Genomic organization, transcriptional mapping, and evolutionary implications of the human bi-directional histidyl-tRNA synthetase locus (HARS/HARSL)

Terrance P. O'Hanlon* and Frederick W. Miller

Environmental Autoimmunity Group, National Institute of Environmental Health Sciences, National Institutes of Health, Bethesda, MD 20892, USA

Abstract

Histidyl-tRNA synthetase catalyses the covalent ligation of histidine to its cognate tRNA as an early step in protein biosynthesis. In humans, the histidyl-tRNA synthetase gene (HARS) is oriented opposite of a synthetase-like gene (HARSL) that bears striking homology to HARS. In this report, we describe the genomic organization of the HARS/HARSL locus and map multiple transcripts originating from a bi-directional promoter controlling the differential expression of these genes. The HARS and HARSL genes each contain 13 exons with strong structural and sequence homology over exons 3–12. HARS transcripts originate from two distinct promoters; a cluster of short transcripts map 15–65 bp upstream of the HARS ORF while a single, longer transcript (352 bp 5′-UTR) maps to a distal promoter. Similarly, multiple HARSL transcripts (mapping 10–198 bp upstream of its ORF) are produced by the shared bi-directional promoter. Human and rodent HARS/HARSL loci are homologous and support a model of inverted gene duplication to explain the emergence of HARSL during mammalian evolution. © 2002 Elsevier Science (USA). All rights reserved.

Keywords: Histidyl-tRNA synthetase; Transcript mapping; Genomic structure; HARS; HARSL

The aminoacyl-tRNA synthetases (aaRSs) comprise a ubiquitous family of structurally and functionally related enzymes that play a fundamental role in synthesizing the full complement of charged aminoacyl-tRNAs as an early step in protein biosynthesis. Besides their critical housekeeping functions, it has become increasingly recognized that various aaRS gene family members play important roles in many physiologic processes beyond protein biosynthesis. Non-traditional aaRS functions include regulation of RNA transcription, splicing, and translation as well as cytokine activities upon proteolytic cleavage during apoptosis [1–3].

Histidyl-tRNA synthetase (HARS), the enzyme that synthesizes His-tRNA, is one member of the aaRS gene family that has been well characterized in a variety of species including humans, owing in part to its targeting by autoantibodies in patients with the autoimmune muscle diseases polymyositis and dermatomyositis [4–6]. In humans and rodents, the HisRS open reading frame (ORF) is preceded by a bi-directional promoter that coordinates the transcription of multiple mRNAs in opposite directions [7,8]. Our laboratory has more recently identified the opposite-strand transcript in humans as a close structural homologue of HisRS [9]. This synthetase-like gene (HARSL), formerly termed HO3, shares approximately 74% nucleotide sequence identity with the ORF of the HisRS gene (HARS). Despite their similarity, the HARS and HARSL genes have a prominent point of divergence within regions encoding the amino-terminal domain of the respective polypeptides (approximately 18% identity over the first 60 aa). While the function of the HARSL protein is unknown, HARS and HARSL transcripts exhibit very distinct patterns of gene expression in a variety of human tissue and cell types [9,10]. Curiously, in 293T cells, the HARSL polypeptide has been shown to localize to both the cytoplasm and nucleus and form specific complexes with HIV matrix protein thereby playing an important role in virion assembly [10].

In humans, a bi-directional promoter is positioned within a 344 bp intergenic sequence separating the

* Corresponding author. Fax: +1-301-480-4127.
  E-mail address: ohanlon@niehs.nih.gov (T.P. O'Hanlon).

HARSL and HARS ORFs [8,9]. This GC-rich domain lacks identifiable TATA boxes and other canonical elements of RNA polymerase II promoters and appears to initiate the synthesis of multiple transcripts with varying lengths of 5′-untranslated region (5′-UTR) [8]. Mapping promoters with multiple transcriptional start points (tsp) by traditional methods (i.e., primer extension and nuclease digestion) is often confounded by the inability to distinguish full-length transcripts from various methodological artifacts such as incomplete extension and/or degradation products [7,11]. To overcome these limitations, we have employed a novel, high-resolution tsp mapping technique, RNA ligase-mediated rapid amplification of cDNA ends (RLM-RACE), to selectively amplify and sequence directly the 5′ ends of full-length, gene-specific transcripts [12,13].

In this report, we describe the complete genomic organization (intron/exon structure) and high-resolution tsp mapping of the human HARSL/HARS bi-directional locus. In addition, a comparison of corresponding rodent and human loci suggests strong evolutionary conservation of HARSL and HARS gene products and further supports a recently proposed model of inverted duplication for the emergence of the HARSL gene.

Materials and methods

Bioinformatics. The genomic organization (intron/exon boundries) of the HARSL and HARS genes was determined by screening a human genomic database (www.doubletwist.com) with the respective full-length cDNA sequences (Accession Nos. NM_012208 and NM_002109). The position and size of corresponding introns and exons were predicted using Doubletwist's Genomic Viewer software (Ver 3.2) on a human chromosome 5 BAC clone (119j3 LBNL H175; Accession No. AC005369). Alignments of cDNA and corresponding genomic exonic sequences were manually edited to determine precise splice site donor and acceptor sites according to established consensus using the human section of the GENIO/seq database (http://genio.informatik.uni-stuttgart.de/GENIO/splice/). The corrected genomic map was identical to and validated corresponding loci within the most recent build (No. 28) of the NCBI human chromosome 5 working draft sequence as viewed by LocusLink (Accession No. NT_006447). Nucleotide sequences spanning the predicted 5′-UTR of the HARSL and HARS genes were used to screen the human expressed sequence tag (EST) database using BLAST (www.ncbi.nlm.nih.gov/blast/). Nucleotide and amino acid alignments between corresponding HARSL and HARS gene sequences (human and murine) were performed using Vector NTI AlignX software (Ver 7.1). Gene-specific PCR primers were designed using Vector NTI oligonucleotide design software (Ver 7.1).

Transcriptional start point mapping. A commercially prepared human kidney RLM-RACE cDNA library (Ambion, Austin, TX) was screened for full-length 5′-UTR sequences using a combination of adapter and gene-specific primers as described below. Briefly, oligo-dT-purified mRNA was treated to remove 5′-phosphate groups from incomplete or fragmented transcripts. The 5′ cap structure protecting full-length transcripts from dephosphorylation was subsequently removed to expose underlying 5′-phosphate acceptor groups. A synthetic RNA adaptor sequence was ligated exclusively to the 5′ end of phosphorylated full-length transcripts prior to random-primed cDNA synthesis. The resulting 5′ adaptor-modified cDNA library was PCR amplified using primers complementary to adaptor and gene-specific sequences (HARSL and HARS) according to manufacturer's recommendations (5′ RLM-RACE Adapter 5′-gctgatggcgatgaatgaacactg-3′; 5′ RLM-RACE Adapter (nested) 5′-cgcggatccgaacactgcgtttgctggctttgatc-3′; 3′ HARSL 5′-tctgcaacctggctttggcaac-3′; 3′ HARSL (nested) 5′-ccaggccctcctgggaagaagtcc-3′; 3′ HARS (short transcripts) 5′-aacaccttctcgcgaactgccatc-3′; 3′ HARS (short transcripts, nested) 5′-tttcatcagg acccagctgtgcct-3′; 3′ HARS (long transcript) 5′-tcgtgactagtgaggtgcgcaaac-3′; 3′ HARS (long transcript, nested) 5′-cctggtgtctgacccgcctcctc-3′). Primary PCR amplification reactions were performed using antisense gene-specific primers (HARSL and HARS) mapping immediately downstream of the respective translational start codons (the HARS long transcript primer was positioned 198–222 bp upstream of the HARS start codon). Primary amplifications (50 μl) were performed for a total of 35 cycles at an annealing temperature (T_a) of 60 °C. Primary amplification products (2 μl) were subjected to a second round of PCR (35 cycles; $T_a = 62$ °C) using nested primers to ensure specificity and adequate yield for analysis. Secondary amplification products were size-fractionated in 2% TBE agarose gels, stained with ethidium bromide, and scanned on a Typhoon model 9200 fluorescent scanner (Amersham Pharmacia, Piscataway, NJ). Amplification products were gel-purified, ligated within the pCR 2.1 T/A cloning vector (Invitrogen, San Diego, CA), and sequenced subsequently on an ABI model 377 automated DNA sequencer as described previously [14].

Results

The genomic organization of the human HARSL/HARS locus (chromosome 5 position q31.3) was determined as described in Materials and methods. As shown in Fig. 1A, both the HARSL and HARS genes are comprised of 13 exons spanning approximately 7.9 and 17.4 kb, respectively. HARSL and HARS intron/exon splice site junctions conformed to known consensus sequences and mapped between in-frame codons with the exception of intron 8 (Figs. 1B and C, respectively). A comparison of HARSL and HARS genomic structure revealed strong conservation of exon size (bp) and primary sequence homology (nucleotide and amino acid) for exons 3–12 despite considerable variations in the sizes of corresponding introns and 3′-UTRs (see Table 1). For exons 4, 6, 7, 9, and 10, amino acid sequence similarities exceeded 90%. In contrast, significant sequence variation was observed between polypeptide coding domains in exons 1, 2, and 13 (amino acid sequence similarities of 20.5%, 40.0%, and 43.5%, respectively).

The detection of murine homologues for HARSL and HARS (Accession Nos. NM_080636 and NM_008214, respectively) suggests that the emergence of the HARSL gene preceded the evolutionary divergence of humans and mice, which has been estimated at approximately 100 million years ago [15]. As shown in Table 2, amino acid sequence alignments of human and murine HARSL and HARS polypeptides revealed high degrees of homology both within and between species. Together, these data illustrate the strong evolutionary conservation of both HARSL and HARS genes.

In an effort to better define the human HARSL/HARS bi-directional promoter, we mapped multiple

A

HARS

HARSL

B

	3′Exon Donor / 5′Intron … 3′Intron / 5′Exon Acceptor
	ATG CCC CTG
(1)	CAA AGC CAG / gtgagcga … tctgccag / GTT GCA GAG
(2)	ACC CCA AAG / gtaatact … gtttccag / GGT ACC AGG
(3)	GAG CTG AAG / gtaagggg … ccctatag / GAA ACC CTG
(4)	GAC CTT ACT / gtatcctt … tatgtgag / GTT CCC TTT
(5)	TGC CAG TGT / gtaagtga … tgctgtag / GAT TTT GAC
(6)	CTC ATT AAG / gtgaggcc … gttctcag / GTA AAT GAC
(7)	CTA GAC AAG / gtaacaaa … ttatgcag / ATG GCT TGG
(8)	CAG TGT CAT G/ gtaagaac … ctctccag / GT GGG GTA
(9)	GCT GAT AAG / gtaagctg … ccccacag / ATC TCC TTT
(10)	AGG ATG AAG / gtaggtcc … cttattag / ACC AAA GGT
(11)	GGA ATC AAG / gtatggtg … cttttcag / GCA GAG ATG
(12)	AGA GAG GAG / gtgagtgg … tgtttcag / GTG GCC ATT
(13)	TCT GAG TCT (**tga**)

C

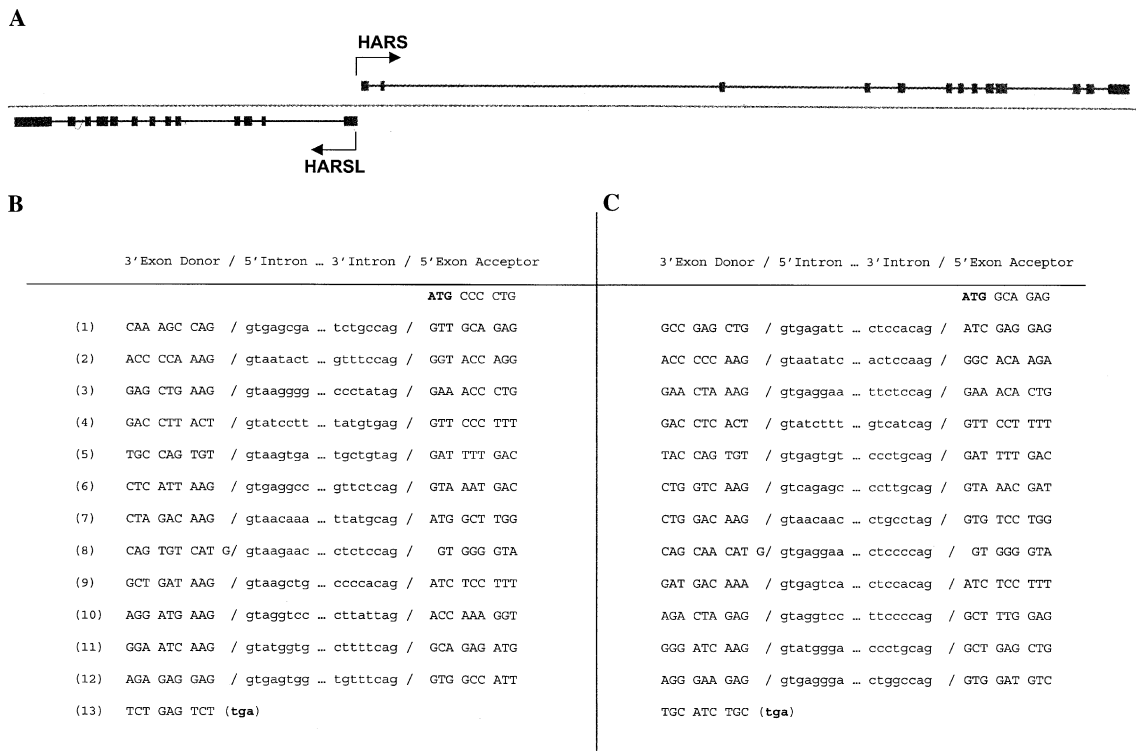| 3′Exon Donor / 5′Intron … 3′Intron / 5′Exon Acceptor |
|---|
| **ATG** GCA GAG |
| GCC GAG CTG / gtgagatt … ctccacag / ATC GAG GAG |
| ACC CCC AAG / gtaatatc … actccaag / GGC ACA AGA |
| GAA CTA AAG / gtgaggaa … ttctccag / GAA ACA CTG |
| GAC CTC ACT / gtatcttt … gtcatcag / GTT CCT TTT |
| TAC CAG TGT / gtgagtgt … ccctgcag / GAT TTT GAC |
| CTG GTC AAG / gtcagagc … ccttgcag / GTA AAC GAT |
| CTG GAC AAG / gtaacaac … ctgcctag / GTG TCC TGG |
| CAG CAA CAT G/ gtgaggaa … ctccccag / GT GGG GTA |
| GAT GAC AAA / gtgagtca … ctccacag / ATC TCC TTT |
| AGA CTA GAG / gtaggtcc … ttccccag / GCT TTG GAG |
| GGG ATC AAG / gtatggga … ccctgcag / GCT GAG CTG |
| AGG GAA GAG / gtgaggga … ctggccag / GTG GAT GTC |
| TGC ATC TGC (**tga**) |

Fig. 1. Genomic organization of the human HARSL/HARS locus. (A) Schematic illustration of the intron/exon distribution of opposite-strand HARSL (7.9 kb) and HARS (17.4 kb) genes is shown. Arrows denote the direction of transcription from the intervening bi-directional promoter. The nucleotide sequence of HARSL (B) and HARS (C) intron/exon splice site junctions was determined as described in Materials and methods. Exon numbers are designated in parentheses left of the corresponding 3′ exon donor sequences.

Table 1
Comparative genomic structure and homology of HARSL and HARS genes

	Intron (bp)		Exon (bp)		Exon homology (%)	
	HARSL	HARS	HARSL	HARS	Nucleotide	Amino acid[a]
1	1833	270	$\leqslant$306 (108)	442/$\leqslant$155[b] (90)	39.8[c]	20.5/10.3
2	269	7622	75	90	55.6	40.0/30.0
3	130	3216	120	120	65.8	77.5/67.5
4	1236	660	96	96	78.1	90.6/87.5
5	104	986	126	126	73.8	83.3/66.7
6	257	154	108	108	79.6	97.2/88.9
7	319	234	99	99	84.8	93.9/81.8
8	329	210	94	94	79.8	83.9/71.0
9	93	92	128	128	83.6	95.3/90.7
10	162	1519	243	243	85.6	93.8/92.6
11	246	191	117	117	70.1	79.5/66.7
12	414	350	147	147	78.2	85.7/79.6
13			813 (57)	422 (69)	52.8[c]	43.5/30.4

Note. Sizes of translated sequences within exons 1 and 13 are shown in parentheses.
[a] Amino acid: similarity/identity.
[b] Denotes the size of exon 1 from corresponding long and short HARS transcripts.
[c] Homology comparison of translated sequences.

transcriptional start sites using a novel, high-resolution technique, RLM-RACE, designed to selectively amplify the 5′-UTR of full-length, capped mRNAs. As described in Materials and methods, we amplified the 5′ ends of full-length HARSL and HARS transcripts from a human kidney cDNA library; a tissue previously determined to express high steady-state mRNA levels of both HARSL and HARS [9]. Nested amplifications of HARSL and HARS transcripts using primers positioned downstream of the corresponding translational start codons (+13–36 and +125–148 relative to the (+1)ATG, respectively) detected multiple mRNA species with varying lengths of 5′-UTR mapping within the shared bi-directional promoter region (Fig. 2, lanes 2

Table 2
Amino acid sequence homology of human and murine HARSL and HARS genes

Sequence alignment (aa)	Similarity/identity (%)
Human HARSL and Human HARS	79.3/70.8
Murine HARSL and Murine HARS	75.6/67.4
Murine HARS and Human HARS	96.3/94.9
Murine HARSL and Human HARSL	88.2/83.9

and 3). Amplified transcripts in the approximate size range 80–280 bp (HARSL) and 200–260 bp (HARS) were purified, ligated within a T/A cloning vector (pCR 2.1), and subsequently sequenced as described in Materials and methods. The resulting clones were aligned with complementary genomic DNA sequences and used to map the 5′ nucleotide (i.e., cap site) of full-length 5′-UTR sequences (Fig. 3). HARSL and HARS transcripts with multiple tsp were detected mapping in a range of 10–198 and 15–65 bp upstream of their respective translational start codons. A strong cluster of opposite-strand HARSL and HARS tsp were detected flanking a 80 bp domain mapping 70–150 bp upstream of the HARS ORF. This domain contained an A/T-rich sequence (AAATAATTTTTT) that might possibly serve a role in transcription. We also detected smaller clusters of transcripts with shorter 5′-UTRs mapping upstream of each ORF. In addition, a subset of alternatively spliced HARSL transcripts was detected that contained a 91 bp 5′-UTR exon mapping 42–133 bp upstream of the HARSL ORF. Perfect splice site consensus sequences were observed flanking the 5′-UTR exon and, in some instances, we detected alternatively spliced transcripts mapping to identical HARSL transcriptional start sites.
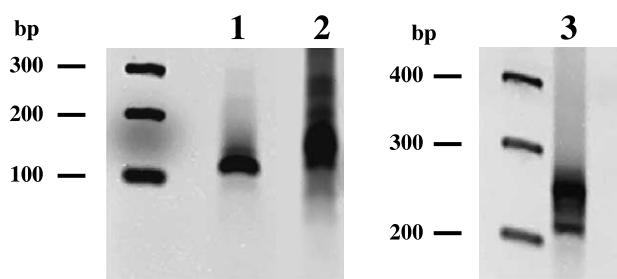

Fig. 2. PCR amplification of full-length 5′-UTRs of HARSL and HARS transcripts by RLM-RACE. A human kidney RLM-RACE cDNA library was PCR amplified using primers complementary to the 5′ adapter and gene-specific sequences mapping immediately downstream of the HARSL (lane 2) and HARS (lane 3) ORFs as described in Materials and methods. Based upon reports of HARS transcripts with longer 5′-UTRs (approximately 380 and 455 bp) [8,16], a similar amplification was performed using nested HARS gene-specific primers mapping approximately 275 bp upstream of the HARS ORF (lane 1). Nested amplification products were size-fractionated in 2% TBE agarose gels, stained with ethidium bromide, and subsequently detected with a fluorescent scanner. A DNA size-standard (100 bp ladder) is shown to the left of each gel. Note that amplified products include an additional 38 bp of synthetic adapter sequence positioned adjacent to the cap site of full-length 5′-UTRs.
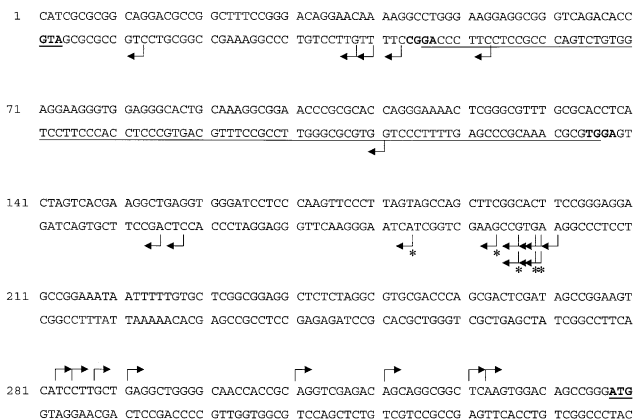

Fig. 3. Transcriptional start site mapping of HARSL and HARS genes. HARSL and HARS 5′-UTR gene-specific amplification products (see Fig. 2) were gel-purified, ligated within the pCR 2.1 T/A cloning vector, and sequenced subsequently as described in Materials and methods. The resulting clones were aligned with complementary genomic DNA sequences and the position of the 5′ cap site (i.e., the junction of 5′ adaptor and gene-specific sequences) was mapped within a 344 bp domain positioned between opposite-strand translational start codons (HARSL (**GTA**) and HARS (**ATG**)). The positions of multiple transcriptional start sites for HARSL (lower strand) and HARS (upper strand) are designated with arrows showing the direction of transcription. A 91 bp alternatively spliced exon mapping within the 5′-UTR of multiple HARSL transcripts is underlined; splice site junctions are highlighted in bold type. Note that, in some instances, alternatively spliced HARSL transcripts (∗) were detected with identical transcriptional start sites.

Previous reports have suggested that in addition to the short class of mRNAs mapping immediately upstream of the HARS ORF, a second class of longer HARS transcripts may derive from a more distal, independent promoter overlapping the first exon of the HARSL gene [8,16]. We performed an RLM-RACE analysis using primers mapping downstream of this putative transcript and successfully amplified a single mRNA species mapping 352 bp upstream of the HARS ORF (Fig. 2, lane 1). These data suggest that at least two promoters separated by approximately 300 bp direct the synthesis of multiple HARS transcripts.

Discussion

We have described the genomic architecture (intron/exon structure) and transcriptional start site map for the human HARSL/HARS locus (chromosome 5q31.3), which contains genes encoding HisRS and a HisRS-like polypeptide. Unlike other "housekeeping" genes, the organization and expression of the human HisRS gene is intimately associated with a close structural homologue, HARSL, which is transcribed in the opposite direction from a shared bi-directional promoter. While the function of the HARSL polypeptide is unknown, it is clear that its genomic organization has been preserved since the estimated divergence of humans and rodents approximately

100 million years ago [15]. A bi-directional promoter element flanking the HisRS gene has been described in both rodents and humans [7,8]. Our laboratory subsequently identified the opposite-strand transcript in humans as encoding a HisRS-like polypeptide formerly designated HO3 [9]. More recently, a full-length murine homologue of HARSL (NM_080636) has been reported that bears striking amino acid sequence homology (88.2%) to its human counterpart (see Table 2) [17].

The HARSL gene apparently emerged as the result of an inverted duplication through the first intron of the HisRS gene [18]. Both HARSL and HARS share a common genomic architecture (exons 2–13) with more distantly related HisRS genes from hamster and the Japanese pufferfish, *Fugu rubripes* [7,18]. In fact, *Fugu* and human HisRS genes share approximately 84% amino acid sequence homology. Although the *Fugu* genome lacks the upstream HARSL homologue, the *Fugu* HisRS gene itself contains a highly divergent first exon that the authors attribute to a translocation event (exon shuffling) through the first intron following the evolutionary split of mammals and fishes approximately 450 million years ago [19]. The authors proposed a similar mechanism to explain the divergent amino terminal domain (mapping largely to exon 1) of the human HARSL gene. Since its emergence, the structural framework and primary sequence of HARSL exons have been largely preserved with HARS (exons 2–13) despite considerable variations in the lengths of the respective introns and 3'-UTRs (see Fig. 1 and Table 1) [8]. Moreover, as seen in Table 2, the inter-species amino acid sequence similarity of HARSL ORFs (88.2%) exceeds intra-species comparisons of both human and murine HARSL and HARS ORFs (79.3% and 75.6%, respectively). Together, these data argue that strong evolutionary pressures have worked to maintain the nature and integrity of the HARSL gene in humans.

It has become increasing recognized that aaRS are multi-functional proteins that play broader roles in physiology than cellular protein biosynthesis. aaRS and polypeptides bearing aaRS-like domains have been associated with a host of non-traditional functions [1–3]. One fascinating example includes the cytokine activities associated with proteolytic fragments of human TyrRS [20]. Under apoptotic conditions, TyrRS is secreted and cleaved by an extracellular protease (leukocyte elastase) liberating amino- and carboxy-terminal fragments with demonstrable cytokine (interleukin-8-like) and chemokine (endothelial monocyte-activating polypeptide II-like) activities, respectively. It is remarkable that the precursor form of the genuine endothelial monocyte-activating polypeptide II molecule (pro-EMAP II) contains an aaRS-like domain, associates with a cytosolic multi-aaRS complex, and helps facilitate aminoacylation [21].

The HisRS polypeptide has been extensively characterized owing, at least in part, to its targeting by autoantibodies (anti-Jo-1) in a subset of patients with the autoimmune muscle diseases polymyositis and dermatomyositis. Patients with anti-Jo-1 autoantibodies are distinguished from other myositis patients by a unique matrix of epidemiologic, clinical, genetic, and prognostic features [6]. Anti-Jo-1 autoantibodies are directed primarily against an evolutionarily conserved B cell epitope mapping within the amino-terminal domain (aa 1–60) of HisRS; a domain that is highly divergent in corresponding regions of the HARSL molecule [22,23]. While these data suggest that the HARSL polypeptide is not the primary target of anti-Jo-1 autoantibodies, it does not rule out the possibility that T cell and/or secondary B cell epitopes shared by HARS and HARSL might play a role in initiating autoimmune responses.

The function of the human HisRS homologue, HARSL, is currently unknown. The conservation of structural motifs associated with class II aaRS and HisRS in particular leads us to speculate that HARSL might encode a mitochondrial HisRS [9]. Conflicting lines of evidence question this hypothesis. A predictive analysis for mitochondrial targeting sequences (http://www.mips.biochem.mpg.de/cgi-bin/proj/medgen/mitofilter) revealed a candidate domain within the hydrophobic amino-terminus of the HARSL polypeptide (aa 1–60). Transient expression of the HARSL cDNA in human 293T cells, however, revealed localization in both cytoplasm and nucleus [10]. Oddly, upon HIV infection of cultured 293T cells, the HARSL polypeptide (formerly designated HO3) forms specific complexes with HIV matrix protein and aids viral maturation [10]. Based upon these findings, it is unclear whether the previously described nuclear localization of HisRS may be in part attributable to antibodies cross-reacting with epitopes shared with the HARSL polypeptide [24]. Mitogen-induced nucleolar localization was recently described for human MetRS where that protein presumably plays a role in rRNA synthesis [25]. It is noteworthy that sequences within the amino-terminal domain (aa 1–60) of HisRS necessary for enzymatic activity are lacking within corresponding regions of the HARSL polypeptide [22]. These data suggest that while the HARSL polypeptide may not be enzymatically active, it may possibly retain properties associated with aaRS (e.g., ATP, amino acid, and/or tRNA binding). On-going studies in our laboratory are examining the possible significance of nuclear trafficking and other functional attributes of the HARSL polypeptide.

HARSL and HARS exhibit distinctive patterns of steady-state mRNA expression in a variety of human tissue and cell types suggesting elaborate mechanisms of transcriptional control [9,10]. The sizes of mature HARSL and HARS mRNAs (approximately 2.5 and 2.0 kb, respectively) are consistent with the sizes of the multiple transcripts identified initiating from a shared bi-directional promoter positioned within a 344 bp region separating the opposite-strand ORFs [8–10]. The

bi-directional promoter region lacks canonical elements of RNA polymerase II promoters (e.g., TATA and CAAT boxes) and directs the synthesis of multiple, opposite-strand transcripts with varying lengths of 5′-UTR [7,8]. Tsui et al. [8] previously described the synthesis of multiple transcripts initiating approximately 38–82 bp upstream of the human HARS ORF and identified a functional bi-directional promoter element mapping within 122 bp of the HARS ORF.

As described in Materials and methods, we have utilized a high-resolution, 5′ cap site mapping technique, RLM-RACE, designed to avoid difficulties often associated with mapping multiple tsp promoters using traditional primer extension and nuclease digestion assays (e.g., incomplete extension and/or degradation products) [11–13]. We have similarly identified multiple HARS transcripts with 5′-UTRs extending 15–65 bp (see Fig. 3). A recent scan of the human EST database (12.8×10^6 independent sequence entries) indicated that a majority of HARS transcripts identified extend 50–75 bp upstream of the ORF (data not shown). In addition, we have mapped multiple tsp extending 10–198 bp upstream of the human HARSL ORF (see Fig. 3). A cluster of HARSL transcripts (5′-UTRs extending 191–198 bp) flank the minimal bi-directional promoter element defined by Tsui et al. [8]. A similar scan of the human EST database suggests that a majority of HARSL transcripts initiate 150–200 bp upstream of the HARSL ORF. Of interest, we identified a subset of HARSL transcripts that contain an alternatively spliced exon mapping within the 5′-UTR. A putative set of alternatively spliced 5′-UTR exons has also been identified upstream of the human gene encoding TrpRS [26]. The extent and significance of these alternative-splicing events are unknown but are currently under investigation.

In addition to the multiple transcripts mapping immediately upstream of the HARS ORF, a second class of longer HARS mRNAs (approximately 380 and 455 bp 5′-UTRs) presumably initiating from more distal promoter elements has been reported [8,16]. Using RLM-RACE, we selectively amplified a HARS transcript initiating 352 bp upstream of the HARS ORF and overlapping the opposite strand of HARSL exon 1. The basis for variations in the reported sizes of these longer HARS transcripts is unclear. It is conceivable that different mRNA source materials (i.e., cultured cell lines (HeLa and lymphoblastoid) vs. human kidney) and/or different methodological approaches may possibly account for these dissimilar findings. Nevertheless, the size of HARS mRNA readily detected on multiple tissue Northern blots (approximately 2.0 kb) is consistent with the detection of shorter HARS transcripts (5′-UTR less than 100 bp) derived form the bi-directional promoter element [9,10]. Moreover, a screen of the human EST database failed to detect transcripts consistent with the longer class of HARS mRNAs identified (352, 380, and 455 bp 5′-UTRs). These data suggest that the longer class of HARS mRNAs expressed from a distal promoter element(s) are lower abundance transcripts.

In summary, our and other data suggest that the use of multiple and bi-directional promoter elements, and possibly alternative splicing of 5′-UTR exons, contribute to complex and coordinated patterns of gene expression from the evolutionarily conserved human HARSL/HARS locus. We are currently exploring the mechanistic basis and regulatory implications of HARSL and HARS gene expression in a variety of human tissues.

Acknowledgments

References

[1] S.A. Martinis, P. Plateau, J. Cavarelli, C. Florentz, Biochimie 81 (1999) 683–700.
[2] K.A. Ivanov, N.A. Moor, O.I. Lavrik, Biochemistry (Mosc.) 65 (2000) 888–897.
[3] P. Schimmel, D.P. Ribas, Trends Biochem. Sci. 25 (2000) 207–209.
[4] W. Freist, J.F. Verhey, A. Ruhlmann, D.H. Gauss, J.G. Arnez, Biol. Chem. 380 (1999) 623–646.
[5] F.W. Miller, K.A. Waite, T. Biswas, P.H. Plotz, Proc. Natl. Acad. Sci. USA 87 (1990) 9933–9937.
[6] F.W. Miller, JAMA 270 (1993) 1846–1849.
[7] F.W. Tsui, L. Siminovitch, Gene 61 (1987) 349–361.
[8] H.W. Tsui, S. Mok, L. de Souza, A. Martin, F.W. Tsui, Gene 131 (1993) 201–208.
[9] T.P. O'Hanlon, N. Raben, F.W. Miller, Biochem. Biophys. Res. Commun. 210 (1995) 556–566.
[10] J. Lama, D. Trono, J. Virol. 72 (1998) 1671–1676.
[11] Y. Lin, T.A. Ince, K.W. Scotto, Biochemistry 40 (2001) 12959–12966.
[12] K. Maruyama, S. Sugano, Gene 138 (1994) 171–174.
[13] B.C. Schaefer, Anal. Biochem. 227 (1995) 255–273.
[14] T.P. O'Hanlon, O.J. Lawless, W.E. Katzin, L.J. Feng, F.W. Miller, J. Autoimmun. 14 (2000) 283–293.
[15] D.W. Burt, C. Bruley, I.C. Dunn, C.T. Jones, A. Ramage, A.S. Law, D.R. Morrice, I.R. Paton, J. Smith, D. Windsor, A. Sazanov, R. Fries, D. Waddington, Nature 402 (1999) 411–413.
[16] N. Raben, F. Borriello, J. Amin, R. Horwitz, D. Fraser, P. Plotz, Nucleic Acids Res. 20 (1992) 1075–1081.
[17] P. Carninci, Y. Shibata, N. Hayatsu, Y. Sugahara, K. Shibata, M. Itoh, H. Konno, Y. Okazaki, M. Muramatsu, Y. Hayashizaki, Genome Res. 10 (2000) 1617–1630.
[18] S. Brenner, L.M. Corrochano, Proc. Natl. Acad. Sci. USA 93 (1996) 8485–8489.
[19] R.B. Walter, D.C. Morizot, Adv. Struct. Biol. 4 (1996) 1–24.
[20] K. Wakasugi, P. Schimmel, Science 284 (1999) 147–151.
[21] Y. Kim, J. Shin, R. Li, C. Cheong, K. Kim, S. Kim, J. Biol. Chem. 275 (2000) 27062–27068.
[22] N. Raben, R. Nichols, J. Dohlman, P. McPhie, V. Sridhar, C. Hyde, R. Leff, P. Plotz, J. Biol. Chem. 269 (1994) 24277–24283.
[23] A. Martin, M.J. Shulman, F.W. Tsui, FASEB J. 9 (1995) 1226–1233.
[24] D. Vazquez-Abad, J.H. Carson, N. Rothfield, Cell Tissue Res. 286 (1996) 487–491.
[25] Y.G. Ko, Y.S. Kang, E.K. Kim, S.G. Park, S. Kim, J. Cell Biol. 149 (2000) 567–574.
[26] L.Y. Frolova, A.Y. Grigorieva, M.A. Sudomoina, L.L. Kisselev, Gene 128 (1993) 237–245.