4344–4347 Nucleic Acids Research, 1993, Vol. 21, No. 18

Cloning and characterization of the *C.elegans* histidyl-tRNA synthetase gene

Yousef G.Amaar and David L.Baillie
Institute of Molecular Biology and Biochemistry, Department of Biological Sciences, Simon Fraser University, Burnaby, BC V5A 1S6, Canada

Received April 19, 1993; Revised and Accepted August 9, 1993 GenBank accession no. L13152

ABSTRACT

In this paper, we report the cloning and sequencing of the *C.elegans* histidyl-tRNA synthetase gene. The complete genomic sequence, and most of the cDNA sequence, of this gene is now determined. The gene size including flanking and coding regions is 2230 nucleotides long. Three small introns (45 – 50 bp long) are found to interrupt the open reading frame. The open reading frame translates to 523 amino acids. This putative protein sequence shows extensive homology with the human and yeast histidyl-tRNA the histidyl-tRNA synthetase gene is a single copy gene. Hence, it is very likely that it encodes both the cytoplasmic and the mitochondrial histidyl-tRNA synthetases. It is likely to be trans-spliced since it contains a trans-splice site in its 5′ untranslated region.

INTRODUCTION

Aminoacyl-tRNA synthetases form a family of enzymes which are essential for protein synthesis. These enzymes transfer amino acids to their cognate tRNA through a two step reaction. First, the enzyme catalyzes the binding of the appropriate amino acid to adenosine triphosphate (ATP) to produce an amino acid adenylate complex. Second, the enzyme catalyzes the transfer of the amino acid to its cognate tRNA (1). There are 20 distinct aminoacyl-tRNA synthetases in every living cell, one for each naturally occurring amino acid (1).

Although aminoacyl- tRNA synthetases have been investigated in a number of prokaryotes and eukaryotes, little is known about them in the nematode *Caenorhabditis elegans*. In this paper we report the cloning and characterization of the *C.elegans* histidyl-tRNA synthetase gene.

MATERIALS AND METHODS

Caenorhabditis elegans var Bristol, strain N2 was the organism used in this study. The nematodes were maintained on NGM plates supplemented with *E.coli* strain OP50 as a food source (2).

Picking random clones from a *C.elegans* lambda Zap cDNA library

A *C.elegans* lambda Zap cDNA library (Stratagene) was provided by R.Barstead and R.Waterston, Washington University, Missouri. Approximately ten thousand phage were plated according to the Stratagene protocol. Fifty phage were randomly picked and plasmids (pBluescript) for twenty-nine of them were excised, using the Stratagene protocol. Twenty-three of these plasmids contained cDNA inserts ranging from 400 to 1600 bp. Plasmid #29 which contained a 420 bp cDNA insert was chosen for sequencing.

Screening the lambda Zap cDNA library

A *C.elegans* Lambda Zap cDNA library was screened with the 420 bp cDNA fragment contained in plasmid #29. Using the method of Benton and Davis (3), three rounds of screening were carried out and the plasmids from two phage isolated were excised according to the Stratagene protocol. The two plasmids obtained were called #3-1 and #6-1 and contained 1.5 and 1.4 kb cDNA inserts, respectively.

Preparation of plasmid DNA

Plasmid DNA was prepared using a modified alkaline lysis protocol (Pharmacia Miniprep Kit plus). The DNA extracted with this method was used for both restriction digests and sequencing reactions.

Preparation of cosmid DNA

Cosmids were provided by A.Coulson and J.Sulston of The Medical Reasearch Council (4), Cambridge, England. Cosmid DNA was prepared using a standard protocol as out lined by Maniatis *et al* (5).

Restriction enzyme digestion

All restriction enzymes were purchased from either Bethesda Research Laboratories (BRL) or from Pharmacia. The amount of restriction enzyme and the reaction buffers used were those recommended by the manufacturer.

* To whom correspondence should be addressed

Southern blots

Southern blots were prepared using a standard protocol (5).

Preparation of DNA fragments for oligolabelling

Cloned DNA for oligolabelling was prepared by applying the low melting Seaplaque GTG agarose gel protocol as described by FMC Bioproducts. All of the probes used in this study were labeled with ^{32}P-dATP using the oligolabelling technique of Feinberg and Vogelstein (6). Random hexamer primers were obtained from Pharmacia. The specific activity of probes varied from 1×108 to 2.0×109 cpm/μg. Hybridization of probes to DNA filters was carried out as described in (7).

Subcloning

Subcloning was conducted by utilizing the shotgun method (8). The ligation reactions of all *Eco*R1 fragments to *Eco*R1 digested plasmid (pBluescript SK$^+$) were carried out as previously described by Snutch (8). The transformation procedure was carried out as described in the BRL transformation protocol.

Oligonucleotides

Oligonucleotide primers were designed from the cDNA sequence as it became available. The oligonucleotides were designed using the program OLIGO (9). The oligonucleotides were on average 18 nucleotides in length and were synthesized on an ABI synthesizer. Commercially available T3 and T7 primers were used to sequence the 420 bp *Eco*RI cDNA fragment (plasmid #29) and the ends of 1.5 cDNA fragment (plasmid #3-1).

Sequencing methodology

DNA sequencing was performed using the dideoxy termination method of Sanger *et al.* (1975) (10). Templates were sequenced as described by Hattori and Sakaki (11) using Sequenase version 2.0 kit reagents(USB), following the recommendations of T.Snutch (personal communication).

All DNA and protein sequence analyses were conducted using the following computer programs: (I) FASTA (12), (II) PC/Gene (the nucleic acid and protein sequence analysis software system: A.Bairoch/University of Geneva; (TM) IntelliGenetics Inc. and Genofit SA., (III) ESEE (The Eyeball Sequence Editor) (13).

RESULTS

Recovery of a cDNA fragment that encodes a portion of the *C.elegans* his-tRNA synthetase gene

A 420 bp cDNA was serendipitously recovered (see materials and methods) from a *C.elegans* lambda Zap cDNA library B.Barstead and R.Waterston, Washington University, Missouri). This 420 bp fragment was found to encode a portion of the *C.elegans* histidyl-tRNA synthetase. This fragment was sequenced; the 420 bp coding nucleotide sequence was translated and the predicted peptide sequence was used to search the translated EMBL and SWISS PROT 13 data banks. The predicted peptide showed extensive homology with the human histidyl-tRNA synthetase amino acid sequence. The fragment showed 70.9% amino acid identity in a 177 amino acid overlap and 88% similarity when conservative changes were taken into

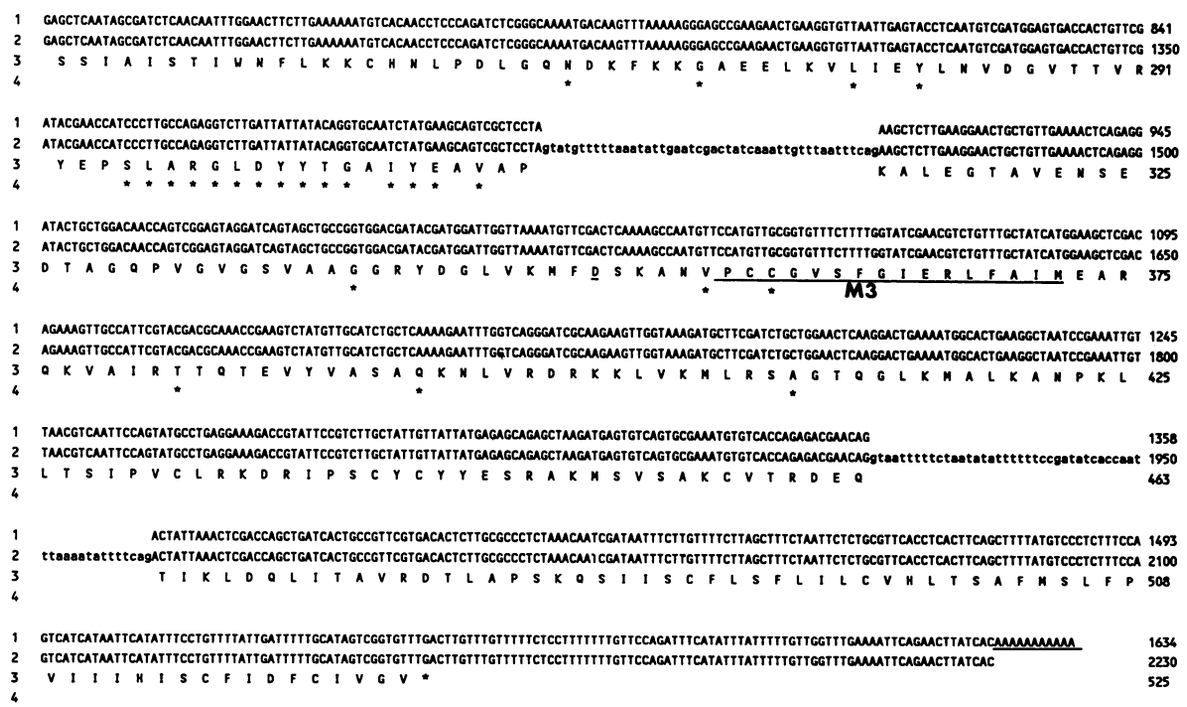

Figure 1. Most of the cDNA sequence (1), the genomic sequence (2), the deduced amino acid sequence (3) of the *C.elegans* histidyl-tRNA synthetase gene are shown. Lower case letters indicate the 5′ untranslated region and the location of the three introns identified. The putative TATA box, (tatta), the CCAAT box, (ccaaat), the trans-splice site (tttttcag), the three motifs (M1, M2, M3) are underlined, as is the poly (A) tail. The (*) (4) shows identical amino acids among *C.elegans*, human, and *S.cerevisiae* histidyl-tRNA synthetases.

consideration. The high degree of similarity suggested that this 420 bp cDNA represents a portion of the coding sequence for the *C.elegans* histidyl-tRNA synthetase gene.

Screening a *C.elegans* cDNA library for larger cDNA fragment(s) using the 420 bp cDNA as a probe

The recovered 420 bp ^{32}P labeled cDNA was used as a probe to screen the *C.elegans* Lambda Zap cDNA library in order to recover additional cDNAs. Approximately twenty thousand phage were screened resulting in the recovery of two cDNAs of 1.4 kb and 1.5 kb (Data not shown). Both cDNAs were found to contain sequences corresponding to the *C.elegans* histidyl-tRNA synthetase gene, however neither cDNA was full length. The 1.5 kb cDNA fragment was chosen and used to continue the sequencing process.

The nucleotide sequence from the 1.5 kb fragment was subsequently found to overlap with the sequence of the 420 bp cDNA fragment.

Probing *C.elegans* genomic DNA digested with *Eco*R1 with the 420 bp cDNA

The search for the genomic DNA fragment containing the *C.elegans* histidyl-tRNA synthetase gene necessitated determination of the size of the genomic fragment(s) containing all or part of the gene. A Southern blot containing *C.elegans Eco*R1 digested genomic DNA, probed with the 420 bp ^{32}P labeled cDNA resulted in probe hybridization to a 5.5 kb *Eco*R1 genomic fragment (figure 1), and suggests that the *C.elegans* histidyl-tRNA synthetase gene is probably present in a single copy.

Cosmid CO4G1 contains genomic DNA corresponding to the *C.elegans* histidyl- tRNA synthetase gene

The overlapping region identified by the four YACs on the physical map was best covered by the cosmid CO4G1 (results not shown). A Southern blot of *Eco*R1 digested cosmid CO4G1 DNA probed with the 420 bp ^{32}P labeled cDNA resulted in hybridization to a 5.5 kb *Eco*R1 genomic fragment. This result was consistent with that obtained from probing *C.elegans Eco*R1 digested genomic DNA with the same probe. Thus, the cosmid CO4G1 contains the genomic DNA corresponding to the *C.elegans* histidyl-tRNA synthetase gene. The 5.5 kb *Eco*R1 genomic fragment containing the gene was subsequently subcloned into pBluescript SK^+ for further analysis.

cDNA sequence

The sequences from both the 420 bp and the 1.5 kb cDNA fragments were found to overlap with each other. The 1.5 kb cDNA fragment was found to contain a poly A tail marking the 3′ end of the gene. Combining the nucleotide sequence from the two cDNA fragments, a sequence of 1770 nucleotides corresponding to the *C.elegans* histidyl-tRNA synthetase gene was obtained. This sequence, however, is missing the 5′ initiating methionine.

Obtaining the genomic DNA sequence corresponding to the *C.elegans* histidyl-tRNA synthetase gene

Synthetic oligonucleotides designed from the cDNA sequence were used to obtain the sequence of interest from the 5.5 kb *Eco*R1 genomic fragment subcloned from cosmid CO4G1. Genomic sequence corresponding to the 5′ untranslated region, coding region, and 3′ untranslated region of the *C.elegans* histidyl-tRNA synthetase gene was obtained (figure 1). The genomic sequence corresponding to the coding region of the gene exactly matches the cDNA sequence except where interrupted by introns. The genomic sequence contains an open reading frame interrupted by three small introns (47−52 bp long, figure 1). The coding sequence of the *C.elegans* histidyl-tRNA synthetase translates to a putative protein sequence of 525 amino acids. The 3′ untranslated region of the cDNA sequence contains a poly(A) tail. However, a consensus poly(A) addition site was not found in the cDNA and genomic sequences corresponding to the 3′ untranslated region.

DISCUSSION

The *C.elegans* histidyl-tRNA synthetase gene is a single copy gene

Southern experiments have shown that the *C.elegans* histidyl-tRNA synthetase gene is present in a single copy in the genome. In eukaryotic cells studied previously two forms of synthetases have been found. Distinct cytoplasmic and mitochondrial synthetases are required for normal development and viability of the organism (14). In yeast, it was shown that a single structural nuclear gene can encode both the cytoplasmic and the mitochondrial histidyl- and valyl-tRNA synthetases (15,16). In the case of *C.elegans* it appears that the histidyl-tRNA synthetase gene is a single copy gene, and probably encodes both the cytoplasmic and the mitochondrial synthetase functions.

Analysis of the genomic DNA sequence of the *C.elegans* histidyl-tRNA synthetase gene

The complete DNA sequence of the histidyl-tRNA synthetase gene from the nematode *C.elegans* has been determined (figure 1). The size of this gene including the flanking and coding sequences is 2230 nucleotides. This gene contains an open reading frame interrupted by three small introns (47−53 bp). The position of each of the three introns has been identified by comparing the genomic sequence to the partial cDNA sequence (figure 1).

Examination of the 5′ untranslated region of the gene shows a putative TATA box, (TATTA), and a putative CCAAT box, (CCAAAT). In most eukaryotic genes, these two consensus sequences indicate the promoter region of the gene (17). Thus, the presence of the TATTA and CCAAAT boxes provides evidence that the 5′ untranslated region contains the promoter region of the *C. elegans* histidyl-tRNA synthetase gene. The 5′ untranslated region has also been found to contain a splice site (TTTTTCAG). This suggests that this gene may be trans-spliced. The trans-splicing phenomena has been well characterized in the protozoan *Trypanosome bruci* where all mature mRNAs contain a 35 nucleotide leader sequence at their 5′ end (18). In *C.elegans* about 10 to 15% of the genes have been estimated to be trans-spliced (19). The *C.elegans* histidyl-tRNA syntheses gene predicts a protein of 525 amino acids which shares extensive homology with both the human and *S.cerevisiae* histidyl-tRNA synthetase protein sequences (figure 1).

The *C.elegans* putative protein sequence exhibits the three signature motifs characteristic of class II synthetases (20) (figure, 1). The presence of these motifs in the *C.elegans* protein sequence further confirms the notion that class II synthetases have arisen from a common ancestral gene (20).

ACKNOWLEDGEMENTS

We would like to thank Dr R.Waterston for kindly providing the genomic mapping filter. We would like also to thank Dinar Suleman, Marcia Wakarchuk, and Steve Jones for their technical assistance, and Marco Mara for his critical review of the manuscript. This work has been supported by a grant to D.L.B. from the NSERC and a scholarship to Y.A. from the Libyan Secretariat of Scientific Research.

REFERENCES

1. Neidhardt, F.C., Parker, J. and Mckeever, W.G. (1975) Ann Rev. Bacteriol. 216–234.
2. Brenner, S. (1974) Genetics 77, 71–94.
3. Benton, W.D and Davis, W. (1977) Science 196, 180–182.
4. Coulson, A., Sulston, J., Brenner, S. and Karn, J. (1986) Proc. Natl. Acad. Sci. USA. 83, 7821–7825.
5. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular cloning: A Laboratory Manual. Cold Spring Harbor University Press, Clod Spring Harbor, New York.
6. Feinberg, A. and Volgelstein, B. (1983) Anal. Biochem. 132, 6–13.
7. Prasad, S.S. and Baillie, D.L. (1989) Genomics 5, 185–198.
8. Snutch, T.P. (1984) Ph.D. Thesis, Simon Fraser University, Burnaby, B.C., Canada
9. Rychlik, W. and Rhoads, R.E. (1989) Nucleic Acid Res. 17, 8543–8551.
10. Sanger, F. and Coulson, A. (1975) J. Mol. Biol. 94, 441–448.
11. Hattori, M. and Sakak, Y. (1982) Anal. Biochem. 152, 232–238.
12. Pearson, W.R. and Lipman, D.J. (1988) Pro. Natl. Acad. Sci. USA. 85, 2444–2448.
13. Cabot, E.L. and Beckenbach, A.T. (1989) CABIOS 5, 233–234.
14. Schneller, J.M., Scheider, C. and Stahl, A.J. (1978) Bi0chem. Biophys. Res. Commu. 85, 1392–1399.
15. Natsoulis, G., Hilger, F. and Fink, G. (1986) Cell 46, 235–243.
16. Chatton, B., Walter, P., Ebel, J-P., Lacroute, F. and Fasiolo, F. (1988) J. Biol. Chem. 263, 52–57.
17. Watson, J.D., Hopkins, N.H., Roberts, J.W., Steitz, J.A. and Weiner, A.M. (1987). Molecular Biology of The Gene (4th ed). The Benjamin/Cumming publishing company, Inc. Menlo park, California. Vol. 1.
18. Borst, P. (1986) Ann. Rev. Biochem. 55, 701–732.
19. Bektesh, S., Van Doren, K. and Hirsh, D.L. (1988) Genes and Development 2, 1277–1283.
20. Eriani, G., Delarue, M., Poch, O.,Gangloff, J. and Moras, D. (1990) Nature 347, 303–206.