Volume 13 Number 14 1985 Nucleic Acids Research

Structural features of the *his*T operon of *Escherichia coli* K-12

Peggy J.Arps, Christopher C.Marvel[1+], Beth C.Rubin, Dean A.Tolan[1], Edward E.Penhoet[1] and Malcolm E.Winkler*

Department of Molecular Biology, Northwestern University Medical School, Chicago, IL 60611, and [1]Department of Biochemistry, University of California, Berkeley, CA 94720, USA

Received 25 April 1985; Revised 17 June 1985; Accepted 19 June 1985

ABSTRACT

The DNA sequence of a 2.3-kilobase segment of the E. coli hisT operon was determined. Analysis of the sequence indicated that the upstream gene in the operon encodes a 36,364-dalton polypeptide, which runs aberrantly on SDS-polyacrylamide gels. The distal hisT gene encodes the tRNA modification enzyme, pseudouridine synthase I, which was shown to have a polypeptide molecular mass of 30,399 daltons. The DNA sequence was consistent with the phenotypes and hisT expression of mutant operons. Analysis of the sequence and genetic complementation experiments demonstrated that the upstream and hisT genes are evolutionarily, structurally, and functionally unrelated; however, translation signals for the two genes overlap, which is consistent with genetic evidence suggesting translational coupling. Codon usage in the upstream gene is radically different from the hisT gene and may underlie the differential expression observed from the operon. Gene-inactivation experiments and S1-mapping of in vivo transcripts indicated that the operon contains an additional upstream gene. S1-mapping experiments also confirmed the presence of an internal promoter, which might be stringently controlled. Taken together, these results show that the structure of the hisT operon is complex and suggest that the operon might be regulated at several levels.

INTRODUCTION

Maturation of stable RNA molecules occurs by multistep processing and modification of primary transcripts. A considerable number of biochemical and genetic studies have outlined the organization of genes that specify stable RNA molecules and the enzymatic cleavage steps involved in processing (reviewed in 1). In contrast, comparatively little is known about modification, even though this process is a dynamic, integral part of stable RNA biosynthesis in both prokaryotes and eukaryotes. In Escherichia coli, greater than 1% of the chromosome encodes the enzymes that modify tRNA and rRNA molecules (2). With the exception of the work of Björk and his associates on tRNA-methyl transferase genes (summarized in 2), the structure and regulation of these genes are largely unknown. In addition, there is currently inadequate knowledge of the functions played by RNA modifications. Modifications undoubtedly play structural roles in stable RNA molecules, and they may influence the interac-

tions between RNA and protein molecules (reviewed in 2 and 3). Physiological experiments have also suggested that the degree of tRNA modification may modulate translation efficiency in response to environmental stress and developmental stage in both prokaryotic and eukaryotic cells (2,3,4,5).

Recently we reported the cloning and characterization of the hisT gene of E. coli K-12 (6). The hisT gene encodes tRNA pseudouridine synthase I (PSUI), which catalyzes the formation of pseudouridine (Ψ) residues in the anticodon stem and loop of at least 30 different tRNA species in E. coli (7). Because PSUI modifies over half of the cellular tRNA molecules, mutations in the hisT gene have widespread pleiotropic effects on metabolism and regulation (see 6). We established that hisT is part of a differentially-expressed operon that contains at least two genes (6). Subcloning, minicell, and transposon-insertion experiments showed that the hisT structural gene, which encodes the 31,000-dalton PSUI polypeptide, is preceded in the operon by a gene (designated temporarily as "usg") that encodes a polypeptide with an apparent molecular mass of 45,000 daltons. Both of these genes seemed to be transcribed from a promoter which was localized to a 569 base pair (bp) restriction fragment. Minicell experiments showed that the polypeptide encoded by the "usg" gene was synthesized at least eightfold more on a molar basis than the PSUI polypeptide.

In this paper, we present the complete nucleotide sequence of the restriction fragment which contains the "usg" and hisT genes. The nucleotide sequence suggests a number of structural features which might account for the differential gene expression observed previously for the wild-type and mutant hisT operons (6). In addition, S1-mapping and gene-inactivation experiments confirm that the hisT gene is part of an operon in the native E. coli chromosome and suggest that the operon may contain additional genes.

MATERIALS AND METHODS

Materials

Restriction endonucleases and T4 DNA ligase were purchased from New England Biolabs. Nuclease S1 and non-radioactive nucleotide triphosphates were from P-L Biochemicals, Inc. A kanamycin resistance cassette was purchased from Pharmacia, Inc. DNA polymerase I large fragment was from Boehringer-Mannheim Corp. Biochemicals and antibiotics were purchased from Sigma Chemical Co. and P-L Biochemicals, Inc. (α-^{32}P)dCTP (≈800Ci/mmol) was from Amersham Corp. (γ-^{32}P)ATP (≈3000Ci/mmol)was purchased from New England Nuclear Corp.

Bacterial Strains, Phages, and Plasmids

Plasmids Ψ300, pNU47, pNU61, and pNU84 were described previously (6). Strain

W3110 wild-type and P1*clr* phage were obtained from C. Yanofsky. Strain JC7623 (*recB21 recC22 sbcB15*) (8) was obtained from B. Nichols. Messing strains JM83, JM103, and JM105 and phages M13mp8, M13mp9, M13mp18, and M13mp19 were obtained from B. Nichols or purchased from P-L Biochemicals, Inc.

DNA Sequence Determination and Analysis

Cloning into M13 phage vectors was completed essentially as described by Messing *et al.* (9). DNA sequence determinations were performed by a modification of the dideoxynucleotide method of Sanger *et al.* (10) provided to us by B. Nichols (unpublished). Changes to the basic Sanger protocol included: (i) substitution of (α-^{32}P)dCTP for (α-^{32}P)dATP with appropriate changes in the composition of the reaction mixtures; (ii) incubation of elongation and chase reactions at 37°C for 10 min instead of at 25°C for 15 min; and (iii) inclusion of reaction mixtures that contained dITP instead of dGTP.

DNA fragments from sequencing reactions were resolved using three loadings on 80 cm gels containing 5% polyacrylamide and 8M urea as described by Sanger and Coulson (11). Sequences were analyzed using computer programs purchased from the University of Wisconsin Genetics Computer Group.

S1-mapping of Transcripts

Total RNA was isolated from strain W3110 wild-type, which was growing exponentially in LB + cysteine medium at 37°C, by variations of the "hot phenol" (12) or "hot SDS" methods (13). Labeling 5' ends of restriction fragments, DNA strand separations, RNA-DNA hybridizations, and nuclease S1 treatment were performed as described by Scarpulla (14). Hybridization reactions in different experiments contained 50-250 µg of total bacterial RNA or purified tRNA control and 100,000-400,000 cpm (Cerenkov) of 5' end-labeled, single-strand DNA.

Other Methods

Bacterial growth media and conditions were previously described in detail (6). Likewise, DNA manipulations and PSUI enzyme assays were completed exactly as before (6). Insertion of a kanamycin resistance (Km^R) cassette into the bacterial chromosome as diagrammed in Fig. 4 was completed by the method of Winans *et al.* (15). Transductions using P1*clr* phage were completed according to Miller (16).

RESULTS

Nucleotide sequence of "usg" and hisT

Previously we showed that the "*usg*" and *hisT* genes are contained in a 2.3-kilobase (kb) *Hin*dIII-*Cla*I restriction fragment, which had been inserted into

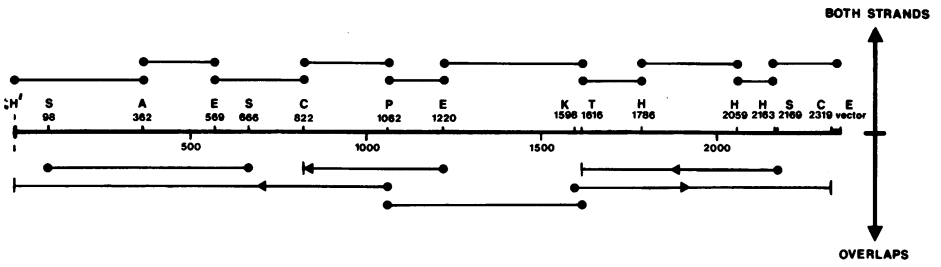


Figure 1. Strategy used to determine the nucleotide sequence of the "usg" and hisT genes. Horizontal lines ending in closed circles indicate clones that were sequenced completely on both strands. Horizontal lines ending in small vertical lines indicate clones sequenced on one strand from which sequences were read to the arrow heads to establish overlaps. Restriction sites are indicated by the following letters: H'=HindIII, S=Sau3A, A=AluI, E=EcoRI, C=ClaI, P=PstI, K=KpnI, T=TaqI, H(internal)=HpaII, E(vector)=EcoRI site from pBR322.

vector pBR322 to form plasmid ¥300 (6). We further showed that the proximal 1.3 kb HindIII-BssHII restriction fragment contains a promoter for the hisT gene and the "usg" gene, while the distal 1.0 kb BssHII-ClaI restriction fragment contains the hisT gene. Figure 1 shows the strategy used to determine the sequence of the HindIII(1)-ClaI(2318) restriction fragment by the dideoxynucleotide method. The entire DNA sequence was determined on both strands and overlap sequences were determined for each restriction site used to clone fragments into the M13 phage vectors. Because the nucleotide composition of this region is somewhat (G+C)-rich (54%), sequencing gel patterns sometimes showed compressions. To circumvent this problem, each clone was sequenced twice: once with regular reaction mixtures containing dGTP and once with reaction mixtures containing dITP instead of dGTP (see Materials and Methods).

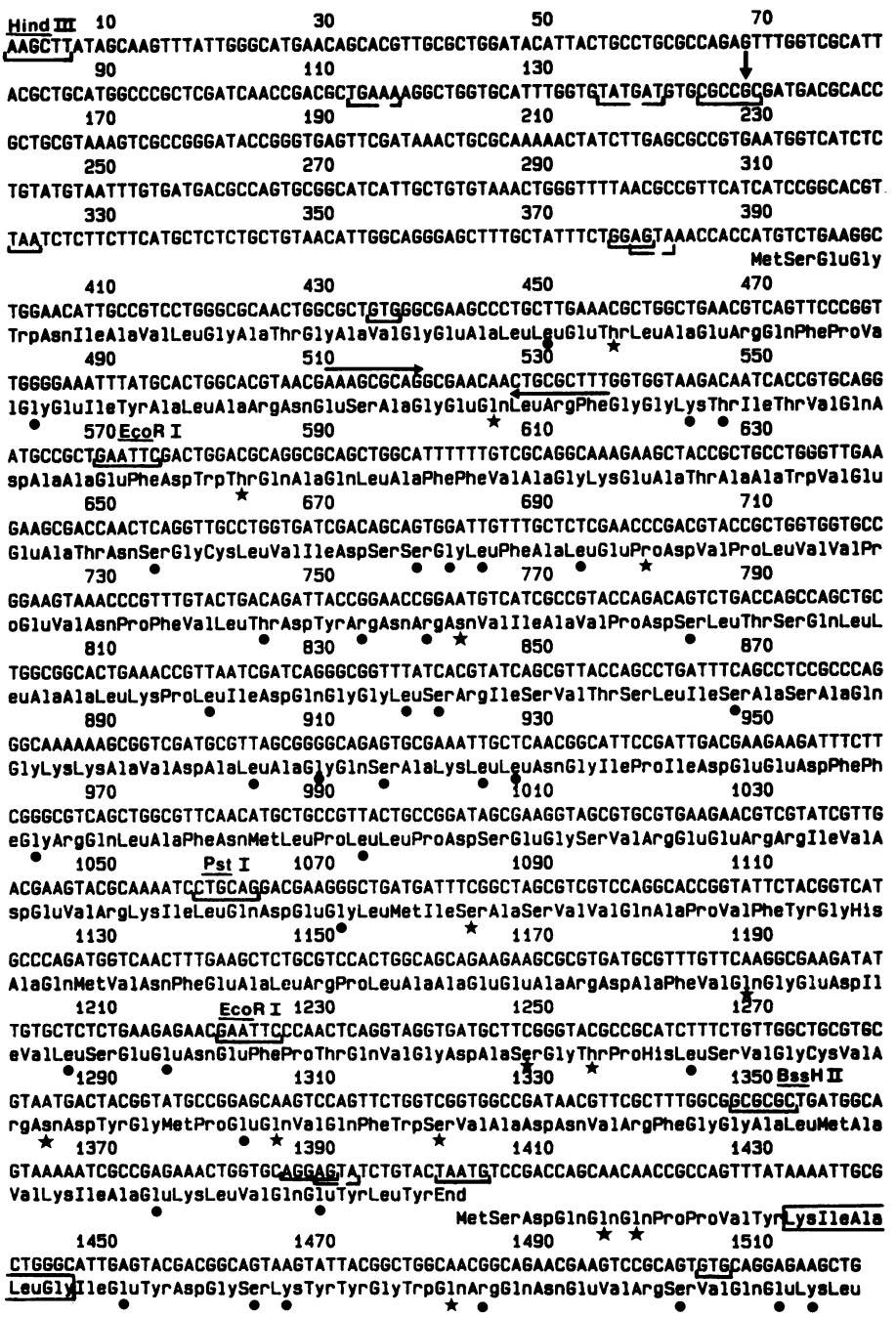
The complete DNA sequence is shown in Figure 2. As indicated in Figure 3A, two long open reading frames extend between nucleotides (nt) 389 and 1402 and nucleotides 1402 and 2214. Based on our analysis of subclones and mutant plasmids (6), these reading frames must correspond to the "usg" and hisT genes, respectively (see below, also). The DNA sequence indicates that the "usg" gene can maximally encode a 36,364-dalton polypeptide containing 337 amino acids. However, minicell experiments demonstrated that the "usg" polypeptide has an apparent molecular mass of 45,000 daltons on SDS-polyacrylamide gels (6). This discrepency shows that the "usg" polypeptide runs aberrantly in this gel system, perhaps because it contains a relatively large number of acidic amino acids (Asp + Glu = 48 vs Arg + Lys = 24), which

are known to increase apparent polypeptide molecular weights on SDS-polyacrylamide gels (17). Presently, we have not determined the amino terminus of the "usg" polypeptide. We favor the start codon indicated in Figure 2 because it minimizes the discrepency between the real and apparent polypeptide molecular weights and because it is preceded by a reasonably good Shine-Delgarno sequence (nt 376, Fig. 2). A second possible start codon at position 434 is not preceded by a Shine-Delgarno consensus sequence.

The translation start at position 1402 for the PSUI polypeptide was empirically determined and is consistent with results from minicell experiments. PSUI protein was isolated to near-homogeneity, and its amino terminus was analyzed in a gas-phase sequenator. For reasons that are not clear, the identity of the first ten amino acids was ambiguous; however, residues 11-15 were clearly determined as Lys-Ile-Ala-Leu-Gly. This pentapeptide exactly matches residues 11-15 predicted from the DNA sequence for a translation start at position 1402 (Fig. 2) and does not occur elsewhere in either the "usg" or hisT coding regions. In addition, a translation start at position 1402 predicts that the PSUI polypeptide will contain 270 amino acids and have a molecular mass of 30,399 daltons, which is close to the apparent molecular mass of 31,000 daltons determined on SDS-polyacrylamide gels (6). In contrast, a translation start at the next most likely initiation codon (1504) would mean that the intercistronic region between "usg" and "hisT" would be somewhat long (>100 nt) and that the predicted molecular weight of the PSUI polypeptide (M_r = 26,364) would not match the apparent molecular weight. Thus, the available protein sequence data, the PSUI polypeptide molecular weight, and the size of the intercistronic region are all consistent with a translation start at position 1402 for the PSUI polypeptide. Having the translation start at this position means that the stop codon for the "usg" polypeptide overlaps the start codon for the PSUI polypeptide (UAAUG; nt 1400, Fig. 2). The AUG start codon for the PSUI polypeptide is still preceded by two possible Shine-Delgarno sequences (nt 1386, Fig. 2); however, the better of the two possible consensus sequences is over ten nucleotides upstream. Possible implications of this gene arrangement are presented in the Discussion.

Sequences of mutant hisT operons

In earlier experiments, the "usg" and hisT genes were located by constructing mutations in plasmid ¥300 (6). From the DNA sequence, we find that all of the internal deletion or insertion mutations that rendered the plasmid hisT$^-$ are located in the PSUI coding region. The sequence also shows that the mutation constructed by removing the EcoRI(569)-EcoRI(1220) fragment should delete the


Hind III
EcoR I
Pst I
EcoR I
Bss H II

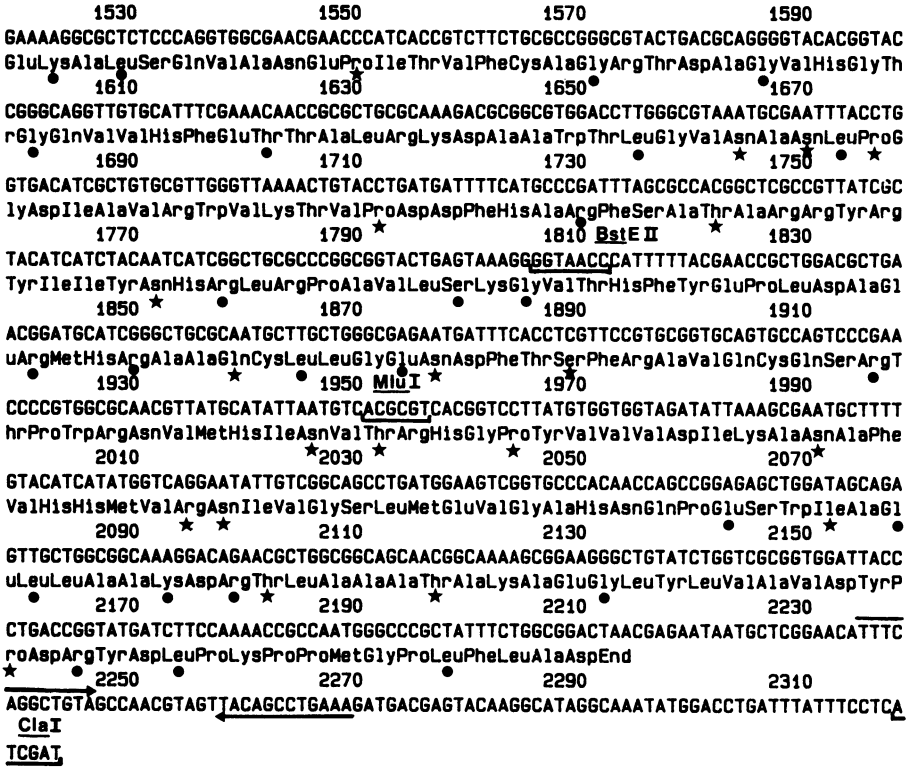


Figure 2. Complete nucleotide sequence of the 2.3 kb-region containing "usg" and hisT. The deduced amino acid sequences for the "usg" and PSUI polypeptides are indicated. Closed circles and stars mark infrequently-used and rare codons, respectively. Other features of the sequence shown in the figure are described in the text.

center of the "usg" gene coding region and create an in-frame translational fusion (Fig. 3B). Our results agreed with this prediction; the mutant plasmid was fully hisT$^+$ and produced a fused, mutant polypeptide which, like the "usg" polypeptide, ran aberrantly on SDS-polyacrylamide gels (predicted M_r = 12,882; apparent M_r = 34,000; 6).

We directly determined the DNA sequence of the mutation that was constructed by removing the BssHII (1347) site from plasmid ¥300 with S1-nuclease (plasmids pNU57 and pNU83, ref. 6). This mutation was a deletion of four base pairs (nt 1348-1351, Fig. 2) and should cause a frame-shift that will result in abrupt translation termination of the "usg" polypeptide (320 amino acids instead of 337; Fig. 3C). Consistent with this prediction, the mutant plasmid produced a truncated "usg" polypeptide in minicells (predicted M_r = 34,399;

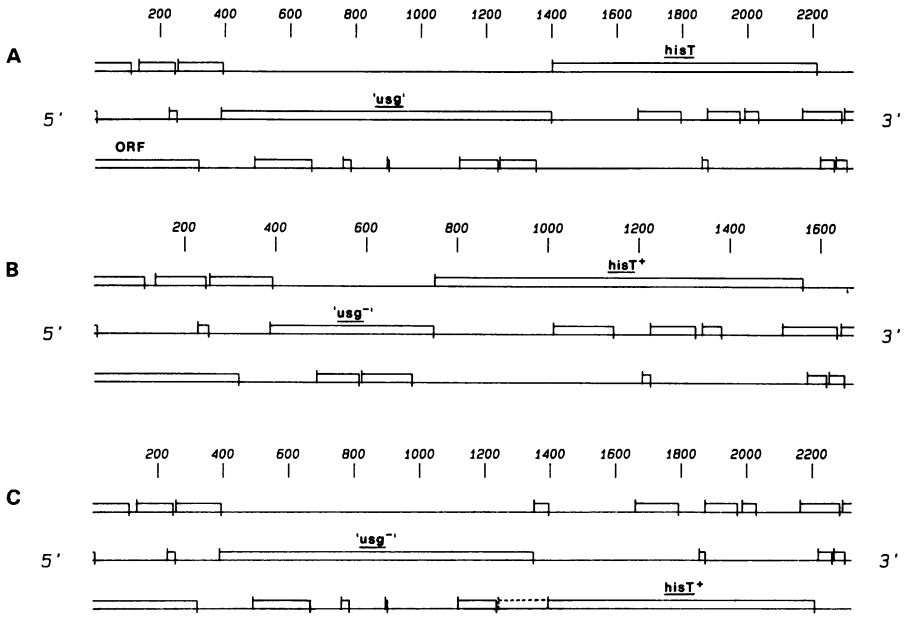


Figure 3. Possible translation reading frames starting with AUG start codons in the wild-type and mutant hisT operons. The mutations and their phenotypes are described in detail in the text. (A) Wild-type operon; "usg" and hisT are indicated. ORF refers to the putative open reading frame for the gene upstream from "usg" in the operon. (B) Mutant operon from plasmid pNU61 in which 675 bp are deleted from the center of the "usg" coding region. (C) Mutant operon from plasmid pNU57 in which 4 bp are deleted from the 3'-terminus of the "usg" coding region. The deletion creates a frame-shift which causes immediate translation termination in the "usg" mRNA (see text). The normal translation start of hisT is indicated by solid lines. An alternate, but unlikely, hisT translation start, which extends into the "usg"- coding region, was created by the deletion and is indicated by broken lines.

apparent M_r = 43,000; 6). In addition, hisT expression from the ΔBssHII mutant plasmid (pNU57) was at least sixfold less than from the wild-type plasmid (¥300; 6), which is the result expected for translational coupling in the synthesis of the "usg" and PSUI polypeptides (see Discussion). Thus, the DNA sequence presented in Figure 2 is completely consistent with the phenotypes and expression of all the mutant hisT operons we constructed.

Relationship between "usg" and hisT

The possibility of translational coupling between the "usg" and hisT genes was somewhat surprising, because the "usg" polypeptide is synthesized to a greater extent than the PSUI polypeptide in minicells (6). Based on the Met content of the two proteins deduced from the DNA sequence (Table 1), we calculate that the

TABLE 1 Comparison of Codon Usage for the hisT Operon with Frequently- and Infrequently-expressed E. coli Genes

		Percentage Use of Codon Synonym								Percentage Use of Codon Synonym					
		NON-REGULATORY[a]	dnaG[a]	uvrC[b]	trmD[c]	"usg"[d]	hisT			NON-REGULATORY[a]	dnaG[a]	uvrC[b]	trmD[c]	"usg"[d]	hisT
Ala	GCU	28	16	17	10	22 (9)	16 (5)	Leu	UUG[I]	8	14	18	18	6 (2)	15 (3)
	GCC	19	19	24	19	24(10)	13 (4)		CUU[I]	9	19	14	11	6 (2)	5 (1)
	GCA	23	24	9	48	27(11)	19 (6)		CUC[I]	7	7	13	4	9 (3)	5 (1)
	GCG	30	40	50	24	27(11)	52(16)		CUA[I]	2	5	5	-	-	5 (1)
Arg	CGU	58	24	48	26	67(10)	25 (5)		CUG	69	37	44	54	69(24)	65(13)
	CGC	35	51	38	32	20 (3)	30 (6)	Lys	AAA	77	60	79	75	89 (8)	64 (7)
	CGA[I]	2	11	5	-	-	10 (2)		AAG[I]	23	40	21	25	11 (1)	36 (4)
	CGG[I]	3	8	14	26	13 (2)	25 (5)	Met	AUG	100	100	100	100	100 (6)	100 (6)
	AGA[I]	1	2	-	11	-	5 (1)	Phe	UUU	44	52	41	50	50 (8)	56 (5)
	AGG[R]	0.25	2	2	5	-	5 (1)		UUC	56	48	59	50	50 (8)	44 (4)
Asn	AAU[R]	24	52	67	50	17 (2)	64 (7)	Pro	CCU[R]	9	24	9	39	-	27 (4)
	AAC	76	48	33	50	83(10)	36 (4)		CCC[R]	6	17	9	8	7 (1)	7 (1)
Asp	GAU	51	51	71	57	56(10)	40 (6)		CCA	20	31	13	23	20 (3)	20 (3)
	GAC	49	49	29	43	44 (8)	60 (9)		CCG	65	27	70	31	73(11)	47 (7)
Cys	UGU	42	71	57	-	-	-	Ser	UCU	27	12	6	-	14 (3)	-
	UGC	58	29	43	100	100 (2)	100 (3)		UCC	26	12	14	21	5 (1)	30 (3)
Gln	CAA[R]	27	40	41	64	17 (3)	33 (4)		UCA[I]	8	12	14	14	14 (3)	-
	CAG	73	60	59	36	83(15)	67 (8)		UCG[R]	11	21	14	21	14 (3)	10 (1)
Glu	GAA	73	61	75	73	87(26)	62 (8)		AGU[I]	6	16	19	14	14 (3)	30 (3)
	GAG[I]	27	39	25	27	13 (4)	38 (5)		AGC	22	25	33	29	38 (8)	30 (3)
Gly	GGU	48	23	30	21	33 (9)	25 (4)	Thr	ACU	24	9	7	17	17 (2)	14 (2)
	GGC	41	48	50	42	48(13)	44 (7)		ACC	51	40	52	33	42 (5)	50 (7)
	GGA[I]	5	16	13	21	4 (1)	-		ACA[I]	6	12	14	-	17 (2)	7 (1)
	GGG[I]	7	11	7	17	15 (4)	31 (5)		ACG[R]	20	38	28	50	25 (3)	29 (4)
His	CAU	39	33	53	67	100 (2)	73 (8)	Trp	UGG	100	100	100	100	100 (4)	100 (5)
	CAC	61	67	47	33	-	27 (3)	Tyr	UAU	41	43	76	33	33 (2)	50 (6)
Ile	AUU	37	36	55	60	47 (7)	50 (5)		UAC	59	57	24	67	67 (4)	50 (6)
	AUC	62	32	42	20	53 (8)	40 (4)	Val	GUU	38	27	26	29	21 (7)	15 (4)
	AUA[R]	1	32	3	20	-	10 (1)		GUC	13	27	26	7	24 (8)	27 (7)
Leu	UUA[I]	6	15	6	14	11 (4)	5 (1)		GUA	23	7	12	14	24 (8)	27 (7)
									GUG	27	38	26	50	30(10)	31 (8)

a,b,c Data from references 18, 19, and 20, respectively. [d] "usg" refers to the upstream gene in the hisT operon The number of codons is in parenthesis. [R] Rarely used codon; [I] infrequently used codon (see 19).

molar excess of the "usg" polypeptide relative to the PSUI polypeptide was actually 10-14 fold. Dot matrix analysis of both the DNA and deduced polypeptide sequences failed to reveal homology between the "usg" and hisT genes; therefore, these two genes seem to be evolutionarily unrelated. Analysis of predicted polypeptide secondary structures and hydrophobicity patterns also failed to reveal any similarities between the "usg" and PSUI polypeptides.

The lack of similarity between the "usg" and hisT genes is further underscored by the data in Table 1 and Figure 2. The codon usage in the "usg" gene closely matches the pattern noted for moderately- to frequently-expressed proteins in E. coli (Table 1; 18). Figure 2 indicates that the 31 infrequently-used and 12 rare codons present in the "usg" gene do not seem to be clustered; in particular, a rare codon adjacent to an infrequently-used codon occurs only twice. In contrast, the codon usage in the hisT gene matches the pattern for infrequently-expressed proteins like the dnaG, uvrC, and trmD gene products (Table 1; 18, 19, 20). In addition, the 33 infrequently-used and 23 rare codons present in the hisT gene are moderately clustered; rare codons appear tandemly or adjacent to infrequently-used codons six times (Fig. 2). The possible implication of this pattern of codon usage to differential gene expression in the hisT operon is presented in the Discussion.

In order to determine whether the "usg" and hisT gene products are functionally related, we needed to construct a chromosomal mutation in the "usg" gene. This was accomplished using a method of insertion mutagenesis devised by Walker and his associates (Fig. 4; 15). The method is based on the observation that linearized plasmid DNA containing cloned inserts is relatively stable in recBC sbc mutants and can recombine with homologous regions of the bacterial chromosome. Using this method, we introduced a kanamycin resistance (Km^R) cassette into the coding region of the "usg" gene (Fig. 4). Following P1 transduction of the Km^R marker into a fresh genetic background (Fig. 4), the resulting strain (NU399) was characterized (Table 2). Strain NU399 lacked PSUI activity (Table 2, line 2); therefore, disruption of the "usg" gene was completely polar on hisT gene expression. This result confirms that "usg" and hisT are part of the same operon in the bacterial chromosome.

Strain NU399 grew with the same doubling time in rich medium as its nearly isogenic parent, strain W3110 wild-type (Table 2). In contrast, NU399 grew 30-60% slower than W3110 wild-type in minimal + glucose or minimal + acetate medium at 30°C, 37°C, or 42°C (data not shown). A plasmid which was "usg"$^-$hisT$^+$ (pNU61, Table 2) restored the growth rate of NU399 to that of W3110 wild-type on minimal medium. Therefore, lack of expression of the hisT gene, but not of the

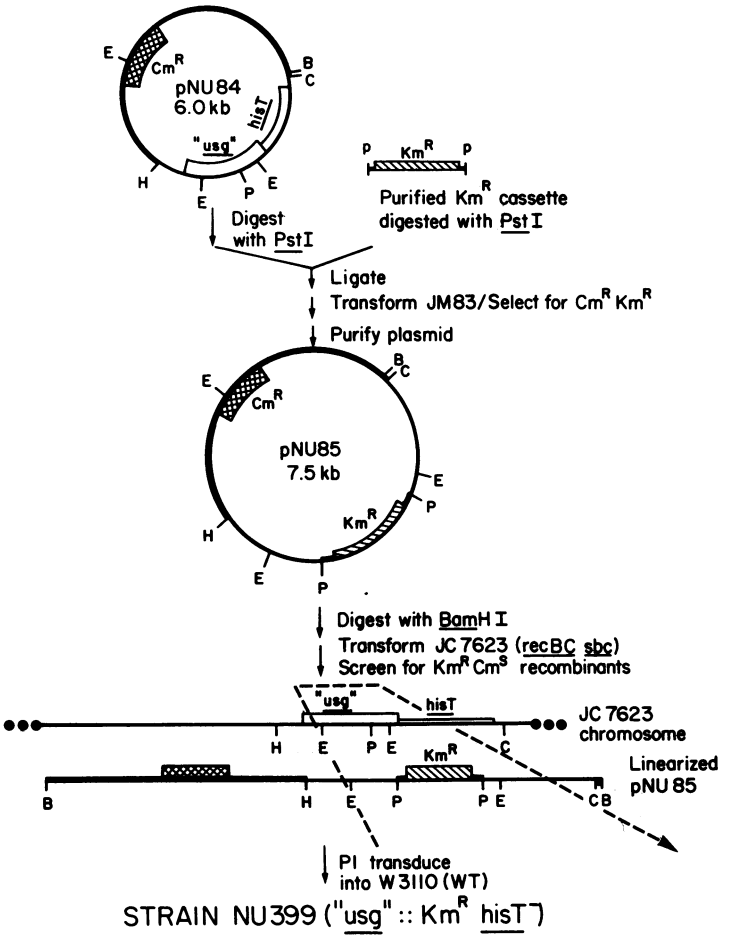


Figure 4. Construction of a mutant containing a Km^R cassette inserted into the "usg" gene in the E. coli chromosome. Details of the construction are presented in the text. Restriction sites are indicated by the same letters used in the legend to Fig. 1.

"usg" gene, is sufficient to reduce the bacterial growth rate. Because NU399 is not an auxotroph, we can conclude that "usg" does not correspond to any of the genes known to map at 50 min in the E. coli chromosome (21).

The results in Table 2 also demonstrate that an intact "usg" polypeptide is not required for PSUI activity in vivo, even though synthesis of the two proteins seems to be coupled. When plasmids that are "usg"$^-$hisT$^+$ were introduced into strain NU399, which is "usg"$^-$hisT$^-$, full hisT$^+$ gene expression was

TABLE 2 *hisT* Gene Expression in Strains Containing a Km^R Cassette Upstream from *hisT* in the Bacterial Chromosome

	STRAIN	DOUBLING TIME (min)[a]	PLASMID GENOTYPE		RELATIVE PSUI SPECIFIC ACTIVITY[b]
			"usg"	*hisT*	
1.	W3110 (wild-type)	31	no plasmid		≡1
2.	NU399 (W3110 *"usg"*:: Km^R)[c]	34	no plasmid		<0.01
3.	NU399/pBR322	36	-	-	<0.01
4.	NU399/Ψ300	36	+	+	18
5.	NU399/pNU61	32	-	+	16
6.	NU399/pNU47	45	-	+	35
7.	NU400 (W3110 *"uusg"*::Km^R)[d]	33	no plasmid		0.5

a Strains were grown in LB + cysteine medium at 37°C with shaking. Starter cultures used to innoculate final cultures were grown in the presence of 50μg/ml of ampicillin and/or kanamycin, where appropriate. Inheritance of plasmid pNU47 was somewhat unstable in the absence of selection, which was detected by the loss of *hisT* gene expression from certain cultures.

b PSUI specific activity was determined in exponentially growing bacteria as described previously (6). The PSUI specific activity was 180 cpm released per min per mg of protein for strain W3110 wild-type. Values in the table are the average of at least two separate determinations. Control experiments established that essentially the same PSUI activities were present in the strains grown under selection for the plasmids. In addition, PSUI activity in strain W3110 (wild-type) containing plasmids Ψ300, pNU61, or pNU47 was the same as shown in the table for strain NU399 containing these same plasmids.

c *"usg"*:: Km^R refers to a Km^R cassette inserted into the bacterial chromosome at the PstI(1062) site (see Fig. 4). The Km^R cassette disrupts the *"usg"* coding region.

d *"uusg"*:: Km^R refers to a Km^R cassette inserted into the bacterial chromosome at the HindIII(1) site.

observed from the plasmids (Table 2, lines 3-6). In strain NU399/pNU61, the positions of the chromosomal insertion and plasmid deletion eliminate the possibility that a functional *"usg"* gene could have arisen by recombination between the plasmid and chromosome. Therefore, disruption of the *"usg"* gene does not affect PSUI function or activity *in vivo*. Protein purification data also supports this conclusion, since PSUI purified to near-homogeneity seems to

consist only of the 31,000-dalton polypeptide (H.O. Kammen and C. Marvel, in preparation). In summary, the "usg" and hisT genes seem to be evolutionarily, structurally, and functionally unrelated, even though they are closely associated in the hisT operon.

Transcripts from the hisT operon

Examination of the DNA sequence shows two perfect dyad symmetries which could form strong secondary structures in hisT operon mRNA (nt 510-535 and 2235-2269, Fig. 2). The possible secondary structure at nucleotide 2235 is immediately downstream from the translation stop codon for the PSUI polypeptide and may play a functional role; however, this possible structure is not followed by the polyU region characteristic of a rho-independent terminator (see below, also). Application of the Zuker fold program at 100 nucleotide intervals in the sequence showed that each segment of the hisT operon transcript could fold into a relatively stable RNA secondary structure (ΔG/structure < -20 kcal); however, extremely stable RNA secondary structures were not predicted around the "usg" and hisT translation start points. This analysis also failed to reveal any tRNA-like structures which could form in this region of the hisT operon transcript.

Previous experiments demonstrated that the HindIII(1)-EcoRI(569) restriction fragment contained a promoter for the "usg" and hisT genes (6); however, it was not possible to tell from these results whether this promoter was primary or internal in the operon. In order to resolve this issue and to examine hisT operon transcription, we completed S1-mapping experiments using 5' end-labeled, single-strand DNA probes isolated from several regions of the operon. When the noncoding strand of the HindIII(1)-EcoRI(569) fragment was used as the probe, no S1-resistant strand was detected (lanes 15-21, Fig. 5). When the coding strand from the HindIII(1)-EcoRI(569) fragment was hybridized to total cellular RNA, two strong bands were detected following treatment with S1-nuclease (lanes 9-11, Fig. 5). The upper band represents full-length protected DNA, suggesting that part of the hisT operon transcription originates upstream from the HindIII (1) site. In some experiments, the single band shown in Fig. 5 appeared as a closely-spaced doublet, which might have been generated by "breathing" of the hybrid in the (A+T)-rich HindIII(1) end. The lower band, which was somewhat less intense than the upper band, most likely represents transcription from the promoter detected in our subcloning and expression vector experiments (6). By running gels considerably longer than shown in Fig. 5, it was possible to determine that the end of the S1-resistant strand was at position 148 ± 1 (arrow, Fig. 2). Sequences upstream from position 148 match the -35 and -10 promoter

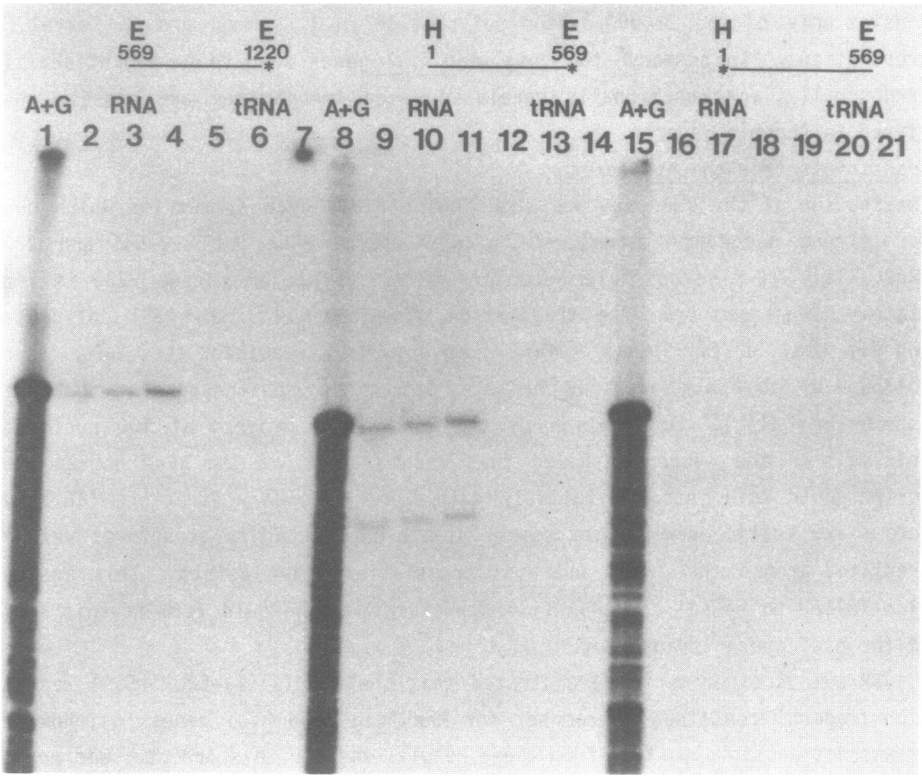


Figure 5. S1-mapping of hisT operon transcripts. Total RNA was isolated by the "hot phenol" method from strain W3110 wild-type growing exponentially in LB + cysteine medium at 37°C. Hybridization mixtures contained the purified, single-strand DNA diagramed above each set of seven lanes and 250 µg of total RNA or tRNA control (see Materials and Methods). Nuclease S1 treatment was carried out on each hybridization mixture for 5, 10, and 15 min before a portion of the reaction mixture was removed and prepared for gel electrophoresis. Lanes 1, 8, and 15 contain standards in which the strands were subjected to the Maxam-Gilbert A+G reaction. Lanes 2-4, 9-11, and 16-18 are from hybridization mixtures that contained total cellular RNA, and lanes 5-7, 12-14, and 19-21 are from hybridization reactions that contained purified control tRNA. Lanes are arranged in the order of increasing time of S1-treatment. Slightly more sample was analyzed in the last lanes of each set. The slight upward curvature of the bands was a property of this particular gel.

consensus sequences reasonably well (Fig. 2), which further supports the notion that this region represents an internal promoter in the hisT operon rather than a transcript processing site. It is perhaps significant that the -4 to +2 region of this putative internal promoter is similar to the discriminator region for stringently-controlled promoters (Fig. 2; 22). Control experiments indicated that no S1-resistant strand was detected when the probe was hybridized to

tRNA (lanes 12-14, Fig. 5). When the single-strand HindIII(1)-EcoRI(569) probe was run directly on a gel, no band corresponding to the lower band in Fig. 5, lanes 9-11 was detected. Finally, the same result was obtained when total bacterial RNA was isolated by two distinct methods (see Materials and Methods). Thus, the S1-mapping experiments strongly support the conclusion that there are at least two promoters for the hisT operon; one upstream from and the other 148 nucleotides downstream from the HindIII(1) site.

Examination of the DNA sequence shows that an open reading frame extends from the HindIII(1) site to position 323 (Fig. 2; line 3, Fig. 3A). This open reading frame could represent another gene in the hisT operon, since the S1-mapping experiments indicated the presence of transcripts that extend from the HindIII(1) site. To test directly the conclusions from the S1-mapping experiments, we inserted a Km^R cassette into the bacterial chromosome at the HindIII(1) site (strain NU400, Table 2). If this region is part of the hisT operon, then the presence of the Km^R cassette might be polar on hisT gene expression. The construction was accomplished by a protocol similar to the one shown in Figure 4 using a plasmid containing DNA upstream from the HindIII(1) site (6). As predicted by the results from the S1-mapping experiments, disruption of the chromosomal HindIII(1) site reduced hisT gene expression (line 7, Table 2). The residual 50% PSUI activity in strain NU400 most likely reflects expression from the internal promoter at position 148. Strain NU400 grows on minimal + glucose medium at 30°C, 37°C, and 42°C; however, the colonies are oddly translucent. Therefore, insertion of the Km^R cassette into the HindIII(1) site seems to have disrupted another upstream gene in the hisT operon.

In a final series of S1-mapping experiments, we found full-length protection of coding strands that were made from the EcoRI(569)-EcoRI(1220), BssHII-(1347)-BstEII(1806), and MluI(1951)-ClaI(2318) restriction fragments (lanes 1-7, Fig. 5, and data not shown). The EcoRI(569)-EcoRI(1220) probe is from an internal region of the "usg" gene coding region and should be fully protected. Likewise, the BssHII(1347)-BstEII(1806) probe spans the "usg" and hisT coding regions. Full length protection of this probe again confirms that these two genes are part of the same operon. Full-length protection of the MluI(1951)-ClaI(2318) probe indicates that the hisT operon transcript extends beyond the possible secondary structure at position 2269, which was noted above. This result indicates that this possible secondary structure does not seem to act as a terminator *in vivo*, and it raises the possibility that the hisT operon may contain additional downstream genes.

DISCUSSION

The results in this paper demonstrate that the hisT gene is part of a complexly structured operon. The hisT gene, which encodes the 30,399-dalton PSUI polypeptide, is located downstream from a gene which encodes a 36,364-dalton polypeptide (Fig. 2; Fig. 3A). Until this upstream gene is identified, we will temporarily refer to it as the "usg" gene. The stop codon for the "usg" polypeptide overlaps the start codon for the PSUI polypeptide, suggesting that synthesis of these two polypeptides may be coupled (Fig. 2, Fig. 3A). Although the "usg" and hisT genes are closely associated in the operon, they seem to be evolutionarily, structurally, and functionally unrelated (Fig. 2, Fig. 4; Table 1; Table 2). In particular, codon usage in the "usg" and hisT genes is radically different; "usg" contains codons found for moderately- to frequently-expressed E. coli proteins, while hisT contains clustered, rare codons found for regulatory and infrequently-expressed proteins (Table 1).

Overlapping translational signals between functionally-related genes have been noted in several bacterial operons. In the trp operon, the stop codon of trpE overlaps the start codon of trpD (23). A similar arrangement of codons exists between trpB and trpA (23). A different type of overlap was found in the pyrB-pyrI operon, in which the stop codon of pyrB overlaps the Shine-Delgarno sequence of pyrI (24). An analogous arrangement was reported between galT and galK in the gal operon (25). In each of these instances, the overlap of translational signals is thought to couple translation and thereby guarantee stoichiometric synthesis of polypeptides that interact to form oligomeric enzymes or that are required in equimolar amounts.

Results from previous genetic experiments support the idea that the overlap between the "usg" and hisT translation signals (Fig. 2; Fig. 3A) also causes translational coupling between the genes (6). A 651-bp deletion, which fuses the amino- and carboxy-terminal portions of the "usg" polypeptide (Fig. 3B), should not disrupt translational coupling (Fig. 3B). Consistent with this expectation, hisT expression is not affected by this deletion in the "usg" gene (Table 2; 6). In contrast, a 4-bp deletion, which causes premature translation termination of the "usg" polypeptide 49 nucleotides upstream from the normal start codon of the PSUI polypeptide (Fig. 3C), greatly reduces hisT expression (6). Since hisT expression is not restored when the mutant plasmid is put into a polarity suppressor strain, it seems unlikely that the 4-bp deletion reduces hisT expression via classical polarity (6). The short distance between the termination codon generated by the 4-bp deletion and the PSUI start codon also argues against polarity. Based on the gene arrangement presented

in this paper, a likely interpretation of this mutant phenotype is that synthesis of the complete "usg" polypeptide is required for maximum translation of the PSUI polypeptide, possibly through the action of a single ribosome, and the 4-bp deletion simply uncouples translation between the genes. Additional experiments are required to rigorously test this model.

From the context provided by the trp, gal, and pyrBI operons, translational coupling in the hisT operon would seem unusual, because the "usg" and hisT genes and polypeptides seem to be unrelated. In addition, previous minicell experiments showed that the "usg" and PSUI polypeptides are not synthesized in equimolar amounts; rather 10-14 times more "usg" polypeptide seems to be produced compared to PSUI polypeptide (6). The codon usage observed in the "usg" and hisT genes (Table 1) may play a role in causing this differential gene expression. It has been proposed that the use of rare codons may limit translation levels (18,26). Genes like dnaG, urvC, and trmD, which are expressed at low levels, have essentially the same codon usage as hisT (Table 1). In addition, dnaG and trmD, like hisT, are members of differentially-expressed operons. The observation that both hisT and trmD, which encodes tRNA (m^1G) methyltransferase (20), are organized into differentially-expressed operons may suggest a common arrangement for genes that encode modification enzymes. Since hisT expression seems to be a function of gene dosage (6), limited translation of hisT would have to reflect an intrinsic property of translating rare codons rather than simple availability of tRNA isoaccepting species (26). Clearly, additional experiments are needed to learn whether codon usage modulates translation levels in the hisT operon.

Gene-inactivation experiments demonstrated that the order and expression deduced previously for the cloned "usg" and hisT genes is the same as in the native E. coli K-12 chromosome (Fig. 4; Table 2). S1-mapping of transcripts and gene-inactivation experiments further revealed the presence of another upstream gene in the hisT operon (Fig. 5; Table 2). An insertion mutation in this gene drastically affected bacterial colony morphology, whereas the same type of insertion mutation in the "usg" gene caused the bacteria to grow more slowly but did not affect colony morphology. Complementation experiments showed that the slower growth rate of strains containing the Km^R insertion in "usg" was caused by a lack of PSUI activity, since transformation of this strain with plasmids that are "usg"-$hisT^+$ returned the growth rate in minimal medium to that of wild-type cells. Complementation experiments also indicated that "usg" and hisT are functionally unrelated. Presently, we do not know the identity of "usg" or the gene upstream from "usg" in the hisT operon; therefore,

we do not yet understand why these two genes are part of the same operon as hisT. The mutants presented in this paper should help answer this question.

The S1-mapping experiments also confirmed the presence of a promoter upstream from "usg" and hisT (Fig. 5). This promoter, which is located in the putative coding region of the gene upstream from "usg" (Fig. 2; Fig. 3), contains a sequence similar to the discriminator sequence found in stringently-controlled promoters (22). An insertion mutation in the gene located upstream from "usg" reduced hisT expression by 50%, which suggests initiations from the internal promoter might account for half of the hisT expression found in bacteria growing in rich medium at 37°C. The nearly equal band intensities observed in the S1-mapping experiments are consistent with this interpretation (lanes 9-11, Fig. 5). Finally, the S1-mapping experiments raised the possibility that the operon may contain additional genes downstream from hisT. Further analysis of transcription and translation should reveal how the complex structural features present in the operon influence regulation and expression of the "usg" and hisT genes.

ACKNOWLEDGEMENTS

We thank our colleagues for the bacterial strains and plasmids listed in the Materials and Methods. We also thank H. Kammen, P. Matsumura, B. Nichols, and R. Scarpulla for advice and helpful discussions. Finally, we thank I. Neal for help with preparation of the manuscript. We thank F. Masiarz of Chiron Corp. for helping us analyze the PSUI amino terminus. This work was supported by NSF grants PCM-8202921 and DMB-8417005 to M.W., grant 82-41 from the Illinois Division of the American Cancer Society to M.W., USPHS grant RR-05370 to Northwestern Medical School and PHS grant CA-21245 to E.P.

+ Present address: Naval Biosciences Laboratory, University of California, Berkeley, CA 94720

*To whom correspondence should be addressed

REFERENCES

1. Apirion, D. (1984) Processing of RNA, CRC Press, Boca Raton, Florida.
2. Björk, G.R. (1984) in Processing of RNA, Apririon, D. Ed., pp. 291-330, CRC Press, Boca Raton, Florida.
3. Kersten, H. (1985) in Progress in Nucleic Acids Research and Molecular Biology, Cohn, W.E. and Moldave, K. Eds., Vol 31, pp. 59-114, Academic Press, New York.
4. Buck, M. and Ames, B.N. (1984) Cell 36, 523-531.
5. Ortwerth, B.J., Lin, V.K., and Wang, R.J. (1984) Nucleic Acids Res. 12, 9009-9023.
6. Marvel, C.C., Arps, P.J., Rubin, B.C., Kammen, H.O., Penhoet, E.E., and Winkler, M.E.(1985) J. Bacteriol. 161, 60-71.

7. Turnbough, C.L., Neill, R.J., Landsberg, R., and Ames, B.N. (1979) J. Biol. Chem. 254, 5111-5119.
8. Oishi, M. and Cosloy, S.D. (1972) Biochem. Biophys. Res. Commun. 49, 1568-1572.
9. Messing, J., Crea, R., and Seeburg, P.H. (1981) Nucleic Acids Res. 9, 309-321.
10. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
11. Sanger, F. and Coulson, A.R. (1978) FEBS Lett. 87, 107-110.
12. Barry, G., Squires, C., and Squires, C.L. (1980) Proc. Natl. Acad. Sci. USA 77, 3331-3335.
13. Winkler, M.E. (1979) J. Bacteriol. 139, 842-849.
14. Scarpulla, R.C. (1984) Mol. and Cell. Biol. 4, 2279-2288.
15. Winans, S.C., Elledge, S.J., Krueger, J.H., and Walker, G.C. (1985) J. Bacteriol., 161,1219-1221.
16. Miller, J.H. (1972) Experiments in Molecular Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
17. Burton, Z., Burgess, R.R., Lin, J., Moore, D., Holder, S., and Gross, C.A. (1981) Nuc. Acids Res. 9, 2889-2903.
18. Konigsberg, W. and Godson, G.N. (1983) Proc. Natl. Acad. Sci. USA 80, 687-691.
19. Sancar, G.B., Sancar, A., and Rupp, W.D. (1984) Nuc. Acids Res. 12, 4593-4608.
20. Byström, A.S., Hjalmarsson, K., J., Wikström, P.M., and Björk, G.R. (1983) EMBO J. 2, 899-905.
21. Bachmann, B.J. (1983) Microbiol. Rev. 47, 180-230.
22. Lamond, A.I. and Travers, A.A. (1985) Cell 40, 319-326.
23. Oppenheim, D.S. and Yanofsky, C. (1980) Genetics 95, 785-795.
24. Pauza, C.D., Karels, M.J., Navre, M., and Schachman, H.K. (1982) Proc. Natl. Acad. Sci. USA 79, 4020-4024.
25. Schümperli, D., McKenney, K., Sobieski, D.A., and Rosenberg, M. (1982) Cell 30, 865-871.
26. Grosjean, H. and Fiers, W. (1982) Gene 18, 199-209.