ELSEVIER

Biochimica et Biophysica Acta 1230 (1995) 81–85

BBA Biochimica et Biophysica Acta

Short Sequence-Paper

Isolation of the rat F_1-ATPase inhibitor gene and its pseudogenes ☆

Derina S. Samuel, John M. Belote, Samuel H.P. Chan *

Department of Biology, Syracuse University, Syracuse, NY 13244, USA

Received 2 January 1995; revised 13 March 1995; accepted 20 March 1995

Abstract

Multiple mitochondrial ATPase inhibitor genes have been identified in the rat genome. The sequences of two genomic clones indicate that one encodes the functional gene, and the other is a processed pseudogene. The ATPase inhibitor gene isolated is about 1.5 kb long and the coding region contains three exons and two introns. The presence of multiple pseudogenes in the rat is suggested by this study and this is unique since in the bovine genome only a single gene has been found, which is also confirmed here. The presence of multiple inhibitor transcripts in the rat suggests that the functional gene might have multiple transcriptional start sites.

Keywords: ATPase, F_1-, inhibitor; DNA sequence; Pseudogene; (Rat)

The mitochondrial ATP synthase (F_1F_0-ATPase) is a multi-functional enzyme complex essential for the final step of oxidative phosphorylation. The enzyme, found in the mitochondria of all known eukaryotic cells as well as in chloroplasts and bacteria, utilizes a transmembrane proton gradient in the synthesis of ATP (for a review, see ref. [1]). Under non-energizing conditions, i.e., in the absence of a membrane potential, resulting in the net hydrolysis of ATP, the reversible binding of the naturally occurring inhibitor protein to the β-subunits of F_1-ATPase is important, since it inhibits ATP hydrolysis. The natural ATPase inhibitor protein was initially isolated from bovine heart mitochondria [2], and has since been isolated from mitochondria of rat liver [3,4], yeast [5], and plants [6]. The amino acid sequence of the bovine inhibitor protein was determined by two groups [7,8], and its corresponding cDNA sequence subsequently was characterized by Walker et al. [9]. The amino acid sequence of the rat liver inhibitor (RLI) protein, however, has not been directly determined since the N-terminal amino acid is blocked (D.S. Samuel and S.H.P. Chan, unpublished observation).

Recently two rat liver inhibitor cDNAs have been cloned, and the corresponding amino acid sequences deduced, by two groups [10,11]. Curiously, the cDNAs differ at their 5′ end (Fig. 1). Both sequences are identical downstream of base +33 but are entirely different in the 5′ non-coding region and 30 bases after the initiator ATG. This discrepancy in the sequences of the two RLI cDNAs does not appear to be due to the presence of two transcribed rat inhibitor genes, since the absolute identity of nucleotides 34 through 439 would not be expected unless the gene duplication was an extremely recent event. A second possible explanation for this difference is that the two cDNAs were formed as a result of alternative splicing of the RLI pre-mRNA. In this case, alternative 5′ exons would be present, yielding RLI polypeptides that differed at their amino terminal ends. A third possible basis for the difference between the two reported cDNAs is that one, or both, of these cDNAs might contain unrelated sequences as an artifact of their cloning. In order to distinguish these possibilities and to further investigate the molecular characterization of the rat ATPase inhibitor gene, we have analyzed the structure and complexity of the RLI gene by genomic Southern blot hybridization and by the isolation and sequencing of cDNA and genomic clones.

The reverse-transcriptase polymerase chain reaction (RT-PCR) technique [12] was used to amplify cDNA corresponding to the region between bases 115 and 363 of the RLI cDNA, based on the published sequence [10]. The reaction was done with 5 ng rat liver total RNA in a final volume of 25 μl and contained 50 μM of each dNTP, 0.5 μM of the 5′ and 3′ primers (primers synthesized at Genosys Biotechnologies (The Woodlands, TX)), 100 U of MMLV-RT, 1.25 U of *Taq* DNA polymerase and 1 × PCR buffer containing 50 mM KCl, 10 mM Tris-HCl (pH

☆ The nucleotide sequence data reported in this paper have been submitted to the GeneBank database under the accession numbers U12250 (for RI-1) and U12251 (for RI-2).

* Corresponding author. Fax: +1 (315) 4432012.

SSDI 0005-2728(95)00049-6

8.3), 1.5 mM $MgCl_2$ and 0.01% gelatin. The reaction was carried out for a total of 35 cycles after preincubation at 42° C for 10 min. The subsequent cycle conditions were as follows: denaturation, 40 s at 95° C; primer annealing, 1 min at 55° C; primer extension, 3 min at 72° C.

The RLI cDNA, obtained through RT-PCR, was used as a probe to screen a λgt11 rat liver cDNA library (Clontech Labs.) and a λEMBL-3 rat genomic library (a kind gift from Dr. Donald Back, Queen's University, Kingston, Ontario) using LE 392 as the host *E. coli* strain [13]. Sequencing of the plasmid clones was done according to the method of Sanger et al. [14], using [^{35}S]dATP and the Sequenase sequencing kit (United States Biochemical). High molecular mass rat genomic DNA was prepared by a modification of Blin and Stafford [15]. Bovine and human genomic DNA were obtained as a gift from Dr. S. Yokoyama (Syracuse University). Genomic DNA (15 μg/gel slot) was digested at 37° C for 16 h with various restriction endonucleases (Promega). Electrophoresis and overnight transfer of DNA fragments to Genescreen Plus membrane (NEN) was essentially as described by Sambrook et al. [13]. RNA was isolated from rat tissue using the basic protocol of Chomczynski and Sacchi [16] with some modifications of Puissant and Houdebine [17]). Polyadenylated RNA was isolated as described by Celano et al. [18]. Both total and polyadenylated RNA samples were analyzed by formaldehyde-agarose gel electrophoresis essentially as described by Sambrook et al. [13].

Since the two published descriptions of RLI cDNAs yielded sequences that differed from each other, we sought to resolve the discrepancy by isolating and characterizing additional RLI cDNA clones. In screening over 10^6 recombinant clones, only one RLI cDNA clone was obtained. It is not known why the RLI cDNAs are so rare in the library, since Northern blot analysis suggests that RLI mRNA is present at reasonable levels. The 0.8 kb *Eco*RI insert of this clone was subcloned into the pGEM-3 plasmid vector and sequenced. Upon analysis, the clone was found to contain the same coding region as the sequence reported by Higuti et al. [11] (Fig. 1).

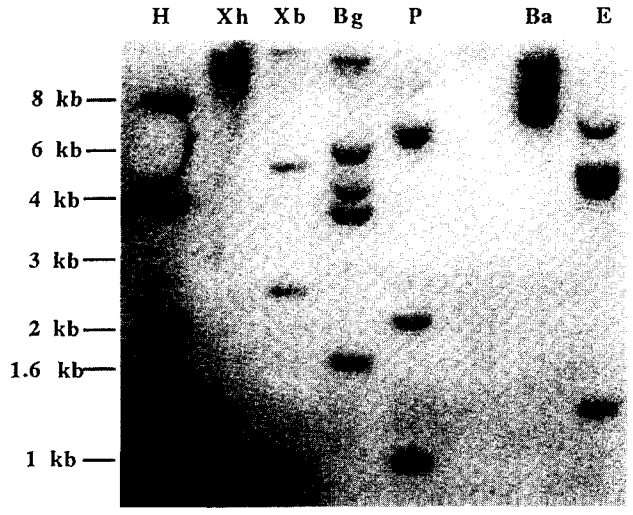

Fig. 2. Genomic Southern. Rat genomic DNA was digested with different enzymes and probed with the RLI cDNA, under high stringency conditions. The enzymes used were: H = *Hin*dIII, Xh = *Xho*I, Xb = *Xba*I, Bg = *Bgl*II, P = *Pst*I, Ba = *Bam*HI, and E = *Eco*RI. The molecular sizes are also indicated.

A Southern blot of rat spleen genomic DNA digested with seven different restriction enzymes was also probed and at least four hybridizing fragments are seen in every lane, under high stringency conditions (Fig. 2). Since this pattern is seen with several different restriction enzymes, it suggests the presence of multiple ATPase inhibitor genes (or pseudogenes) in the rat. This was surprising, since Walker et al. [9] have reported the presence of only one gene in both bovine and human genomic DNA, using the bovine ATPase inhibitor cDNA as a probe. To confirm that there is indeed a difference in the number of RLI sequences in rat as compared to the bovine genomes, *Hin*dIII and *Eco*RI digests of rat and bovine genomic DNA were probed with the PCR-amplified RLI cDNA

```
(i)     ctaca accca ggcac aaatc tcatc aggat aagtc tgaca tttct
(ii)                      caac cacgc ccgca acgga gctga gcaac
(iii)             gagaa g---- ----- ----- ----- ----- -----

                   M   T   K   S   C   R   I   E   A   S   T
(i)     gtctt ccaac ATG ACA AAA TCC TGT AGA ATC GAA GCT AGT ACG
(ii)    gccga agata ATG GCA GGC TCG GCG TTG GCG GTT CGG GCT CGG
(iii)   ----- ----- --- --- --- --- --- --- --- --- --- --- ---
                   M   A   G   S   A   L   A   V   R   A   R

         L   G   V   W   G   M   R   V   L   Q   T   R   G   F
(i)     CTC GGT GTC TGG GGT ATG AGG GTC CTG CAA ACC CGA GGC TTC
(ii)    --- --- --- --- --- --- --- --- --- --- --- --- --- ---
(iii)   --- --- --- --- --- --- --- --- --- --- --- --- --- ---
```

Fig. 1. Comparison of the cDNA sequences isolated. The 5′ end of the two published sequences are shown: (i) Lebowitz and Pedersen [10] and (ii) Higuti et al. [11]. The 5′ end of the cDNA clone isolated in the present study (iii) is aligned with that of Higuti et al. [11]. The dashes shown indicate regions of identity between the sequences.

under reduced stringency conditions (Fig. 3). For each enzyme (data not shown for *Eco*RI), only one hybridizing fragment is seen in the bovine DNA, which confirms the earlier finding. However, in rat DNA there are at least 8 fragments containing cross-hybridizing sequences. Since the intensity of the extra bands is lower, it is likely that these sequences are less homologous to the functional gene than are the four identified under high stringency conditions.

Since the presence of multiple inhibitor-like sequences was suggested by our Southern blot analysis, and since two different RLI cDNA sequences had been reported [10,11], we sought to isolate genomic clones containing the RLI and possible RLI-like genes by screening a λEMBL-3 rat genomic library. Approximately 10^6 plaques were screened and 20 positive recombinant clones were obtained. Restriction enzyme and hybridization analyses revealed that they fell into three groups. Detailed sequence analysis was performed on each of the three classes of clones and this identified two different classes, RI-1 and RI-2 (Fig. 4). The third clone was found to be identical to RI-2 with a deletion in the 3′ end of the third exon and differences in the 5′ end (data not shown). RI-1 contains the rat ATPase inhibitor gene and encodes the cDNA described above. The gene contains two introns: intron I is 87 bp long and interrupts the coding region after codon 29; intron II is 905 bp long and is positioned after the second nucleotide of codon 60. Both introns follow the obligatory GT/AG rule at the 5′ and 3′ ends, respectively. As mentioned above, one possible explanation for the two different reported cDNA sequences is that there might be alternatively spliced forms of RLI transcripts. If so, one would expect the presence of intron sequences immediately preceding codon 11, when the cDNA and genomic sequences are compared. The absence of intron sequence at this position indicates that the different published cDNA sequences are not explained by there being alternatively spliced transcripts.

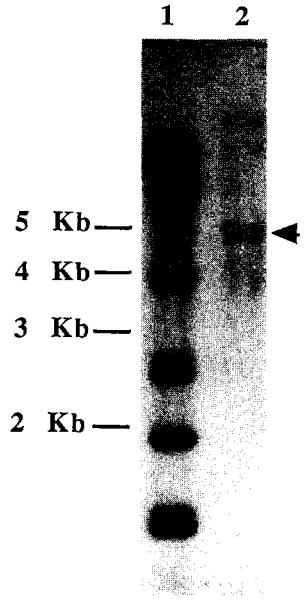

Fig. 3. Zooblot. Rat and bovine genomic DNA were digested with *Hin*dIII and probed with the RLI cDNA, under low stringency conditions. Lane 1 contains rat DNA and Lane 2 contains bovine DNA. The arrow indicates the single hybridizing band in the bovine DNA. The molecular markers are indicated.

The nucleotide sequence of the other clone, RI-2, is very similar to RI-1 (Fig. 4). However, it exhibits characteristics of a processed pseudogene: the intron sequences have been precisely removed and there is a long tract of A-residues, starting 13 bp after the consensus polyadenylation signal. Comparison of RI-2 and RI-1 indicates that RI-2 was probably derived from RI-1, since the sequence homology is quite high. However, it no longer has the characteristics of a functional gene. This is clearly seen when the rates of synonymous (K_s) and non-synonymous substitutions (K_a) are calculated per site. Between RI-1 and RI-2 the K_s value is 0.0593 and that of K_a is 0.0666 (K_s/K_a ratio is approximately 1). This indicates that there has been no evolutionary pressure on the rates of mutation in the RI-2 gene. In addition, the insertion of two extra nucleotides in RI-2 has interrupted the RLI reading frame, leading to a premature termination signal. Taken together, these observations strongly suggest that RI-2 is a nonfunctional pseudogene. This situation is not unusual, as processed pseudogenes are usually found to be transcriptionally inactive [19]. Considering that these pseudogenes appear to be inserted randomly into the genome, the chances that the insertion occurs correctly, 3′ to RNA polymerase II promoters to permit normal transcription, is improbable. It is likely that the other hybridizing fragments seen in our genomic Southern analysis represent additional pseudogenes of the RI. Processed pseudogenes have been shown to be fairly common in vertebrates and found to be part of many different gene families [19,20]. In rat, the best characterized of such a gene family is that of cytochrome *c* [21] which has more than 25 cytochrome *c*-like genes. The present study is the first report of multiple pseudogenes of the ATPase inhibitor protein.

Assuming that RI-2 is the result of an RI-1 encoded mRNA that was reverse-transcribed and then inserted into the genome, comparing the two sequences might reveal certain structural features such as the position of putative introns, polyadenylation and transcription initiation sites. At the 5′ end, sequence homology with RI-1 ends 6 bp from the initiator ATG, and this could indicate the presence of an intron in the 5′ noncoding region. However, at the point where the homology ends in RI-1 there is no 3′ consensus splice sequence that would indicate the end of an intron. In addition, the 5′ sequence of the cDNA continues in its homology with RI-1 for another 33 bp. It therefore seems more probable that RI-2 is an incompletely reverse-transcribed sequence. At the 3′ end, RI-2 has remnants of a poly-A tail 12 bp from the putative polyadenylation signal reported in the cDNA sequences

[10,11], confirming that this is indeed an authentic polyadenylation signal.

In order to further characterize the rat ATPase inhibitor gene, an analysis of its transcripts was undertaken. The liver RNA was enriched for poly(A) RNA and this was probed with the RLI cDNA. As is seen in Fig. 5, two transcripts are seen, at approximately 0.8 and 0.4 kb. One possible basis for this is that there may be two different transcriptional start sites in the functional gene, RI-1. Analysis of bovine liver and heart RNA also shows the presence of two transcripts (data not shown). As there is only one bovine inhibitor gene [9], this further suggests that the two rat transcripts are products of the same gene.

Since the 3′ sequences, of the two cDNAs, are 100% homologous, it seems unlikely that there are two different genes, since one would expect a certain number of synonymous nucleotide substitutions due to evolutionary divergence of the genes. Our results also rule out the possibility that the two published RLI cDNA sequences represent alternatively spliced forms of the RLI gene transcript, since examination of the genomic DNA sequence shows that there is no intron junction at the position where the two cDNAs differ. The most reasonable explanation for the difference in the 5′ end of Lebowitz and Pedersen's sequence [10] is that it is due to a cloning artifact.

Comparing the sequence of the processed pseudogene, RI-2, with RI-1, it is possible to deduce its origin, since RI-2 has a sequence highly homologous to RI-1, with the introns spliced out correctly and the remnants of a poly(A) tail.

In conclusion, this study represents the presence of multiple inhibitor genes in the rat genome and multiple

```
RI-1  cgctg tttcc ctgct cgtcc cgcct cccgc cgtct agatc tggtt ccagc cgccc ctgca
RI-2  tccag ctcaa gactt tgctt gtagt atacc agcca ccagc cgtag aagct tccag ttgtg

RI-1  gttag cgcgc tagga gagaa gcaac cacgc ccgca acgga gctga gcaac gccga agata
RI-2  caaaa gctaa cgtaa gaaag gagac aactt acctt catta gaaat tccaa gcta- -----

       M   A   G   S   A   L   A   V   R   A   R   L   G   V   W   G   M   R
RI-1  ATG GCA GGC TCG GCG TTG GCG GTT CGG GCT CGG CTC GGT GTC TGG GGT ATG AGG
RI-2  --- --- --- -T- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

       V   L   Q   T   R   G   F   G   S   E   S         (87 bp)       S   E
RI-1  GTC CTG CAA ACC CGA GGC TTC GGC TCG GAC TCG gt...Intron I...ag TCG GAG
RI-2  --- --- --- --- --- --- --- -T- --A --- -T- [_________________] --- ---

       S   M   D   S       G   A   G   S   I   R   E   A   G   G   A   F   G
RI-1  AGC ATG GAT TCG ** GGC GCT GGC TCC ATC CGA GAA GCT GGT GGG GCC TTC GGG
RI-2  --- --- --- A-A GG --- --- --- --- --T --- --- --- --- --- --- --- A--

       K   R   E   K   A   E   E   D   R   Y   F   R        (905 bp)        E
RI-1  AAA CGA GAG AAG GCT GAA GAG GAT CGG TAC TTC CGgt...Intron II...agA GAG
RI-2  --- --- --- --- --- --- --- --- --- --- --- --[__________________]- -G-

       K   T   R   E   Q   L   A   A   L   K   K   H   H   E   D   E   I   D
RI-1  AAG ACT AGA GAG CAG CTG GCT GCC TTG AAG AAG CAC CAT GAA GAT GAG ATT GAC
RI-2  --- --- --- --- --- --- --- CG- --- --- --- --- --- --- --- --- --- ---

       H   H   S   K   E   I   E   R   L   Q   K   Q   I   E   R   H   K   K
RI-1  CAC CAT TCG AAG GAG ATA GAG CGT CTG CAA AAA CAG ATC GAA CGG CAT AAG AAG
RI-2  -G- --- --- --- --- --- --- -A- --- --C --- --A --- --- --- --- --- ---

       K   I   K   Y   L   K   N   S   E   H
RI-1  AAG ATT AAA TAC CTA AAG AAT AGT GAG CAT TGA gtcac acagt cagtc tctca cagag
RI-2  --- --- --- --- --- --- -G- -A- --C --- -A- ----- ----- ----- ----- -----


RI-1  tggcc agtat cattc cccac ttcta tagag atgtt gatga ttgat gttga ctgtg tgcta
RI-2  ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----


RI-1  ctaac agata ataaa attat cacca ggata ctttc ttgat gcctt ctact gtctag
RI-2  ----- ----- ----- ----- ----- ---ga aaaaa aaaaa gaaat tccaa gctaaa
```

Fig. 4. Sequence comparison of the two rat genomic clones, RI-1 and RI-2. RI-1 contains two introns: Intron I is 87 bp long and Intron II is 905 bp long. The introns have been excised in RI-2 and this is denoted by [__]. Only the nucleotides that are different in RI-2 are indicated when compared to RI-1. The stars (*) shown in RI-1 are to align the sequences and indicate an insertion of two nucleotides in RI-2. The two stop codons and the polyadenylation signal are highlighted in bold letters.

Fig. 5. Northern. Rat liver poly(A+) RNA was probed with the RLI cDNA under high stringency conditions. The arrows indicate the two transcripts that are recognized by the probe. The two ribosomal RNAs (28S and 18S) are faintly seen as a result of non-specific hybridization. The blot was developed after an overnight exposure.

transcripts, possibly from a single functional gene. It should be further studied in order to understand the complex regulation at the transcriptional level.

This study was supported in part by a Grant-in-aid from the American Heart Association.

References

[1] Penefsky, H.S. and Cross, R.L. (1991) Adv. Enzymol. 64, 173–214.
[2] Pullman, M.E. and Monroy, G.C. (1963) J. Biol. Chem. 238, 3762–3769.
[3] Chan, S.H.P. and Barbour, R.L. (1976) Biochim. Biophys. Acta 430, 426–433.
[4] Cintron, N.M. and Pedersen, P.L. (1979) J. Biol. Chem. 254, 3439–3443.
[5] Satre, M., deJerphanion, M.B., Huet, J. and Vignais, P.V. (1975) Biochim. Biophys. Acta 387, 241–255.
[6] Norling, B., Tourikas, C., Hamsur, B. and Glaser, E. (1990) Eur. J. Biochem. 188, 247–252.
[7] Frangione, B., Rosenwasser, E., Penefsky, H.S. and Pullman, M.E. (1981) Proc. Natl. Acad. Sci. USA 78, 7403–7407.
[8] Dianoux, A.C., Tsujita, A. and Przybylski, M. (1984) 174, 151–156.
[9] Walker, J.E., Gay, N.J., Powell, S.J., Kostina, M. and Dyer, M.R. (1987) Biochemistry 26, 8613–8619.
[10] Lebowitz, M.S. and Pedersen, P.L. (1993) Arch. Biochem. Biophys. 301, 64–70.
[11] Higuti, T., Kuroiwa, K., Kawamura, Y., Morimoto, K. and Tsujita, H. (1993) Biochim. Biophys. Acta 1172, 311–314.
[12] Kawasaki, E.S. (1990) in PCR Protocols: A Guide to Methods and Applications, Academic Press, pp. 21–27.
[13] Sambrook, J., Fritsch, E. and Maniatis, T. (1989) Molecular Cloning, 2nd Edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[14] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[15] Blin, N. and Stafford, D.W. (1976) Nucleic Acids Res. 3, 2303–2308.
[16] Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. 162, 156–159.
[17] Puissant, C. and Houdebine, L.M. (1990) Biotechniques 8, 148–149.
[18] Celano, P., Vertino, P.M. and Casero Jr., R.A. (1993) Biotechniques 15, 27–28.
[19] Vanin, E.F. (1985) Annu. Rev. Genetics 19, 253–272.
[20] Wilde, C.D. (1986) Crit. Rev. Biochem. 19, 323–352.
[21] Scarpulla, R.C. (1984) Mol. Cell Biol. 4, 2279–2288.