generated with the exon 1 primer had corresponding products that were produced with the exon 3 primer, suggesting that all cDNAs with exon 1 sequences also have exon 3 sequences (data not shown). Similarly, all cDNAs with exon 3 sequences appear to contain exon 1 sequences, since 5′-directed primers from exon 1 or exon 3, in combination with vector primers, generated corresponding PCR products.

Finally, since *Ta* cDNA had been isolated in two forms, which varied by the presence or absence of nucleotides 958–1000 (Ferguson et al. 1997; Srivastava et al. 1997), we tested the possibility that the human cDNA also might have a 42-bp variant. Primers closely flanking the 42-bp region were used to amplify the human cDNA library. Only a single 190-bp cDNA product, which included the 42-bp variable region, was identified. Analysis of the intron/exon boundaries shows that these 42 bp are contiguous with exon 8 in genomic DNA; thus, they do not represent a separate exon but, rather, result from the use of an alternate splice donor site.

Mutations Identified

Mutation analysis was conducted on genomic DNAs isolated from 18 families with XLHED. The families were selected to include some with vertical transmission of the disease and others with sporadic cases. Every exon, including flanking splice recognition sequences,

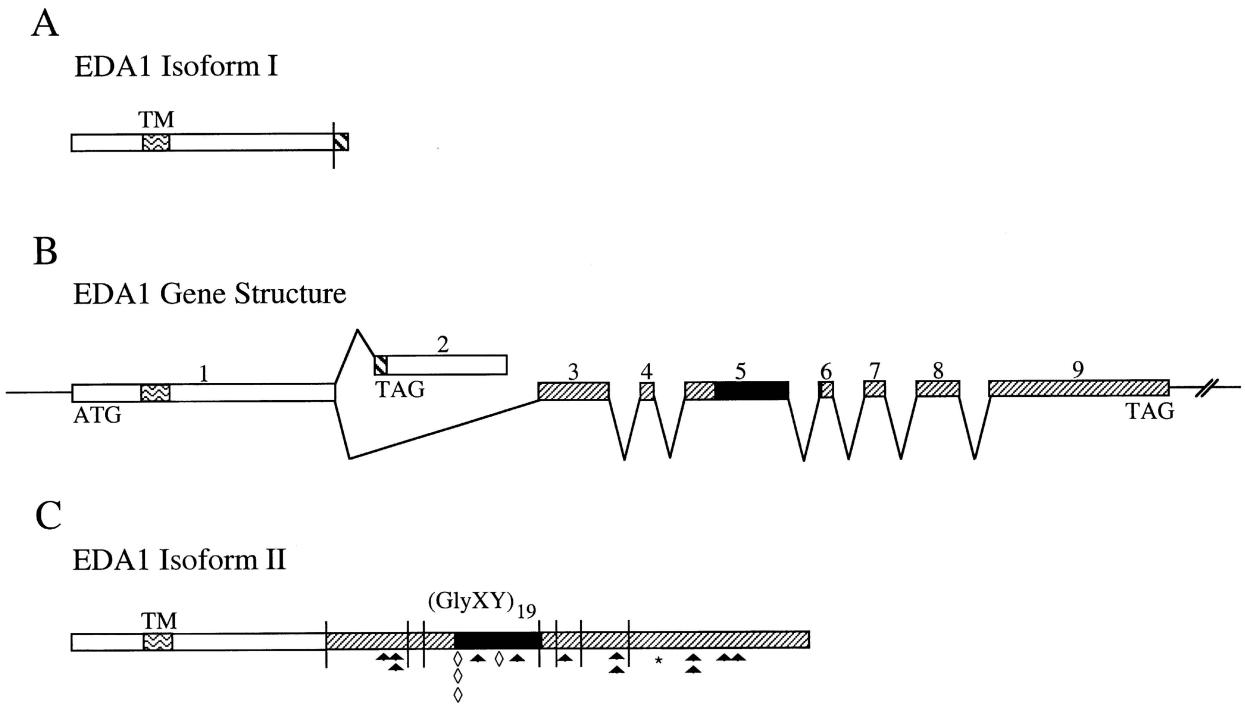

Figure 2 Gene structure and predicted protein products of *EDA1*. The anticipated protein products of the two splice forms are depicted in panels *A* and *C*. The vertical lines separate protein regions encoded by different exons. Transmembrane ("TM") and collagen-like (Gly-X-Y) domains are shown. The relative positions of the mutations identified in this article are depicted in panel *C*, with missense mutations indicated by blackened triangles, deletions indicated by unblackened diamonds, and a nonsense mutation indicated by an asterisk (*). Mutations in exon 1, found in a previous study, are not included in this diagram. The *EDA1* gene structure is depicted in panel *B*, with the numbered boxes representing exons and the connecting lines representing intronic regions. The start and stop codons are indicated by "ATG" and "TAG."

was PCR amplified and sequenced. Sequence variants were confirmed by sequencing the opposing strand of DNA. DNA changes were confirmed in 17 individuals (tables 3 and 4). Four individuals were found to have genomic deletions, two of which also produced frameshift mutations. Twelve individuals had a single base-pair change that was predicted to generate a missense mutation, and one person had a base-pair substitution that created a premature termination codon. To determine whether any of the putative missense mutations identified are present in an unaffected population and thus are likely to be a nonpathogenic polymorphism, we used ASO hybridization to survey 60 X chromosomes. None of the variants was found to be present in the control population. Three presumed mutations (Del794–829, G1136A, and G1285A) each were identified twice among the 18 affected individuals. For two of these cases (Del794–829 and G1285A), a de novo mutation was found in one of the families (ED1204 and ED1126, respectively), which argues against inheritance from a common ancestor (fig. 4; for G1285A, data not shown). Additional sequence variants were identified within intronic regions. Because these changes did not appear to affect the splice recognition sites and also were present in unaffected individuals, they were classified as nonpathogenic polymorphisms (table 4).

Discussion

Novel EDA1 *Splice Form*

We have identified a novel *EDA1* cDNA containing a complete ORF that we suggest encodes a novel isoform of the EDA1 protein (isoform II). Isoform II is highly homologous to the recently described Ta protein. The full-length transcript is likely to include a longer 3′ UTR, since no polyadenylation signal sequence was identified, and northern analysis indicates that the transcript is 5–6 kb in length. The *Ta* transcript is also 5–6 kb in length (Ferguson et al. 1997) and has a 3′ UTR of ~3.5 kb (Srivastava et al. 1997).

This novel cDNA is expected to represent the major *EDA1* transcript, since only a single hybridization band was detected by northern blot analysis. The transcript

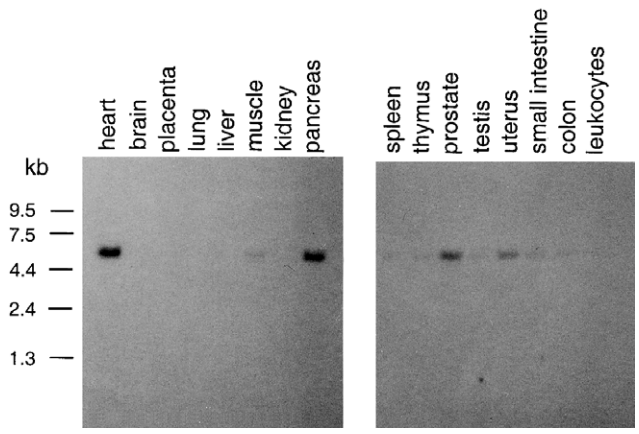

Figure 3 Northern analysis of *EDA1* gene expression in human tissues. An *EDA1* probe, including exons 1–9, was hybridized with polyA-enriched RNA extracted from the tissues indicated above the figure. The migration of RNA size markers is shown on the left.

is expressed in a variety of organs that are not clinically associated with hypohidrotic ectodermal dysplasia, suggesting that the *EDA1* protein may serve a redundant function in some tissues. No other *EDA1* splice products were identified in this study, including the original 0.8-kb full-length transcript encoding *EDA1* isoform I. The possibility that other splice forms exist in other tissues cannot be excluded. In the mouse, only the 5-kb homologous cDNA and its variants have been identified. One variant appears to result from the use of a cryptic splice donor site at the 3′ end of exon 7 (Ferguson et al. 1997), whereas two other variants include portions of intron 1 or intron 2 and are missing all downstream exons (Srivastava et al. 1997). The fact that 95% of XLHED mutations were detected within the newly identified exons suggests that isoform II is of critical importance and supports the notion that no additional exons in the *EDA1* gene are likely to be critical for normal development. The function of EDA1 isoform I, or whether it actually is translated, remains to be determined.

Predicted Protein Structure

The predicted EDA1 isoform II protein includes a single transmembrane domain and, like EDA1 isoform I, is likely to be a cell-surface protein with a type II orientation (the C-terminus projecting extracellularly) (Ezer et al. 1997). The predicted collagen domain $(Gly\text{-}X\text{-}Y)_{19}$ may form a triple helix with either itself or the collagen domains of other proteins, resulting in the formation of either homotrimeric or heterotrimeric complexes. The three cysteine residues that present in the C-terminal domains of both the mouse and the human proteins may serve to stabilize such trimeric complexes. Although the protein sequence itself does not suggest a function, we expect that the EDA1 protein may play a role in intercellular signaling, as either a membrane-bound ligand or a cell-surface receptor or coreceptor, or that it may serve a function in cell adhesion or cell migration, through interactions with the extracellular matrix. Other membrane-associated proteins with collagen domains include collagen XIII and XVII, a macrophage scavenger receptor, and a bacteria-binding macrophage receptor (Rohrer et al. 1990; Acton et al. 1993; Gatalica et al. 1997; Peltonen et al. 1997).

An overall high degree of sequence conservation (94% identity) was found between the predicted EDA1 isoform II and Ta proteins. This conservation is particularly striking in the C-terminal 211 amino acids, which begin with the Gly-X-Y domain and in which only a single, conservative substitution at position 241 is present. The extreme conservation in the C-terminus suggests that it is of critical importance to protein function and that it may interact with other proteins that also are highly conserved.

Functional Implications of Identified Mutations

This study identified mutations within exons 3–9 of *EDA1* in 94.4% (17/18) of the families studied. When considered with the previous mutation analysis of exon 1, which found mutations in ~7%–8% of XLHED families (Ferguson et al. 1998), it was estimated that 95% of mutations in XLHED patients are located within the 1,173-bp coding region of this splice form. Additional mutations may be located in the remaining 3′ UTR se-

Table 3

***EDA1* Mutations in XLHED Patients**

Family	Sequence Change	Exon	Predicted Effect[a]
ED1081	C704T	3	R155C
ED1095[b]	C707T	3	R156C
ED1039	G708A	3	R156H
ED1011	C867T	5	P209L
ED1019	G912C	5	G224A
ED1050	Del794–829	5	Del185–196
ED1204[b]	Del794–829	5	Del185–196
ED1018[b]	Del803–830	5	Del188–197, FS 198, Ter 280
ED1097[b]	Del904–938	5	Del221–233, FS 234, Ter 240
ED1197	A986T	7	H252L
ED1007	G1136A	8	G299S
ED1002	G1136A	8	G299S
ED1001	G1202T	9	E321Ter
ED1021	G1285A	9	A349T
ED1126[b]	G1285A	9	A349T
ED1073	C1308A	9	A356D
ED1022	G1311C	9	R357P

[a] FS = frameshift; Ter = termination.
[b] De novo mutation.

Table 4

Polymorphisms Detected in the *EDA1* Gene

Intron	Position (5′+…−3′)	Variant
3	+28	T or A
4	+11, +12	CT or Del
5	−11	C or T

quence, in regulatory elements, or in unsequenced intronic regions or may include large genomic rearrangements that would have been undetectable by the techniques used in this study.

Although the identified putative mutations are heterogeneous, more than half the individuals (11/18) were found to have changes within the coding regions of exons 5 and 9. These two exons account for ~40% of the ORF. Six mutations were detected in exon 5, including four deletions 28–36 bp in length. Exon 5 encodes the Gly-X-Y domain, and deletions within it may be enriched because of the high frequency of repeated sequences. Indeed, one deletion (Del794–829) identified in two unrelated families is flanked by a 17-bp repeated sequence containing only two nucleotide differences. This deletion is predicted to truncate the collagen domain of the protein but otherwise leaves the protein intact. If the resulting mutant protein is stable, the implication is that the collagen domain is required for normal protein function. Two missense mutations may also disrupt the collagen domain. In one case, the glycine of a Gly-X-Y triplet is substituted by an alanine. In another variant, a third-position proline (Gly-X-Pro) is substituted by a leucine. The third-position proline plays an important role in stabilizing the collagen helix (Brodsky and Shah 1995). The coding region of exon 9 also contains a disproportionate share of mutations (all missense). Three changes cluster within an 8–amino-acid region, highlighting the importance of this domain for either protein stability or function. Also of note, this region overlaps with the C-terminal cysteine domain, which may stablilize intermolecular associations.

Finally, a mutation identified twice in exon 8, G1136A, is located within the 42-bp region previously found to be deleted in one of the *Ta* splice forms (Ferguson et al. 1997). This suggests that the 14–amino-acid region encoded by the sequence is important in at least some, if not all, tissues.

Diagnostic Applications

Linkage-based molecular diagnostic testing for XLHED has been available for ~10 years and can significantly improve carrier-risk estimates (Zonana et al. 1988, 1989, 1992). However, linkage analysis is limited to families with two or more affected individuals, and multiple family members must be available and cooperative. The results of this work, combined with those of our previous study of exons 1 and 2, indicate that ~95% of mutations can be detected in individuals with presumed XLHED and that direct mutation screening is now a practical option.

The finding that approximately half of all identifiable mutations are within exon 5 and the coding region of exon 9 suggests that initial analysis of these two exons would provide the most efficient approach in a mutation-detection protocol. Since these two coding regions are relatively small (181 bp and 252 bp, respectively), they would be appropriate for direct sequencing, screening by SSCP (Sheffield et al. 1993), dideoxy fingerprinting (Liu et al. 1996), or other mutation-detection methods (Grompe 1993). Amplification of exon 5 may reveal deletions large enough to be detectable by agarose-gel analysis, as was seen in four families in this study (fig. 4). If no mutations are present in exons 5 and 9, screening or sequencing of exons 1, 3, 7, and 8 should be performed and may detect an additional 40% of mutations. No mutations were detected in exons 4 and 6; however, they represent only ~5% of the coding sequence, and analysis of a larger set of patients may reveal mutations in these exons as well. In families in which mutations are not detected, linkage analysis still can be used when the family is informative.

Five families had a negative family history and an unaffected mother and, thus, potentially were informative as to whether the mutation was de novo. In all five families, the mutations (three deletions and two missense) were found to be de novo, thus indicating that they arose either during oogenesis or postzygotically. With regard to the latter possibility, there was no sign of mosaicism when sequence analysis was performed.

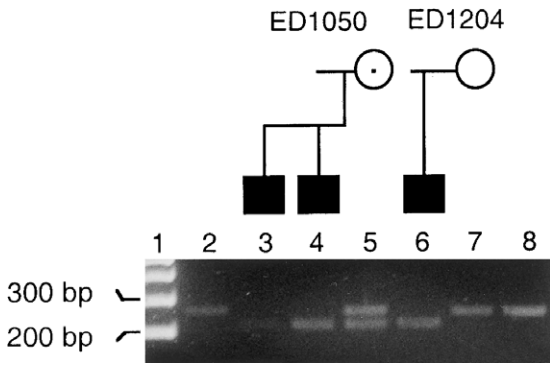

Figure 4 Agarose-gel analysis of Del794–829. Exon 5 was PCR amplified from samples from ED1050 and ED1204 family members and then was separated on a 3% agarose gel and stained with ethidium bromide. Circles represent females, with a dot indicating XLHED carrier status. Blackened boxes represent affected males. Molecular size markers are shown in lane 1, and control DNAs from unrelated, unaffected males are shown in lanes 2 and 8.

This is in contrast with the results of a previous study (Zonana et al. 1994) that indirectly assessed new mutations by use of marker haplotypes, which indicated that 7 of 9 de novo mutations occurred either during spermatogenesis or postzygotically. Although the numbers are small, if the results of the two studies are combined, there appear to be equal rates of mutation during oogenesis and spermatogenesis. The direct identification of de novo mutations significantly decreases the carrier risk of female relatives, barring gonadal mosaicism.

Putative mutations were detected in 9 of 10 families with two or more affected generations. The one family in which a mutation was not detected is the largest we have studied, with a history of seven-generation involvement, classic clinical findings of XLHED, and a multipoint LOD score of 2.70 for linkage to the *EDA1* locus (Zonana et al. 1992). Presumptive mutations were detected in all eight families with a single affected generation, a pedigree type that also would be compatible with the rarer, autosomal recessive form of the disorder. This finding supports previous observations that the autosomal recessive form of the disorder is infrequent, compared with the X-linked form. Review of clinical features showed no obvious phenotype/genotype correlation between individuals with missense mutations and those with either deletions truncating the collagen domain or frameshift mutations that may yield either an abnormal truncated protein or no protein at all. This is similar to our previous analysis of mutations within exon 1, which showed no difference in phenotype between patients with missense mutations and those with exon deletions or frameshifts (Kere et al. 1996; Ferguson et al. 1998).

The availability of direct testing for XLHED mutations is a significant advance in diagnostic capabilities. It will allow carrier detection in families with only a single affected male and will help differentiate the X-linked disorder from the rarer, autosomal recessive form of the disorder. It also should be helpful in distinguishing sporadic carriers of XLHED from females with an autosomal dominant form of isolated hypodontia (Arte et al. 1996). Finally, testing also should be useful in the diagnosis of XLHED in affected male infants, especially in sporadic cases, before abnormal tooth and hair involvement is apparent. Early diagnosis can avert the significant morbidity and mortality due to uncontrolled hyperthermia that is associated with the disorder (Clarke et al. 1987).

Acknowledgments

The authors are grateful to the many clinicians and families who participated in this study. We also thank R. Wildin and S. Hayflick for their careful reading of the manuscript. This work was supported by grant DE11311 from the National Institute of Dental Research (to J.Z.) and by the National Foundation for Ectodermal Dysplasias (support to J.Z.).

Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

GenBank, http://www.ncbi.nlm.nih.gov/Web/Genbank (for *EDA1* cDNA [AF060999]; *EDA1* exons 2–9, including the flanking intronic sequences [AF060992–AF060998]; and the murine gene *Ta* [AF004435])

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for XLHED [MIM 305100])

References

Acton S, Resnick D, Freeman M, Ekkel Y, Ashkenas J, Krieger M (1993) The collagenous domains of macrophage scavenger receptors and complement component C1q mediate their similar, but not identical, binding specificities for polyanionic ligands. J Biol Chem 268:3530–3537

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Arnold C, Hodgson IJ (1991) Vectorette PCR: a novel approach to genomic walking. PCR Methods Appl 1:39–42

Arte S, Nieminen P, Pirinen S, Thesleff I, Peltonen L (1996) Gene defect in hypodontia: exclusion of EGF, EGFR, and FGF-3 as candidate genes. J Dent Res 75:1346–1352

Brodsky B, Shah NK (1995) Protein motifs. 8. The triple-helix motif in proteins. FASEB J 9:1537–1546

Clarke A (1987) Hypohidrotic ectodermal dysplasia. J Med Genet 24:659–663

Clarke A, Phillips DI, Brown R, Harper PS (1987) Clinical aspects of X-linked hypohidrotic ectodermal dysplasia. Arch Dis Child 62:989–996

Ezer S, Schlessinger D, Srivastava A, Kere J (1997) Anhidrotic ectodermal dysplasia (EDA) protein expressed in MCF-7 cells associates with cell membrane and induces rounding. Hum Mol Genet 6:1581–1587

Ferguson BM, Brockdorff N, Formstone E, Ngyuen T, Kronmiller JE, Zonana J (1997) Cloning of Tabby, the murine homolog of the human EDA gene: evidence for a membrane-associated protein with a short collagenous domain. Hum Mol Genet 6:1589–1594

Ferguson BM, Thomas NS, Munoz F, Morgan D, Clarke A, Zonana J (1998) Scarcity of mutations detected in families with X linked hypohidrotic ectodermal dysplasia: diagnostic implications. J Med Genet 35:112–115

Gatalica B, Pulkkinen L, Li K, Kuokkanen K, Ryynänen M, McGrath JA, Uitto J (1997) Cloning of the human type XVII collagen gene (COL17A1), and detection of novel mutations in generalized atrophic benign epidermolysis bullosa. Am J Hum Genet 60:352–365

Grompe M (1993) The rapid detection of unknown mutations in nucleic acids. Nat Genet 5:111–117

Holbrook KA (1988) Structural abnormalities of the epidermally derived appendages in skin from patients with ectodermal dysplasia: insight into developmental errors. Birth Defects 24:15–44

Kere J, Srivastava AK, Montonen O, Zonana J, Thomas N, Ferguson B, Munoz F, et al (1996) X-linked anhidrotic (hypohidrotic) ectodermal dysplasia is caused by mutation in a novel transmembrane protein. Nat Genet 13:409–416

Liu Q, Feng J, Sommer SS (1996) Bi-directional dideoxy fingerprinting (Bi-ddF): a rapid method for quantitative detection of mutations in genomic regions of 300–600 bp. Hum Mol Genet 5:107–114

Munoz F, Lestringant G, Sybert V, Frydman M, Alswaini A, Frossard PM, Jorgenson R, et al (1997) Definitive evidence for an autosomal recessive form of hypohidrotic ectodermal dysplasia clinically indistinguishable from the more common X-linked disorder. Am J Hum Genet 61:94–100

Pearson WR (1994) Using the FASTA program to search protein and DNA sequence databases. Methods Mol Biol 24: 307–331

Peltonen S, Rehn M, Pihlajaniemi T (1997) Alternative splicing of mouse alpha1(XIII) collagen RNAs results in at least 17 different transcripts, predicting alpha1(XIII) collagen chains with length varying between 651 and 710 amino acid residues. DNA Cell Biol 16:227–234

Pinheiro M, Freire-Maia N (1979) Christ-Siemens-Touraine syndrome: a clinical and genetic analysis of a large Brazilian kindred. I. Affected females. Am J Med Genet 4:113–122

Rohrer L, Freeman M, Kodama T, Penman M, Krieger M (1990) Coiled-coil fibrous domains mediate ligand binding by macrophage scavenger receptor type II. Nature 343: 570–572

Sheffield VC, Beck JS, Kwitek AE, Sandstrom DW, Stone EM (1993) The sensitivity of single strand conformation polymorphism analysis for the detection of single base substitutions. Genomics 16:325–332

Srivastava AK, Montonen O, Saarialho-Kere U, Chen E, Baybayan P, Pispa J, Limon J, et al (1996) Fine mapping of the EDA gene: a translocation breakpoint is associated with a CpG island that is transcribed. Am J Hum Genet 58: 126–132

Srivastava AK, Pispa J, Hartung AJ, Du Y, Ezer S, Jenks T, Shimada T, et al (1997) The Tabby phenotype is caused by mutation in a mouse homologue of the EDA gene that reveals novel mouse and human exons and encodes a protein (ectodysplasin-A) with collagenous domains. Proc Natl Acad Sci USA 94:13069–13074

Zonana J (1993) Hypohidrotic (anhidrotic) ectodermal dysplasia: molecular genetic research and its clinical applications. Semin Dermatol 12:241–246

Zonana J, Clarke A, Sarfarazi M, Thomas NST, Roberts K, Marymee K, Harper PS (1988) X-linked hypohidrotic ectodermal dysplasia: localization within the region Xq11-21.1 by linkage analysis and implications for carrier detection and prenatal diagnosis. Am J Hum Genet 43:75–85

Zonana J, Jones M, Browne D, Litt M, Kramer P, Becker HW, Brockdorff N, et al (1992) High-resolution mapping of the X-linked hypohidrotic ectodermal dysplasia (EDA) locus. Am J Hum Genet 51:1036–1046

Zonana J, Jones M, Clarke A, Gault J, Muller B, Thomas NS (1994) Detection of de novo mutations and analysis of their origin in families with X linked hypohidrotic ectodermal dysplasia. J Med Genet 31:287–292

Zonana J, Sarfarazi M, Thomas NS, Clarke A, Marymee K, Harper PS (1989) Improved definition of carrier status in X-linked hypohidrotic ectodermal dysplasia by use of restriction fragment length polymorphism–based linkage analysis. J Pediatr 114:392–399