
Isolation, structure and expression of mammalian genes for histidyl-tRNA synthetase

Florence W.L. Tsui* and L.Siminovitch

Mount Sinai Hospital Research Institute, Toronto, Ontario M5G 1X5 and Department of Medical Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada

Received November 25, 1986; Revised and Accepted February 24, 1987

ABSTRACT

A full length cDNA clone that codes for human histidyl-tRNA synthetase (HRS) and cDNA clones that span the full length transcript of hamster HRS have been isolated. The full length human HRS cDNA was expressed after transfection into Cos 1 cells and a CHO ts mutant defective in the gene for HRS. The complete nucleotide sequence of the hamster and human gene were obtained and extensive homologies were observed in three regions on comparing these sequences between themselves and with the sequence of HRS derived from yeast. These results provide unequivocal evidence that we have indeed cloned the hamster and human gene for HRS.

Three overlapping phage recombinants containing the complete hamster chromosomal gene for HRS have also been isolated. The genomic HRS is divided into 13 exons. The precise locations of each of the 5' and 3' exon-intron boundaries were defined by sequencing the appropriate regions of the cloned genomic DNA and aligning them with the sequence of HRS cDNAs. These studies provide the basis for future structural and functional analysis of the gene for HRS. In particular, it will be of interest to examine if different exons of HRS correlate to different domains of the HRS polypeptide.

INTRODUCTION

Aminoacyl-tRNA synthetases play a crucial role in protein synthesis. They are of particular interest as enzymes because they are involved in two steps during the aminoacylation reactions, first resulting in activation of the amino acid and then its attachment to the 3' end of its cognate tRNA. Although there is considerable information on the structure-function relationships of the bacterial aminoacyl-tRNA synthetases (1,2,3,4,5,6), much less is known about the mammalian enzymes, especially at the molecular level. Since development of such information necessarily first involves cloning of the genes, we recently initiated studies directed to the isolation of the cDNA

of one of these enzymes, histidyl-tRNA synthetase (HRS) from Chinese hamster ovary (CHO) cells. Accordingly, we developed histidinol resistant lines in which the gene for HRS was amplified 30 fold, and using these lines, we were able to isolate a 250 bp cDNA for the HRS gene (7).

This clone has provided the opportunity to investigate the structure, function and expression of the gene for HRS. In this paper, we report on (a) the isolation of full length cDNAs for hamster and human HRS, (b) the expression of the human HRS cDNA in recipient cells, (c) comparison of the nucleotide sequences of CHO and human cDNAs for HRS, and (d) comparison of the primary structure of four species of HRS --- bacteria, yeast, CHO and human. Since this gene has the same essential function through the animal kingdom, conservation of any part of the gene probably reflects important functional domains. We also report the isolation and characterization of the CHO chromosomal gene for HRS. We show that the complete gene spans 18 kb and contains 13 exons.

MATERIALS AND METHODS

Cloning of cDNAs for CHO and human histidyl-tRNA synthetase

The CHO cDNA library and the human SV40 transformed fibroblast cDNA library (both constructed in the pCD vector) were provided by Okayama (7). The CHO cDNA library in the pCD vector was screened by a 3' CHO cDNA probe (pHRS1, 250 bp) described previously (8). A second small CHO library (30,000 recombinants) was constructed using mRNA from histidinol resistant CHO cells, using the procedure of Gubler and Hoffman (9). This latter CHO cDNA library and the human cDNA library were screened by a more 5' genomic fragment (3 kb) of the HRS chromosomal gene.

cDNA probes were nick translated using the Amersham nick translation kit and (α - 32 P)dCTP (specific activity 3000Ci/mmol; Amersham) to a specific activity of $1-2 \times 10^8$ cpm/ μ g. Hybridization conditions were as described previously (8).

Screening of the genomic library for the CHO HRS chromosomal gene

A wild type CHO genomic library prepared in the λ phage EMBL3 was provided by Shotwell and Ray. Genomic clones were screened for inserts corresponding to HRS by in situ plaque hybridization (10). For more detailed analysis, genomic fragments

generated by digestion with BamH 1, Sal 1 or Hind III were subcloned into either psp64, psp65, GEM 1 or GEM 2 vectors.

Restriction maps and DNA sequence analysis

The location of restriction endonuclease cleavage sites in the inserted DNA was determined by digesting the DNA with several restriction enzymes under the conditions recommended by the manufacturer. All DNA sequencing was done by the Sanger dideoxy chain termination method (11). Dideoxynucleotide triphosphates, deoxynucleotide triphosphate were from the Bethesda Research Laboratory. DNA polymerase (Klenow fragment) was from Boehringer Mannheim. The cDNAs were either sequenced directly as double stranded restriction fragments or exonuclease III generated fragments subcloned in the vector GEM 1 or 2 (Promega) (12), using a Sp6 promoter primer or a T7 promoter primer (Promega), or sequenced using the 17mer universal primer (New England Biolabs) after subcloning into the phages M13mp18 and mp19 (Bethesda Research Laboratory). DNA sequences were established by analysis of both strands and analysis of fragments across the restriction sites. Genomic DNA fragments were sequenced directly as double stranded exonuclease III generated fragments subcloned in the vectors GEM 1 or 2 (Promega) (12), using a Sp6 promoter primer or a T7 promoter primet (Promega).

Transient expression of human HRS cDNA in Cos 1 cells

Cos 1 cells were transfected with recombinant plasmid DNA (human HRS cDNA in the expression vector pCD) in a calcium phosphate precipitate (13). After 72 hours, the transfected cells were trypsinized, washed with phosphate buffered saline, and lysed by vortexing in cold buffer (50 mM Tris hydrocholride, pH 7.4, 1 mM DTT, 5 mM EDTA, 20% glycerol, 0.5 mM phenylmethyl-sulfonyl fluoride (PMSF)). The lysates were centrifuged at 10,000 x g for 15 min. Determination of histidyl-tRNA synthetase activity was as reported previously (8).

Cell lysates were also denatured by boiling in 2% SDS and 0.5 mM DTT for 3 min., and run on a 7.5% SDS-acrylamide gel (14). The proteins were transferred to nitrocellulose in a Trans-blot cell (Biorad). The protein blots were then pretreated with 50 mM Tris HCl, pH 7.5, 0.15 M NaCl and 3% BSA (Sigma) for four hours and then treated with anti-HRS from polymyositis patient serum diluted

in 20 mM sodium phosphate pH 7.3, 0.15 M NaCl and 3% BSA. The blots were then washed five times with wash buffer containing 20 mM sodium phosphate, pH 7.3, 0.15 M NaCl, 0.25% Triton and 1% BSA. 2×10^6 cpm of I¹²⁵ protein A (New England Nuclear) was then added to the above wash buffer and the blots were incubated at 4°C overnight. Finally, the blots were washed extensively with the above wash buffer. Autoradiography of the filters was carried out at -70°C with XAR-5 Kodak film and an intensifying screen.

Selection of long term transfectants

D10-5 cells (CHO mutants that have a temperature sensitive HRS) (15) were co-transfected with recombinant plasmid DNA (human HRS cDNA in the expression vector pCD) and pSV₂ Neo plasmid DNA (19) by the calcium phosphate method (13). Forty-eight hours after transfection, the cells were doubly selected by growing in 400 µg/ml Geneticin (G418 sulfate; Gibco) at 39°C. Cells mock transfected or transfected with only pSV₂ Neo plasmid DNA were used as control. Single cell colonies were picked and expanded and RNA preparations were made as described below. No viable colonies were found in control transfected cells.

Northern blot analysis

Total RNA was isolated from cell cultures using the guanidinium isothiocyanate method followed by CsCl centrifugation (16). RNA preparations were run on 1.5% agarose gels containing formaldehyde (17). After transfer of RNA onto nitrocellulose, RNA blots were probed with a nick translated 3' Bgl I - BamH I restriction fragment of the human cDNA for HRS.

RESULTS

Isolation and characterizations of cDNA clones for histidyl tRNA synthetase

Since our cDNA clone only consisted of 250 bp (8), our first objective was to isolate full length clones of the hamster gene. A CHO cDNA library constructed with the pCD vector was screened with the 250 bp (pHRS1) clone (8) as the hybridization probe (see Materials and Methods). Ten different positive cDNA clones were isolated but the longest insert was only 660 bp (pHRS2). Restriction endonuclease and southern blot analysis of the restriction enzyme digested 660 bp pHRS2 (see Fig.1) indicated that pHRS1 was located close to the 3' end of the cDNA.

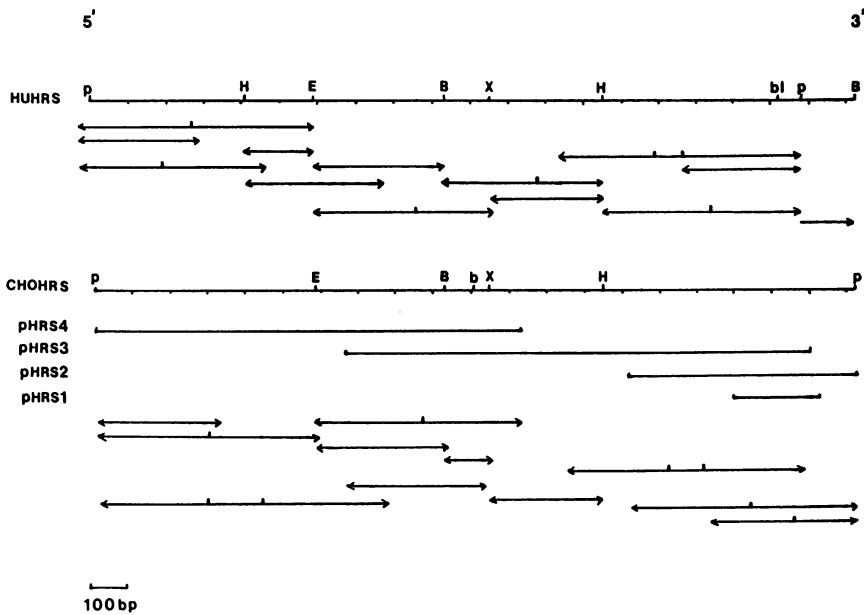


Figure 1. Restriction endonuclease map of human HRS cDNA clone (HuHRS) and a composite restriction endonuclease map of CHO HRS cDNA clones (pHRS1, pHRS2, pHRS3 and pHRS4). Enzymes represented are B, BamH I; bI, Bgl I; b, Bgl II; E, EcoR I; H, Hind III; P, Pst I; X, Xho I. At the bottom of each map are strategies used to determine the nucleotide sequence of the cDNA clones.

As a second approach and to obtain more clones, we decided to profit from the fact that our histidinol resistant CHO lines contained 30 fold increased levels of mRNA (8). We therefore constructed a small cDNA library using this mRNA (see Materials and Methods).

In work on the HRS genomic DNA (Fig. 6), using pHRS1 as probe, we had isolated a unique Sal I - BamH I fragment (from a phage clone called HRS λ 2) from a genomic library prepared in phage EMBL3. Because this fragment was about 3 kb in length and because it was known to be at the 5' end of the gene (this fragment hybridized to the HRS transcript in Northern blot analysis but did not hybridize to the cDNA pHRS2 which contained the complete 3' end of the HRS gene), we used it to probe the second library. It would be expected that we would probably identify cDNA clones with more 5' HRS sequences.

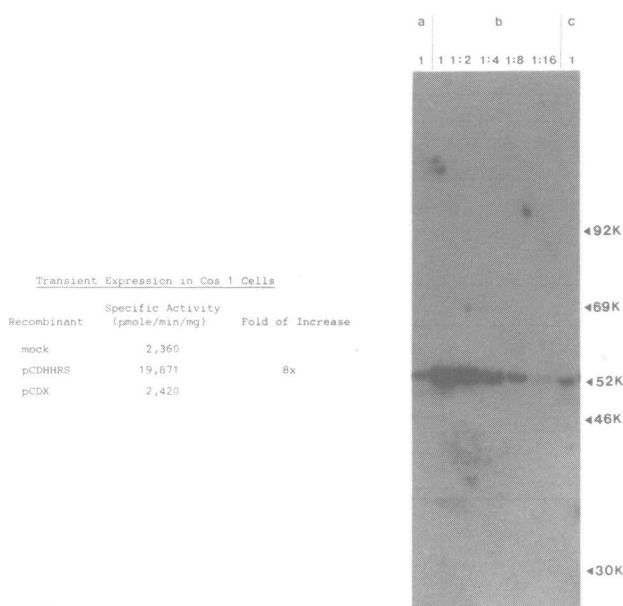


Figure 2. Transient expression in Cos 1 cells - immunoblotting of the transfected Cos 1 cells. lane a : HRS polypeptide from mock transfected Cos 1 cells; lanes b : serial dilution of HRS polypeptide from cells transfected with the full length human cDNA for HRS (pCDHRS); lane c : HRS polypeptide from cells transfected with control plasmid (pCDX).

Twelve positive recombinant plasmid clones were obtained from the screening of 30,000 recombinant clones. Restriction enzyme mapping revealed two clones with inserts of 1.14 kb and 1.13 kb (pHRS3 and pHRS4), respectively, (Fig. 1). The inserts of these two cDNA clones overlapped with each other and spanned 1.9 kb of the 5' end of cDNA for HRS. The 3' end of pHRS3 also overlapped with pHRS2 (Fig. 1). In total, 2.08 kb of cDNA sequences for HRS were obtained. This is the length to be expected from the size of the HRS gene transcript (see later) and it indicates that we have obtained the full length cDNA hamster gene transcript.

Since we were interested in the comparative analysis of the gene for HRS from various organisms, we then carried out experiments to isolate the human cDNA for HRS. A human SV40-transformed fibroblast library, constructed with the pCD vector (7), was

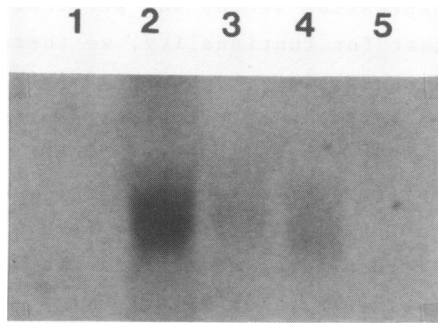


Figure 3. Northern blot analysis of long term transfectants. lane 1 : mRNA from D10-5 cells grown at permissive temperature; lanes 2-4 : mRNAs from three independently selected lines transfected with full length human cDNA for HRS; lane 5 : mRNA from a D10-5 revertant line grown at 39°C. mRNA preparations were subjected to electrophoresis in a 1.5% RNA-formaldehyde gel. The mRNA was transferred to nitrocellulose and hybridized with a nick translated 3' Bgl I - BamH I restriction fragment of the human cDNA for HRS in the presence of 50% formamide at 65°C.

screened using the Sal I - BamH I genomic fragment from HRS2 as hybridizing probe. 1.4×10^6 recombinant plasmids were screened and twenty independent positive cDNA clones were isolated. We were thus able to obtain a long insert of about 2.1 kb. Using this 2.1 kb fragment as a probe for Northern blot analysis of human lymphoblast mRNA, we identified a 2.1 kb RNA transcript, indicating that the human HRS cDNA is probably full length.

Thus we have succeeded in isolating full length human and hamster cDNAs for the gene for HRS. We have compared the restriction maps of the two cDNAs and these are shown in Fig. 1. There are clearly some common, as well as different, sites between the cDNAs of the two species, as would be expected from their nucleotide sequence comparison (see later).

Expression of human HRS cDNA

Considerable evidence was obtained in our earlier work (8) that the original cDNA clone represented a fragment of the HRS gene. However the availability of full length cDNAs provided the opportunity to obtain much more definitive proof for this contention, that is, we were now able to conduct a functional assay for the cloned gene.

The fact that the Okayama and Berg library was constructed in

the pCD mammalian expression vector was particularly useful for this purpose. To test for functionality, we therefore transfected the pCD plasmid DNA containing the human cDNA full length fragment into Cos 1 cells and assayed for transient expression of the gene. Cos 1 cells were used since the pCD vector contains the SV40 promoter and can replicate efficiently in this host in response to the endogenous large T antigen, and other permissive factors. As controls, cells were mock-infected or transfected with a control plasmid. Seventy-two hours after transfection, the HRS activities of the transfected Cos 1 cells were measured. As shown in Fig. 2, cells transfected with the human HRS cDNA showed an eight fold increase in HRS enzyme activity compared to that of mock transfected cells or cells transfected with control plasmid. The HRS proteins made by these transfected cells were also examined by immunoblotting. Total cell extracts were run on SDS-PAGE, blotted onto nitrocellulose, probed with polymyositis patient serum containing autoantibodies specific for HRS (18), followed by ^{125}I protein A. The results are portrayed in Fig. 2. The HRS protein of mock transfected cells and that of cells transfected with control plasmid are shown in lanes a and c, respectively. The proteins from cells transfected with the human HRS cDNA were serially diluted before loading on the gel. As shown in Fig. 2 lane b, there was also an eight fold increase in HRS protein when cells were transfected with human HRS cDNA.

In order to show that the result obtained in the transient expression study was not due to activation of the monkey HRS gene by transactivating factors, long term transfectant lines were selected and analysed. For this experiment, a CHO ts mutant (D10-5) (15) containing a temperature sensitive HRS gene was used as recipient. The cells were cotransfected with DNA from both the human cDNA clone and pSV₂ Neo (19) by calcium phosphate precipitation. Transfectants were then doubly selected by growth in G418 at 39°C. Single cell colonies were picked and expanded, and mRNA prepared from these transfectants. A Northern blot of these mRNAs was probed with a human HRS specific probe under conditions in which the probe did not cross hybridize with DNA from the CHO HRS gene. As shown in Fig. 3, the human HRS mRNA transcripts were only visible in the transfectants (lanes 2, 3 and 4) and not in

the parental ts mutants (lane 1) or revertants of the ts mutants (lane 5). These results provide unequivocal evidence for the identity between the human cDNA we have isolated and the gene for histidyl-tRNA synthetase. Parenthetically, they provide evidence that we have isolated a full length cDNA.

Nucleotide sequences of the mRNA coding for hamster and human HRS

To facilitate analysis of the HRS genomic DNA (17) and to allow an analysis of homologies between HRS from various species, we have sequenced the total human and hamster cDNAs. These sequences are shown in Fig. 4. As may be seen, both species of mRNAs for HRS have a 5' untranslated region of 77 nucleotides. The first Met codon was found at residues 78-81 in both species. The open reading frame for the CHO HRS mRNA lies from +1 (the first ATG codon) to +1524. The human HRS mRNA also contains a 1524 bp open reading frame. The predicted proteins encoded by the CHO and human HRS cDNA's both consist of 508 aa residues. The 3' untranslated regions of the CHO HRS mRNA contains 425 bp including 18 A's located at 13 bp downstream of the polyadenylation consensus sequence AATAAA. The 3' untranslated region of the human HRS mRNA contains 413 bp including a tail of 62 A's located at 14 bp downstream of the AATAAA sequence. Comparison of the CHO HRS and human HRS cDNAs at the nucleotide level shows that there is 71% homology at the 5' untranslated region, 82% homology at the protein coding region and 38% homology at the 3' untranslated region.

Homology among species of histidyl-tRNAs synthetase from E.coli, yeast, hamster and human

Since the nucleotide sequences of HRS from yeast and E.coli were available (20,21), it was next of interest to compare the homologies among the various enzymes species. This was done by examining the derived amino acid sequences. As may be seen in Fig. 5, there is 56% homology with both the hamster and human HRS from aa 20 to aa 272 of the yeast HRS. The second region of homology lies between aa 280 and 340 of the yeast HRS. This region is 53% homologous to the corresponding regions of the hamster and human HRS. The third region of homology lies from aa 396 to 473 of the yeast HRS, giving 42% homology with the corresponding carboxyl-terminal regions of hamster and human HRS.

AGCCGGAAGTCATCCGAACCTAGGCTAGGTCTACCAAAGCGGGTTAGAC AGCCGGAAGTCATCCCTTGCTGAGGCTGGGCAACCAACCAGGTCAGAC	CHOHS -28 HUHS -28
+1	
AGCAGGCTGTTCCTTTGGCAGCCAGGATGGCATCCCTTGGCTGGAAGA AGCAGGCTGCTGAACTTCAAGGAGAGCGCTGCAAGCCCTCCCTCGA	CHOHS 23 HUHS 23
CTTCTGCTGAATCCAGGCACAGACTGCTCAGCGCCCTCAAACAGCAGA GAGCCTGCTGAACTTCAAGGAGAGCGCTGCAAGCCCTCAAACAGCAGA	CHOHS 73 HUHS 73
AAGCCTCCGCTGACACGATCAGGAGGAGGTTGGCCAAACTCCTAAAAGT AGGCCAGCGCCAGCTGATCAGGAGGAGGTTGGCCAAACTCCTAAAAGT	CHOHS 123 HUHS 123
AAGGCACAGCTGGGCGATGATGAAAGCAAACAGAAATTTGCTCAAAAC AAGGCACAGCTGGGCTCTGATGAAAGCAAACAGAAATTTGCTCAAAAC	CHOHS 173 HUHS 173
CCCCAAGGGCACAAGAGACTACAGTCCCGGCAGATGGCAGTTCCGCAAA CCCCAAGGGCACAAGAGACTAAGTCCCGGCAGATGGCAGTTCCGCAAA	CHOHS 223 HUHS 223
AGGTCTTTCAGCTATCATCTGCTGCTTCAAAGCCATGGCCAGCAAGT AGGTCTTTCAGCTATCATCTGCTGCTTCAAAGCCATGGCCAGCAAGT	CHOHS 273 HUHS 273
ATTGACACTTCTCTCTTCTGACTAAAGAAACACTGATGGGAAAGTATGG ATTGATACACTCTCTCTTCTGACTAAAGAAACACTGATGGGAAAGTATGG	CHOHS 323 HUHS 323
GCAAGACTGCAAACTATCTATGACCTGAAAGATCAGGGCGGGGAGCTAC GCAAGACTGCAAACTATCTATGACCTGAAAGATCAGGGCGGGGAGCTAC	CHOHS 373 HUHS 373
TGTCCCTTCGCTATGACCTCACTGTCCCTTTGGAAGGTATCTGGCAATG TGTCCCTTCGCTATGACCTCACTGTCCCTTTGGAAGGTATCTGGCAATG	CHOHS 423 HUHS 423
AATAACTTGACCAACATTAACAGCTACCACATAGCAAAGGTATATCGCCG AATAACTTGACCAACATTAACAGCTACCACATAGCAAAGGTATATCGCCG	CHOHS 473 HUHS 473
GGATAACCCCTGCTATGACCAAGAGGCGGATATCTGAATTCATCACTGTGG GGATAACCCAGCTATGACCAAGAGGCGGATATCTGAATTCATCACTGTGG	CHOHS 523 HUHS 523
ATTTTGACATCGCTGGCCAGTTTGATCCCATGATTCCTGATGCAGAGTCC ATTTTGACATCGCTGGCCAGTTTGATCCCATGATTCCTGATGCAGAGTCC	CHOHS 573 HUHS 573
CTGAAGATCATGTCTGAGATCCTGAGTTCACTTCAGATAGGCAACTTCCT CTGAAGATCATGTCTGAGATCCTGAGTTCACTTCAGATAGGCAACTTCCT	CHOHS 623 HUHS 623
GGTCAAGCTAAATGACAGGCGCATCCTAGATGGAATGTTTGGCTGCTCTGTG GGTCAAGCTAAATGATCGGCGCATCCTAGATGGAATGTTTGGCTGCTCTGTG	CHOHS 673 HUHS 673
GTGTTCTGATAGCAAGTTCCGTACCATCTGCTCCTCAGTGGACAAACTA GTGTTCTGATAGCAAGTTCCGTACCATCTGCTCCTCAGTGGACAAACTA	CHOHS 723 HUHS 723
GATAAGGTTTCCTGGGAGGAAGTAAAGAAAGATGAGTGGTGGGAGAGAAGG GATAAGGTTTCCTGGGAGGAAGTAAAGAAAGATGAGTGGTGGGAGAGAAGG	CHOHS 773 HUHS 773
CCTTGACCCAGAAGTGGCTGATCGCATTGGGGACTATGTCCAGCAACACG CCTTGACCCAGAAGTGGCTGATCGCATTGGGGACTATGTCCAGCAACACG	CHOHS 823 HUHS 823
GTGAGGCTCTCTGCTGTAACACCACTGCTCCAGGATCCTAAACTCTCTCAA GTGAGGCTCTCTGCTGTAACACCACTGCTCCAGGATCCTAAACTCTCTCAA	CHOHS 873 HUHS 873
AACAAACAGGCTGTGGAGGGCTTGGGAGACCTAAAGCTCTCTTTGAATA AACAAACAGGCTGTGGAGGGCTTGGGAGACCTAAAGCTCTCTTTGAATA	CHOHS 923 HUHS 923
TCTAACCTCTTTGGCATTGATGACAAAGATCTCCTTTGACCTGAGCCTTG CTTACCTCTTTGGCATTGATGACAAAGATCTCCTTTGACCTGAGCCTTG	CHOHS 973 HUHS 973

CTCAGGGCTGGACTACTATCTCTGGGTCTCTATCTGGCACTGCTGCTA	CHOHR1023
CTCAGGGCTGGATTACTACTCTGGGTCTCTATCAAGCACTGCTGCTA	HUHR1023
CAGATGCCAATCCGAGCTGGGAAGAACCTGGTGTGGCACTGTGCTG	CHOHR1073
CAGACCCAGCCAGGAGGGGAAGAACCTGGTGTGGCACTGTGCTG	HUHR1073
CTGGAGGGCTATGATGGGCTCTGGGCATGTTGACCCCAAAGCGCA	CHOHR1123
CTGGAGGGCTATGATGGGCTAGTGGGCATGTTGACCCCAAAGCGCA	HUHR1123
AAATGGCATATGCGTGGGCTCAGTATTCAGTACAAGGATCTTCTCCATT	CHOHR1173
AGCTCGCCATGTGTGGGCTCAGCAATTCGGCTGACGGATTTCTCCATC	HUHR1173
CTAAGCAACGATTAAGAGCTTTAGAGGAAAAAGTTCCGGACCCTGAGAG	CHOHR1223
CTGGAAACAGACACTAGAGGCTTTGAGGAAGAAATACGGACCCTGAGAG	HUHR1223
ACAGGTCCTTGTGGCATCTGCACAGAAAAAGGTGGCTGAGGACAGACTA	CHOHR1273
ACAGGTCCTTGTGGCATCTGCACAGAAAAAGGTGGTACAGGAAGACTA	HUHR1273
AAGCTTGTCTCCAGTGTGGGATGCTGGGATCAAGGCTGAGGTGCTCTAC	CHOHR1323
AAGCTTGTCTCAGACTGTGGGATGCTGGGATCAAGGCTGAGCTGCTCTAC	HUHR1323
AAAAAGAACCCAAAGTACTGAACCAGTTCCAGTACTGTGAGGAAACAGG	CHOHR1373
AAAGAAACCCAAAGCTACTGAACCAGTTCCAGTACTGTGAGGAGGCAGG	HUHR1373
CATCCCACTGGTAGCCATCATTCGGTACGAGGAGCTCAAGGATGGTGTCA	CHOHR1423
CATCCCACTGGTGGCTATCATTCGGGAGCAGGAACCTCAAGGATGGGTCA	CHHR1423
TCAAGCTCCGTCACGTGCCTAGCAGAGAAAGAGGTGGATGTCCGAAGAGAA	CHOHR1473
TCAAGCTCCGTCACGTGACGAGCAGGCAAGAGGTGGATGTCCGAAGAGAA	HUHR1473
GACCTTGTGGAGGAGATCAGAAGAAAGAACAAATCAGCCCTCTACGCTTG	CHOHR1523
GACCTTGTGGAGGAAATCAAAGGAGAAACAGCCCTCTCTGACTCTG	HUHR1523
CTGAAGGAAGCAGACTAGCAGAGGAAACGAGGATTCGCCCTTATTTGAACTA	CHOHR1573
CTGAAGCTGAACAACACTATCAGAGGAAAGGAAATGGGACTGGCACTATTTG	HUHR1573
AAACAATACTGCATATGTACTTAAATAAGTTGCACATTTCTGATCCAGCA	CHOHR1623
AGGTTAAAGACAACTGCATATGTACTTCAAATGCTTTGCACTTTTCCGTT	HUHR1623
GAAACTGAAGGGTGTCTCGGTACAGAGGCGCTTTTTTTTCTTTTCTTTT	CHOHR1673
TCAGCCGAAGACCTGAAGAGTGGTCAGAACAGAGCCCTTTGATTTTTATTA	HUHR1673
ATTGTTATTATTCGTAATCACTGGCAAAAAGAAAAAAGAAAAAAGAA	CHOHR1723
TGGTTATTTTATTGATTATTAAGTGGCAAAAACGGCCAGGTTACAA	HUHR1717
AAGGAACAGAAAAAGAAAAAAGGCCAGGTGTACAGTTTCCTAGTCGGAA	CHOHR1773
TAGGGAGGAGCAGTATAGCTCAGTATAGGTCGATTAAACAAGACTGACTC	CHOHR1823
CACCTTTTTATACAAAGCCACGAGGCTTAGTCCAGTCTGTGCTCCTGG	HUHR1767
CTAAACTCCTGTATCAACCTCAGATGGCTTCTTCCAGACCCACATAACAG	CHOHR1873
GCTACAAAGACCCACCTGACATGGTCCCATCTCCAGCCCGGCACAGT	HUHR1827
ATGAATGATCCCTCCCTACAACTGCCCAAAAAGTCCAAATAAATCTG	CHOHR1923
TGCAACCATACCTCCCAACACCAATTGCCAAAAGTCCAAATAAATGCC	HUHR1877
TCCTCCCTTAAAAAATAA	CHOHR1942
TCAAECACGAAAAAATAA	HUHR1896

Figure 4. Nucleotide sequence alignment of (a) hamster and (b) human cDNA for HRS. The nucleotide position number is listed on the right margin. Gaps correspond to computer generated displacement of sequences to maximize alignment. Identical bases at corresponding positions are boxed in.

On comparing the three eukaryotic HRS (yeast, hamster and human) with the prokaryotic HRS (*E.coli*) much less homology was found. However, a significant homology was located in a middle region of all the four species of HRS (Fig.5). Within this region, the hexapeptide arg-glu-leu-asp-tyr-tyr represented the longest contiguous amino acid match among all four species of HRS.

Isolation of HRS recombinants from a CHO genomic library

As a first step in determining the structure of genomic HRS, approximately 1×10^6 phage of a CHO genomic library prepared in the λ phage EMBL3 were screened using the 2590 bp CHO cDNA clone isolated previously, pHRS1, as the hybridization probe (1). Two positive phage clones (HRS λ 1, HRS λ 2) were isolated by repeated plaque purification. One of the positive phage clones was found to hybridize also to a cDNA probe pX75, (350 bp) which coded for the previously described 75K polypeptide of unknown function that is coamplified with HRS in histidinol resistant lines (8). A restriction enzyme map was constructed (Fig. 6) based on restriction endonuclease analysis of the two recombinant phage isolates. The position of pHRS1 and pX75 on the inserted DNA of phage clone HRS λ 1 was assigned by Southern blot analysis of the restriction enzyme digested recombinant phage. In order to span the whole HRS gene, we obtained more overlapping genomic phage clones as follows. An unique sequence containing a Sal I - BamH I fragment of HRS λ 2 was used as a probe to "chromosome walk". An additional overlapping phage clone (HRS λ 3) was obtained in this way and the restriction enzyme map of this recombinant phage was constructed as shown in Fig. 6. In total, three overlapping phage clones were obtained encompassing 29 kb of the genomic DNA.

Mapping of exons and introns of the chromosomal gene for HRS

Exon sequences were tentatively localized in six regions by hybridization of ^{32}P labelled HRS cDNA fragments to enzyme digests of genomic DNA inserts from the phage clones (Fig. 6). The precise locations of each of the 5' and 3' exon-intron boundaries were defined by sequencing the appropriate regions of the cloned genomic DNA and aligning them with the sequence of HRS cDNAs. As shown in Fig. 6, the chromosomal gene for HRS is divided into 13 exons and spans 18 kb. Fig. 7 summarizes the sequences of the

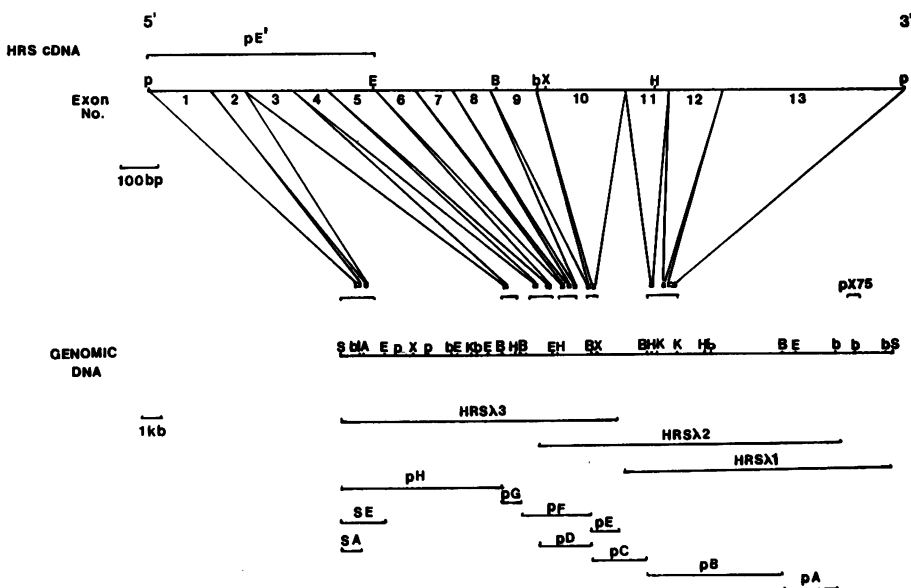


Figure 6. The organization of the hamster HRS gene. A. Restriction endonuclease map of HRS cDNA. The numbers above the map denote exon number. B. Restriction endonuclease map of genomic HRS. The three overlapping λ phage recombinants (HRS λ 1, HRS λ 2, HRS λ 3) that were used to characterize the gene are indicated below the map. The subcloned genomic fragments used for detailed mapping are shown below the recombinants. Enzymes represented are A, AvaI; B, BamHI; bI, BglI; b, BglII; E, EcoRI; H, HindIII; K, KpnI; p, PstI; X, XhoI.

exon-intron boundaries. The splice junctions conform to the consensus sequences described in the literature. The polyadenylation signal was located 20 bp 5' of the site of polyadenylation at the 3' end of the gene.

DISCUSSION

In this paper we have described the isolation and sequencing of full length cDNA clones encoding CHO and human histidyl-tRNA synthetases and have compared these sequences between themselves and with the sequences of HRS derived from yeast and E.coli. The homologies we have observed between the HRS from hamster, human, yeast, together with the expression of the full length human HRS cDNA after transfection into Cos 1 cells and a CHO ts mutant defective in the gene for HRS, provides very convincing evidence

SEQUENCES OF EXON-INTRON JUNCTIONS

		5' Boundary		3' Boundary			
Consensus		C	a	tt	c		
From literature		AG	gt agt	n ag	G		
		A	g	cc	t		
Intron Number						Intron Size (kb)	
1	GCT GAC CAG	gtg	gacta ...	cctccacag	ATC GAG GAC	0.24	
	90				91		
		Gln (30)			Ile (31)		
2	ACC CCC AAG	gta	aactat ...	actctgaag	GCC ACA AGA	7.5	
	180				181		
		Lys (60)			Gly (61)		
3	GAG CTA AAG	gtg	agggt ...	ttttccag	GAA ACA CTC	1.8	
	300				301		
		Lys (100)			Glu (101)		
4	GAC CTC ACT	gtc	accctc ...	tgttatcag	CTC CCT TTT	1.1	
	396				397		
		Asp (132)			Val (133)		
5	T ATC ACT GT	gtg	agtgtt ...	tccttgccag	G GAT TTT GA	1.0	
	521				522		
		Val (174)			Val (174)		
6	CTG GTC AAG	gtc	agaact ...	tcctaaccag	GTA AAT GAC	0.12	
	630				631		
		Lys (210)			Val (211)		
7	CTA GAT AAG	gta	aataac ...	tctgcccag	GTT TCC TGG	0.35	
	729				730		
		Lys (243)			Val (244)		
8	CAC GGT GAG	gta	aagcc ...	gctcccag	GTG TGT CTG	2.9	
	828				829		
		Glu (276)			Val (277)		
9	GAT GAC AAG	gta	agaag ...	tctccatag	ATC TCC TTT	0.1	
	951				952		
		Lys (317)			Ile (318)		
10	CGA TTA GAG	gtc	ccggca ..	gcgtcctag	GCT TTA GAC	0.18	
	1191				1192		
		Glu (397)			Ala (398)		
11	GGC ATC AAG	gtc	accaggag ...	catttgccag	GCC GAG CTC	0.3	
	1398				1399		
		Lys (436)			Ala (437)		
12	AGA GAA GAG	gta	aggagg ...	cctggcccag	CTG GAT GTC...		
	1545				1546		
		Glu (485)			Val (486)		
		GTC CAA TAA ATG TCT CCC CCG TAA caccctgtct					

Figure 7. Organization of exon-intron boundaries of the hamster HRS gene. Exon sequences are in capital letters, intron sequences are in lowercase letters. The number shown below the DNA sequence indicates the nucleotide position at which the intron interrupts the HRS mRNA. The numbers shown in parenthesis indicates the position of the corresponding amino acid residue in the HRS protein.

that we have indeed cloned the hamster and human genes for HRS. However, there remains some controversy about the molecular weight of the mammalian gene product. From the inferred amino acid sequences, the predicted proteins encoded by the CHO and human HRS cDNAs would consist of 508 amino acid residues, and therefore a protein of molecular weight of about 57,000. In contrast, in our earlier work with polymyositis sera known to precipitate HRS (anti-Jo 1 antisera), we consistently immunoprecipitated, in the presence of PMSF, a 52,000 polypeptide with histidyl-tRNA synthetase activity. Disparities in molecular weight of HRS from

mammals have been reported previously. Kane et al. have purified HRS from rabbit reticulocytes as a dimer of identical subunits of 64,000 molecular weight (22). Yang et al. reported multiple polypeptides with molecular weight 52,000 to 64,000 in HRS from rats prepared by using a Bio Rex 70 column (23). The immunoprecipitates of HRS from HeLa cells obtained by using polymyositis anti-Jo 1 antibody, as reported by Mathews and Bernstein (18), contained a 50,000 polypeptide. Molecular weight values for HRS which were significantly smaller than the range of values normally observed have also been reported by Walker et al. (24). It has been generally assumed that the 52,000 polypeptide is probably a proteolytic fragment of HRS. From the predicted amino acid sequences of both CHO and human HRS, the second Met residue lies at the +70 aa residue. If translation were to start from this second Met residue, the predicted molecular weight of this smaller protein would be about 50,000. Mirande et al. (25) have speculated that a number of the eukaryotic synthetases contain a polypeptide domain which can be removed by proteolysis with no loss of activity or subunit structure, resulting in a protein of similar size and subunit structure to the corresponding synthetase in prokaryotes. Resolution of this problem will require further investigation.

Recently, Natsoulis et al (20) have isolated and sequenced the yeast gene for histidyl-tRNA synthetase HTS1. They reported the unusual finding that this HTS1 gene codes not only for the cytoplasmic histidyl-tRNA synthetase (a shorter transcript) but also for the mitochondrial histidyl-tRNA synthetase (a longer transcript). In our case, we believe that the mammalian genes we have cloned represent the genes for hamster and human cytoplasmic HRS since no leader sequences, characteristics of mitochondrial proteins, were found at the 5' end of the coding regions. We do not know whether there are one or two genes that code for mammalian HRS. Further analysis will be necessary to obtain more information on mammalian mitochondrial HRS.

A number of *E. coli* and yeast genes for aminoacyl-tRNA synthetases have now been cloned (2,6). Comparison of aminoacyl-tRNA synthetases (for the same amino acid) between yeast and *E. coli* have revealed information regarding structural organiza-

tion. With both methionyl-tRNA synthetase (26) and glutaminyl-tRNA synthetase(27), an alignment can be made between a central region of the E. coli and yeast proteins. Nothing is known as yet of the functional significance of the conserved regions of these enzyme pairs. Recently, Regan et al (28) identified a common immunological determinant in E. coli and silkworm alanyl-tRNA synthetase. By analysis of ten fragments of the E. coli enzyme the cross-reacting epitope has been mapped to that part of the enzyme which is essential for alanyl adenylate synthesis. In the case of histidyl-tRNA synthetase, on comparing four species of HRS throughout the animal kingdom (ie E. coli, yeast, hamster and human), a significant homology was located in a middle region of all the four species of HRS. We are in the process of testing whether this conserved area represent the core catalytic domain of the enzyme. The availability of the full length cDNAs and the transfection assays will allow us to study the structural and functional relationship of this enzyme.

A number of genes for aminoacyl-tRNA synthetases in yeast have been cloned (20,26,27,29). None of these were shown to be split into exons. From the studies in bacteria concerning structural and functional relationship of the genes for aminoacyl-tRNA synthetase, there is considerable evidence that there are different domains in the structure of these enzymes (1,4,5,30, 31). The information on the intron-exon boundaries of the 13 exons of the chromosomal HRS gene will enable us to address the question of whether these exons correlate to different domains of the HRS polypeptides.

ACKNOWLEDGEMENTS:

This work was supported by the grants from the Medical Research Council and National Cancer Institute of Canada. F.T. was a recipient of an Ontario Graduate Scholarship and an University of Toronto Open Scholarship.

We thank Dr. P. Berg and Dr. H. Okayama for their gift of the cDNA libraries, Dr. L. Thompson for his gift of the cell line D10-5, Dr. B. Williams for his gift of Cos 1 cells, Dr. E. Keystone for his gift of autoimmune sera from polymyositis patients, and Dr. G.R. Fink for providing the sequence of the

yeast histidyl-tRNA synthetase. We also thank Drs. M. Shotwell and P. Ray for their gift of the CHO genomic library.

*To whom correspondence should be addressed. Present address: Rm 1419, North Pavilion, Toronto Western Hospital, 399, Bathurst Street, Toronto, Ontario M5T 2S8, Canada

REFERENCES

1. Bodouelle, H. and Winter, G. (1986) *Nature (Lond.)* 320, 371-373.
2. Blow, D.M., Bhat, T.N., Metcalfe, A., Risler, J.L., Beunie, S., and Zelwer, C. (1983) *J. Mol. Biol.* 171, 571-576.
3. Fersht, A.R., Shi, J.P., Knill-Jones, J., Lowe, D.M., Wilkinson, A.J., Blow, D.M., Brick, O., Carter, P., Waye, M.M.Y., and Winter, G. (1985) *Nature (Lond.)* 314, 235-238.
4. Jasin, M., Regan, L., and Schimmel, P. (1983) *Nature (Lond.)* 306, 441-447.
5. Jasin, M., Regan, L., and Schimmel, P. (1984) *Cell* 36, 1089-1095.
6. Webster, T., Tsai, H., Kula, M., Mackie, G.A., and Schimmel, P. (1984) *Science* 226, 1315-1317.
7. Okayama, H. and Berg, P. (1983) *Mol. Cell. Biol.* 3, 280-289.
8. Tsui, F.W.L., Andrusis, I.L., Murialdo, H., and Siminovitch, L. (1985) 5, 2381-2388.
9. Gubler, U. and Hoffman, B.J. (1983) *Gene* 25, 263-269.
10. Benton, W.D., and Davis, R.W. (1977) *Science* 196, 180-182.
11. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* 74, 5463-5467.
12. Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatus, T., Zinn, K. and Green, M.R. (1984) *Nucl. Acid. Res.* 12, 7035-7056.
13. Graham, F. and Van der Eb, A., (1973) *Virology* 52, 456-457.
14. Laemmli, U.K. (1970) *Nature (Lond.)* 227, 680-685.
15. Thompson, L.H., Lofgren, D.H., and Adair, G.M. (1978) *Somatic Cell Genet.* 4, 423-435.
16. Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J., and Rutler, W.J. (1979) *Biochemistry* 18, 5294-5299.
17. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) *Molecular Cloning, a laboratory manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
18. Mathews, M.B. and Bernstein, R.M. (1983) *Nature (Lond.)* 304, 177-179.
19. Gorman, C. (1985) In *DNA cloning*, Glover, D.M. Ed. vol. II. pp.143-190. IRL press, Washington, D.C.
20. Natsoulis, G., Hilger, F., and Fink, G.R. (1986) *Cell* 46, 235-243.
21. Freedman, R., Gibson, B., Donovan, D., Biemann, K., Eisenbeis, S., Parker, J., and Schimmel, P. (1985) *J. Biol. Chem.* 260, 10063-10068.
22. Kane, S.M., Vugrincic, C., Finbloom, D.S., and Smith, D.W.E. (1978) *Biochemistry* 17, 1509-1514.

23. Yang, D.C.H., Dang, C.V., and Arnett, F.C. (1984) *Biochem. Biophys. Res. Comm.* 120, 15-21.
24. Walker, E.J., Treacy, G.B., and Jeffrey, P.D. (1983) *Biochemistry* 22, 1934-1941.
25. Mirande, M., Cirakoglu, B., and Waller, J.P. (1983) *Eur. J. Biochem.* 131,163-170.
26. Walter, P., Gangloff, J., Bonnet, J., Bodulanger, Y., Ebel, J.-P., and Fasiolo, F. (1983) *Proc. Natl. Acad. Sci. U.S.A.* 80, 2437-2441.
27. Ludmerer, S.L. and Schimmel, P. (1985) *J. Bacteriol.* 163 763-768.
28. Regan, L., Dignam, J.D., and Schimmel, P. (1986) *J. Biol. Chem.* 261, 5241-5244.
29. Pape, L.K., and Tzagoloff A. (1985) *Nucl. Acid Res.* 13, 6171-6183.
30. Barker, D.G., Ebel, J.P., Jakes, R., and Bruton, C.J. (1982) *Eur. J. Biochem.* 127,449-457.
31. Waye, M.M.Y., Winter, G., Wilkinson, A.J., and Fersht, A.R. (1983) *EMBO J.* 2, 1827-1829.