SHORT COMMUNICATION

Genomic Organization of the ATM Gene

TAMAR UZIEL, KINNERET SAVITSKY, MATTHIAS PLATZER,* YAEL ZIV, TAL HELBITZ, MICHAEL NEHLS,† THOMAS BOEHM,† ANDRE ROSENTHAL,* YOSEF SHILOH, AND GALIT ROTMAN[1]

*Department of Human Genetics, Sackler School of Medicine, Tel Aviv University, Ramat Aviv 69978, Israel; *Department of Genome Analysis, Institute of Molecular Biotechnology, D-07745 Jena, Germany; and †German Cancer Research Center, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany.*

Received December 5, 1995; accepted January 18, 1996

The ATM gene was recently identified and found to be responsible for the human genetic disorder ataxia-telangiectasia. The major ATM transcript is 13 kb. Using long-distance PCR, we determined the genomic structure of this gene and identified all of its exon–intron boundaries. The ATM gene spans approximately 150 kb of genomic DNA and consists of 66 exons. The initiation codon falls within exon 4. The last exon is 3.8 kb and contains the stop codon and a 3′-untranslated region of about 3600 nucleotides.

Ataxia-telangiectasia (A-T) is an autosomal recessive disorder characterized by progressive cerebellar degeneration, oculocutaneous telangiectasia, immunodeficiency, and cancer predisposition manifested primarily by lymphoreticular malignancies. A-T cells are hypersensitive to ionizing radiation and radiomimetic chemicals and are defective in cell cycle checkpoints activated by DNA damage (reviewed in Refs. 7, 8 and 15). A-T heterozygotes show moderate radiosensitivity and an increased susceptibility to cancer (5, 17, 18).

The gene responsible for A-T was recently cloned in our laboratory using a positional cloning approach and was designated ATM (13). The open reading frame of the ATM transcript is 9168 nucleotides and predicts a putative large protein of 350 kDa consisting of 3056 amino acids (14). The ATM gene product is a member of a novel family of large proteins that share a highly conserved carboxy-terminal region of about 300 amino acids that shows high similarity to the catalytic domain of PI-3 kinases (14, 19, and references therein).

The major ATM transcript of 13 kb (13, 14, and unpublished data) is observed in every tissue tested to date. Northern blots of several tissues, and analysis of cDNA clones and RT-PCR products from various tissues, failed to disclose any evidence of alternative forms within the coding sequence. However, in different tissues we have found extensive alternative splicing of noncoding 5′ exons resulting in a multitude of 5′-untranslated regions (UTRs) (K. Savitsky *et al.*, in preparation).

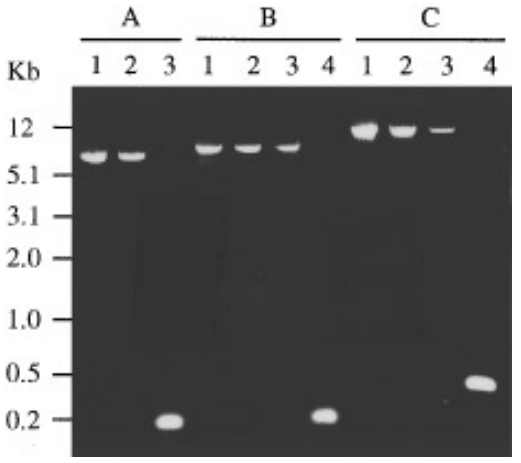

FIG. 1. Products of long-distance PCR using primers flanking introns 61 (A), 58 (B) and 63 (C). The following templates were used: Lanes 1 and 2, genomic and YAC DNA, respectively. Lanes B3 and C3, cosmid DNA. Lanes A3, B4, and C4, cDNA. Long-distance PCR was performed using the Expand Long Template PCR system (Boehringer Mannheim, Germany). Electrophoresis was performed on a 0.7% agarose gel.

We report here the genomic organization of the ATM gene and present its exon–intron boundaries. These boundaries and the intron sizes were determined using long-distance PCR (2, 3, 6). Primers were designed based on the ATM cDNA sequence (13, 14) at 200- to 300-bp intervals. Templates for these reactions were cosmid and YAC clones and human genomic DNA. PCR products were obtained in all cases, including those that span the largest intron of 11 kb. An example is shown in Fig. 1. In the large majority of cases, PCR products of the same size were obtained with all templates, and those obtained from genomic DNA were used for sequencing the exon–intron junctions. Following initial reactions, new primers were designed as needed, based on the evolving knowledge of the gene structure. Exon–intron boundaries were determined at the sites where genomic and cDNA sequences diverged. Typical splice acceptor and donor sequences were found

[1] To whom correspondence should be addressed. Telephone: 972-3-6408584. Fax: 972-3-6407471.

GENOMICS 33, 317–320 (1996)
ARTICLE NO. 0201
0888-7543/96 $18.00

TABLE 1

Exon–Intron Organization of the ATM Gene

Exon No.	5′ intronic sequence	Exon first base[a]		Exon length (bp)		Exon last base[a]	3′ intronic sequence	Size of 3′ intron (kb)
1a[b]		−915	AGGTAG	120	GCAGTG	−796	gtaggggcgcggaggcaacgcagcggcttc	0.18
1b[b]		−795	TCCTCC	634	TTTGAG	−162	gtatgagcggaagaagagatatcaggagac	0.09
2	tgatcattgctaacatttgctgtgtttcag	−161	GCACTG	43	ATAGAG	−119	gtaggtactagtattgtttttccttttatc	0.65
3	ttttctattactgtgtttttgtttcctcag	−118	AGGCAT	88	ATGAGG	−31	gtaggatttgtatctgtttagttcattatt	2.75
4	tatatatacctatatgtatttttttttacag	−30	ACAGTG	102	CGAAAG	72	gtagtaaattacttaaattcaatttttcct	0.08
5	aacccattattatttccttttttattttcag	73	AAAGAA	113	TTTTAG	185	gtattctattcaaatttattttactgtctt	1.30
6	tttctgaaattgcattttgttttcttgaag	186	ATTTTT	146	ACAGAA	331	gtaagtgatgttataaattataaataaatggc	6.35
7	gtttttctttatttgtttattttgaaatag	332	GAGCAC	165	GGTTAG	496	gtatgttttgaaggttgttgtttgtgaatttttt	8.10
8	catgactaataatttttttttttttttaag	497	AATTGT	166	TGCGAG	662	gtaatctaatctctttttctttttgttttgt	0.67
9	cccagttgagcttgtttgtttcttcacag	663	ACAAGA	239	AAAAAG	901	gtataaaggaaatgtttactgttttgaattt	1.94
10	aaaaattacattttaatttttggattacag	902	GTGCTT	164	CACCAG	1065	gtacagtaagtaggtcatgtcacatttaga	1.80
11	gaaaaaagtggatttattttttatttacag	1066	GTTTTT	170	GCCTTG	1235	gtaaagtgttaccattttctcattcagtgt	1.60
12	ttcaaattatccttttttttttttttttag	1236	GCTACA	372	TTCATG	1607	gtaagttcagcatgcattatgtctgactta	0.80
13	ttttcacaattgtcctttgttttgttatag	1608	TCCTGC	195	TCACAG	1802	gtaatttaagttcattagcatgctgctgtt	0.80
14	ctaagtgaagctttttgttttttctttgtag	1803	TAATTT	96	AGAATG	1899	gtatgttatctaataatgctctttatcatt	0.90
15	ttatatattaaagatcttactttcttgaag	1899	TGAACA	226	TCTGAG	2124	gtgagatttttttaaaaaaagaactaagctt	2.20
16	tatatatttttatttgtggtttactttaag	2125	ATTACA	126	GCCAAC	2250	gtaggagaatttatactaataaagtttcgg	1.15
17	aatttgcatttttccttctattcacaatag	2251	TCTCTA	126	ACCAAG	2376	gtaagattttcttcttcttgttttgttttt	1.40
18	ttgcttggttctttgtttgtcttaattgcag	2377	AAGAGT	90	AGTTTA	2466	gtaagtatgcttcctgttttgctatcatat	8.10
19	cttgaacatctttgtttctcttccttgaag	2467	GCATCC	172	CCATAG	2638	gtaaatacatatttactacttgggatttct	1.10
20	ttagtgttaatgagtgctttttatttttag	2639	GTGCCA	200	CATATG	2838	gtgagttacgttaaatgaagaagctcttgg	2.50
21	cctgatttttttccctcctaccatcttag	2839	TATCTA	83	ACTATC	2921	gtaagaaattaaaaccttatgttatgttca	0.10
22	aagttgaacttttttttttttttttaccacag	2922	CAATGT	156	ATTTTG	3077	gtaggtacagtctattttgtggtcctatt	1.20
23	tttaactttggaaaacttacttgatttcag	3078	GCATCT	76	CTTGAG	3153	gtgagtttttgcatttttttagtaagatct	0.10
24	tcatatttaaccacagttcttttcccgtag	3154	GCTGAT	131	CAATAG	3284	gtaatgggtcaaatattcatgaagtatttg	7.00
25	tttgtttgtttgtttgcttgcttgttttaag	3285	ATTGTT	118	GAAATG	3402	gtaattttaagtaacatgtatttgctgtta	1.30
26	ttacaatttttttttaaatttctttttaag	3403	TCCCAT	174	AAAAAG	3576	gtatatatggatgagtattttattagaagc	1.50
27	cttaacacattgactttttggttcgtgcag	3577	GTTTTA	170	CTATAG	3746	gtaagtttatacatgacatatgtgaaattt	1.35
28	aacctgtatatttttaaatttttctatttttag	3747	ATCTTG	247	AAACAG	3993	gtatggcttcaatttttatgtacttttcat	3.12
29	taaatatattttaattttgtgcccttgcag	3994	ATTGAT	116	TTCAGG	4109	gtatgtacattttaaacttagagaactagc	1.30
30	tgactgtatttttttcccttaactctgttag	4110	GGATTT	127	AGCCCT	4236	gtaagtatacatgatgagtttaataataga	0.50
31	aagtttttactaaatctgtttatttttctag	4237	GATTCC	200	CCAAAG	4436	gtaaataacatatttagaccaatatataag	2.80
32	ttgttgttgttttttttttctccctatattag	4437	GCCTTC	175	AAACAG	4611	gtaattttctgactcatcttcaaaatggta	0.52
33	tataatttttttcttttttaaattatatttag	4612	GTATTG	165	TTGGAG	4776	gtaataaaaatttcatcatctactatttttt	1.45
34	gttaaaagcaagttacattttctcttttag	4777	GAAATT	133	CTCAGG	4909	gtgctaattttaaatgacatgggctattt	2.25
35	ttaaactaatttttttaaaaaattatttctag	4910	ATAATC	96	TTCTAG	5005	gtaaactacagtcatgcgctgcgtgacattt	2.35
36	actctactgaaatagaatttctatatgtag	5006	AGGCTG	172	AGATTG	5177	gtgagtatttattgataccttatatgtaat	1.76
37	cttgataggcatttgaattgtttttttcag	5178	TGTCAA	142	AAAAAG	5319	gtctcttaagtaataaatgtttattgaata	1.05
38	atttacattttctaatccctttctttctag	5320	TTTTTA	177	TGTGAA	5496	gtaagaagattaattagtctgatataattc	1.65
39	tattgggtggatttgtttgtatattctag	5497	GTGAAA	178	ATTCAG	5674	gtattctattaaatttttaacattaatact	3.05
40	ggactgaggggagatatttttgtttgtcag	5675	AGTCAG	88	AAAGAG	5762	gtaatgtaatgagtgttgcttcttacgttt	2.15
41	tgaatgacattatatctcatttttctttag	5763	ACCTTC	156	GAAAAG	5918	gtaatggaatttagaatttttggtttttaa	2.10
42	cattaaaagaggtgttcttgtgacaaacag	5919	AAGTCT	88	TTACAG	6006	gtaaatattagaggctctattatttatgac	3.30
43	cttcaatttttgttgtttccatgttttcag	6007	GATCTT	89	TACTAG	6095	gtaaattgcatttttctaaacaacggtatag	0.10
44	cccaaagctattttcacaatcttttcttatag	6096	ACTACG	103	ATTCAG	6198	gtacatttttttcccagatttggtaaagcca	1.26
45	aacttaaaaacaaaataactcctgtttag	6199	GCCTTG	149	CGTCAG	6347	gtaagaaatttgacttgatttttttttttt	2.50
46	gtatatttttttctttgacttatctcacag	6348	CAAAGA	105	TGCCAG	6452	gtattatgaaaagacaaagttactgtatttt	1.25
47	tttcagagtgtctttttctttttgctactag	6453	AGTAAA	120	CTCAAG	6572	gtatgtaattcgtatgactttgttatccta	4.00
48	cttacatgaactctatgtcgtggcattcag	6573	ATCAGT	235	ACTCAG	6807	gtaaatacaatttaaaactatgtcatctta	0.51
49	atttattcccatatgtcattttcatttcag	6808	CTCCCT	168	GCAGCG	6975	gtttgttttttttattggctggattagtgt	1.40
50	tatattttaagattttgcctttcttatacag	6976	AACAAT	114	GAAAAG	7089	gtaagatttttggagcaacccttaagatag	1.30
51	tataatttaaattggttgtgttttcttgaag	7090	GCAGTA	218	AAACAG	7307	gtaactaggtttctacaagtgacaatttta	1.00
52	ttgtgttttaccttaattattctatgcaag	7308	ATACAC	208	ATGAAG	7515	gcaagtgttactcagcccaatattctaccc	1.00
53	cttattttgtgtctttttttaatggtag	7516	AGAGAC	114	AATAAT	7629	gtaagtaaacctgaaaatcaaaccacaata	0.32
54	tgcataaatctaatagttcttttcttacag	7630	CTAATC	159	GATGAG	7788	gtatttggattaaacatacgtacctttag	0.70
55	tatgtaatgtttttttgttttttattaatag	7789	GATCGA	139	AGAGAA	7927	gtatgtttttttttaaagaagaaacgttact	1.00
56	tcactaaaatctcttcatttttaaatacag	7928	AAGGCA	83	ATTAAG	8010	gtaatttgcaattaactcttgatttttttt	1.00
57	ctattatcaatcatgtttatacttttattag	8011	GTGGAC	141	GTTAAG	8151	gtgagccttcccttctctggcttagcccttt	0.80
58	acttgtttattcatgcttaattattctgaag	8152	GGCCGT	117	TATAAG	8268	gtaactatttgtacttctgttagttcacca	7.50
59	aattaaaaggtatttaatctgtaactccag	8269	GTGGTT	150	ATGATG	8418	gtgagtgacacccaaaattaaaggttattg	2.40
60	aaaataattatatatattctctatttaaag	8419	GAGGTG	166	CTATTG	8584	gtaatcttcttgtacatatagtagattgag	1.40
61	tttcagattgtttgtttctttttttctccag	8585	TTGGTT	87	ATCTAG	8671	gtaagtaataaaatctatgtatctattctt	6.00
62	cctcctaacttcactgtattctttacttttag	8672	GTGTTG	115	CAGAAG	8786	gtaagtgatatgaagtaaaggagggaaat	1.00
63	atccgtatttataatgtgtttgactctag	8787	ATGCTG	64	GTAGAG	8850	gtaaagtattttataaggaagactttatttt	11.00
64	cagatgttctctctgtttag	8851	GTCCTT	137	TCTCAG	8987	gtgagcagtattttaagaaggtcctgttgt	0.10
65	actgaaaccttgtgtttttgtccttag	8988	TGATAT	~3800				

[a] The first nucleotide of the open reading frame was designated +1.

[b] 1a and 1b are alternative leader exons.

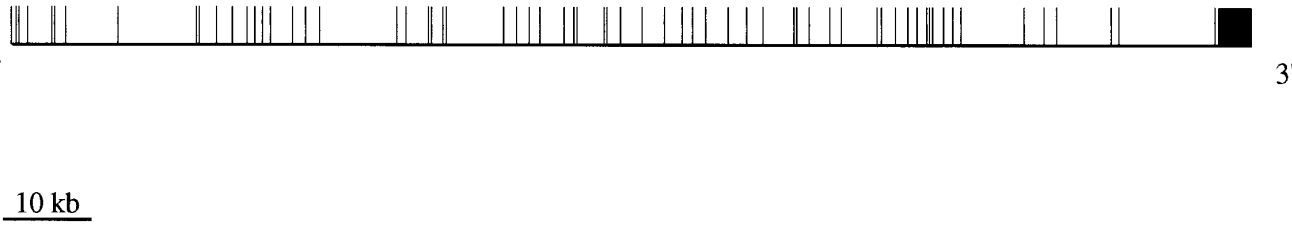

FIG. 2. A schematic representation of the exon–intron organization of the ATM gene. Vertical lines denote the positions of the ATM exons. The 3′ exon and all introns (except very small ones) are drawn to scale.

around these sites in all cases. During the search for the A-T gene, six exons were isolated by exon trapping (16) using the vectors pSPL3 (4) and λGET (10, 11). Their boundaries coincided with those obtained by long-distance PCR.

Parallel with these experiments, an effort was initiated to sequence the entire ATM gene. The sequence of several cosmids, spanning about 120 kb, has been completed (unpublished data). Comparison of this genomic sequence to the cDNA sequence yielded exon boundaries that coincided with those obtained by long-distance genomic PCR.

The ATM gene is composed of 66 exons (Fig. 2 and Table 1). The first two are alternative leader exons and were designated 1a and 1b (K. Savitsky *et al.*, in preparation). The first methionine of the open reading frame is located in exon 4, whereas the stop codon is located in the 3′ and largest exon of 3.8 kb. This exon includes a 3′ UTR of about 3600 nucleotides. With the exception of the 3′ exon, ATM exons range from 43 to 634 bp, with an average of 152 bp. The introns vary considerably in size, from 100 bp to about 11 kb, with the majority in the range 1–3 kb. The consensus dinucleotides GT and AG were found at the donor and acceptor splice sites of all introns, except for a variant donor site with a GC dinucleotide (reviewed in Ref. 9) present in the intron 3′ to exon 52.

The ATM gene contains one of the largest numbers of exons reported to date for a human gene. However, these exons are spread over a relatively compact genomic region of about 150 kb. The dystrophin gene, for example, consists of 79 exons spanning 2.4 Mb of genomic DNA (12), while the Huntington disease gene consists of 67 exons spread over 180 kb (1).

This report demonstrates the use of genomic long-distance PCR for rapid analysis of gene organization, even when relatively long introns are involved. The data presented here are indispensable for the amplification of ATM exons from genomic DNA and should be particularly useful for the identification of mutations in splice junctions and in regulatory elements.

ACKNOWLEDGMENTS

The authors are indebted to Drs. Francis Collins and Dan Tagle for valuable discussions. This study was supported by research grants from the A-T Children's Project, the A-T Medical Research Foundation, the Thomas Appeal (A-T Medical Research Trust), The United States–Israel Binational Science Foundation, and the National Institute of Neurological Disorders and Stroke (NS31763).

REFERENCES

1. Ambrose, C. M., Duyao, M. P., Barnes, G., Bates, G. P., Lin C. S., Srinidhi, J., Baxendale, S., Hummerich, H., Lehrach, H., Altherr, M., Wasmuth, J., Buckler, A., Church, D., Housman, D., Berks, M., Micklesm, G., Durbin, R., Dodge, A., Read, A., Gusella, J., and MacDonald, M. E. (1994). Structure and expression of the Huntington's disease gene: Evidence against simple inactivation due to an expanded CAG repeat. *Somat. Cell Mol. Genet.* 20: 27–38.
2. Barnes, W. M. (1994). PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proc. Natl. Acad. Sci. USA* 91: 2216–2220.
3. Cheng, S., Fockler, C., Barnes, W. M., and Higuchi, R. (1994). Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc. Natl. Acad. Sci. USA* 91: 5695–5699.
4. Church, D. M., Stotler, C. J., Rutter, J. L., Murrell, J. R., Trofatter, J. A., and Buckler, A. J. (1994). Isolation of genes from complex sources of mammalian genomic DNA using exon amplification. *Nature Genet.* 6: 98–104.
5. Easton, D. F. (1994). Cancer risks in A-T heterozygotes. *Int. J. Radiat. Biol.* 66: S177–S182.
6. Foord, O., and Rose, E. A. (1994). Long-distance PCR. *PCR Methods Appl.* 3: S149–S161.
7. Gatti, R. A. (1991). Ataxia-telangiectasia: An interdisciplinary approach to pathogenesis. *Medicine* 70: 99–117.
8. Harnden, D. G. (1994). The nature of ataxia-telangiectasia: Problems and perspectives. *Int. J. Radiat. Biol.* 66: S13–S19.
9. Jackson, I. J. (1991). A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res.* 19: 3795–3798.
10. Nehls, M., Pfeifer, D., and Boehm, T. (1994). Exon amplification from complete libraries of genomic DNA using a novel phage vector with automatic plasmid excision facility: Application to the mouse neurofibromatosis-1 locus. *Oncogene* 9: 2169–2175.
11. Nehls, M., Pfeifer, D., Micklem, G., Schmoor, C., and Boehm, T. (1994). The sequence complexity of exons trapped from the mouse genome. *Curr. Biol.* 4: 983–989.
12. Roberts, R. G., Coffey, A. J., Bobrow, M., and Bentley, D. R. (1993). Exon structure of the human dystrophin gene. *Genomics* 16: 536–538.
13. Savitsky, K., Bar-Shira, A., Gilad, S., Rotman, G., Ziv, Y., Vanagaite, L., Tagle, D. A., Smith, S., Uziel, T., Sfez, S., Ashkenazi, M., Pecker, I., Frydman, M., Harnik, R., Patanjali, S. R., Simmons, A., Clines, G. A., Sartiel, A., Gatti, R. A., Chessa, L., Sanal, O., Lavin, M. F., Jaspers, N. G. J., Taylor, A. M. R., Arlett, C. F., Miki, T., Weissman, S. M., Lovett, M., Collins, F. S., and Shiloh, Y. (1995). A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science* 268: 1749–1753.
14. Savitsky, K., Sfez, S., Tagle, D. A., Ziv, Y., Sartiel, A., Collins, F. S., Shiloh, Y., and Rotman, G. (1995). The complete sequence

of the coding region of the ATM gene reveals similarity to cell cycle regulators in different species. *Hum. Mol. Genet.* 4: 2025–2032.

15. Shiloh, Y. (1995). Ataxia-telangiectasia: Closer to unraveling the mystery. *Eur. J. Hum. Genet.* 3: 116–138.
16. Shiloh, Y., Ziv, Y., Savitsky, K., Bar-Shira, A., Gilad, S., Vanagaite, L., Uchenik, V., Smith, S., Patanjali, S. R., Tagle, D., Simmons, A., Clines, G., Collins, F. S., Weissman, S., Lovett, M., and Rotman, G. (1994). Genetic, physical and functional analysis of the ataxia-telangiectasia locus on chromosome 11q22–23. 44th Annual Meeting of the American Society of Human Genetics, Montreal. *Am. J. Hum. Genet.* 55: A49.
17. Swift, M., Morrell, D., Massey, R. B., and Chase, C. L. (1991). Incidence of cancer in 161 families affected by ataxia-telangiectasia. *N. Engl. J. Med.* 325: 1831–1836.
18. Weeks, D. E., Paterson, M. C., Lange, K., Andrais, B., Davis, R. C., Yoder, F., and Gatti, R. A. (1991). Assessment of chronic gamma radiosensitivity as an in vitro assay for heterozygote identification of ataxia-telangiectasia. *Radiat. Res.* 128: 90–99.
19. Zakian, V. A. (1995). ATM-related genes: What do they tell us about functions of the human gene. *Cell* 82: 1–20.