Codon Pair Utilization Biases Influence Translational Elongation Step Times*

Becky Irwin, J. Denis Heck‡, and G. Wesley Hatfield§

From the Department of Microbiology and Molecular Genetics, College of Medicine, University of California, Irvine, California 92717

Two independent assays capable of measuring the relative in vivo translational step times across a selected codon pair in a growing polypeptide in the bacterium Escherichia coli have been employed to demonstrate that codon pairs observed in protein coding sequences more frequently than predicted (over-represented codon pairs) are translated slower than pairs observed less frequently than expected (under-represented codon pairs). These results are consistent with the findings that translational step times are influenced by codon context and that these context effects are related to the compatabilities of adjacent tRNA isoacceptor molecules on the surface of a translating ribosome. These results also support our previous suggestion that the frequency of one codon next to another has co-evolved with the structure and abundance of tRNA isoacceptors in order to control the rates of translational step times without imposing additional constraints on amino acid sequences or protein structures.

While it is known that translational elongation rates are discontinuous and influenced by codon context, the reasons for variations in translational step times across individual codon pairs are not well understood. The prevailing theory has been that translational rates reflect the correlation between the species-specific usage of a given codon and the abundance of its cognate tRNA (Sorensen et al., 1989). However, more recent experimental results support the idea that translation rates are influenced by the compatabilities of adjacent tRNAs in the A- and P-sites on the surface of translating ribosomes (Smith and Yarus, 1989; Yarus and Curran, 1992). In support of this idea, we previously described an extreme, species-specific, codon pair utilization bias in bacteria, yeast, and mammals (Hatfield and Gutman, 1989; Hatfield and Gutman, 1992). We showed that some codon pairs are used in protein coding sequences much more frequently than expected from the usage of the individual codons of these pairs (over-represented codon pairs), and that some codon pairs are observed much less frequently than expected (under-represented codon pairs). Similar results were obtained by Kolaskar and Reddy (1986) who analyzed codon pair bias in the protein coding sequences of Escherichia coli together with nine coliphages.

For E. coli, our codon pair utilization analysis (Gutman and Hatfield, 1989; Hatfield and Gutman, 1992) was performed on a collection of 237 nonredundant protein coding sequences containing 75,403 codon pairs taken from the GENBANK data base (Release 40.0). The usage frequencies of the 61 nonterminating codons were determined and used to calculate the expected values for the random occurrence of each of the 3721 (61^2) codon pairs in these sequences. The actual occurrence of each codon pair in the data set was tabulated, and these expected and observed values were used to calculate a \( \chi^2 \) value for each codon pair (CHISQ1). These values represent the degree of bias of codon pair usage, and we arbitrarily identify \( \chi^2 \) values associated with under-represented codon pairs as negative numbers. A second set of expected values, which removes the component of codon pair bias associated with the small bias for amino acid nearest neighbors in E. coli (Gutman and Hatfield, 1989) was calculated and used to generate a second set of \( \chi^2 \) (CHISQ2). A third set of expected values, further corrected for the well known III-I dinucleotide bias in protein coding sequences (Fluck et al., 1977; Bossi and Roth, 1980; Colby et al., 1976; Fienstein and Altman, 1977), was calculated, and yet another set of \( \chi^2 \) values (CHISQ3) was generated. Thus, the bias represented by CHISQ3 (used in this report) cannot be the consequence of the bias in amino acid nearest neighbors, the bias of adjacent nucleotides between codons, or the bias of codon usage (since the actual codon frequencies were used to calculate the expected values). These CHISQ3 values range from 125.7 for the most over-represented GAA CUG, Glu-Leu, codon pair to –52.5 for the most under-represented CUG CAG, Leu-Gln, codon pair.

We also previously demonstrated that codon pair biases are directional (the bias associated with codon pair A-B is independent of the bias associated with codon pair B-A) and restricted to adjacent codons, and that genes expressed at high levels tend to avoid over-represented codon pairs (in addition to their well known avoidance of infrequently used codons (Sharp and Li, 1986; Ikemura, 1992)). These observations suggested that at least a portion of codon pair bias is related to the translation process. This conclusion is consistent with the hypothesis that codon pair bias is correlated with translational step times which, in turn, might be related to the compatibility of adjacent tRNA molecules on the surface of a translating ribosome. It is, therefore, possible that the use of one codon next to the other may have co-evolved with the structure and abundance of tRNA isoacceptors to control the rates of translational step times without imposing constraints on amino acid sequences or protein structure (Hatfield and Gutman, 1992).

In this report, we describe two independent assays capable of measuring the relative in vivo translational step times of specific codon pairs in a growing polypeptide chain. We have used each of these assays to demonstrate that over-represented...
codon pairs are translated more slowly than under-represented codon pairs. One assay is based on the observation that a ribosome pausing at a site near the beginning of an mRNA coding sequence can inhibit translation initiation by physically interfering with the attachment of a new ribosome to the message (Liljenstrom and von Heijne, 1987; Bergmann and Lodish, 1979). The other assay is based on the fact that the transit time of a ribosome through the leader polypeptide coding region of the leader RNA of the trp operon sets the basal level of transcription through the trp attenuator (Landick and Yanofsky, 1987).

**EXPERIMENTAL PROCEDURES**

Materials—Restriction endonucleases, T4 DNA ligase, and T4 polynucleotide kinase were purchased from New England Biolabs. ONPG, acetyl coenzyme A, and DTNB were purchased from Sigma. The DNA probe used for the Southern analyses was radiolabeled using a nick translation kit purchased from Amersham Corp. DNA sequencing was performed using an Amersham Sequenase kit. DNA oligonucleotides were synthesized on an Applied Biosystems PCR Mate DNA synthesizer using phosphoramidite chemistry according to the manufacturer’s instructions.

**Plasmid Constructions**—A translational fusion plasmid containing the trc promoter upstream of the lacZ translational initiation sequence, unique NcoI and BamHI restriction endonuclease sites positioned at the first and ninth codons of a 3′-truncated lacZ structural gene, and a complete copy of the lacI′ repressor gene was constructed as follows. A synthetic double-stranded DNA oligonucleotide (34 bp) with Nhel and SalI sticky ends and containing the trpA transcriptional terminator was inserted between the Nhel and SalI sites of a PstI-pBR322 plasmid (pBR322 bp positions 229 and 651, respectively) to yield pBR-T-3. A 1618-bp EcoRI-XmnI DNA fragment containing the trc promoter and the lacZ coding sequences (3′-truncated at the M1uI site) was excised from a pKK232–2 based plasmid (pED512, this laboratory) and inserted into the 3′-3′ EcoRI sites of a pBR322 DNA fragment obtained from pJ118EH (Furste et al., 1986) and inserted into the filled-in EcoRI site of pDH-4 to yield the final construct pDH-3.

A plasmid containing the trpLap (Landick et al., 1990) leader-attenuator region of the E. coli tryptophan operon transcriptionally fused to a 3′-truncated lacZ gene with unique PstI and EcoRI restriction endonuclease sites in the trp leader polypeptide coding sequence was constructed as follows. Two deletions were made in the Simons transcriptional fusion vector pRS551 (Simons et al., 1987). A 3853-bp BclI restriction endonuclease fragment containing the 3′ portion of the lacZ gene and a 520-bp XhoI-HindIII endonuclease restriction fragment from the kanamycin resistance gene were removed to form plasmid pHX1. The unique PstI site in the ampicillin resistance gene of pHX1 was replaced by the Bsal-Scal restriction endonuclease fragment of this plasmid with the analogous Bsal-Scal fragment from pUC19. The unique EcoRI site in pHX1 was eliminated by digestion of the plasmid with EcoRI and end-filling with the Klenow fragment of DNA polymerase I and self-ligation. A 490-bp Sau3A restriction endonuclease fragment containing the trpLap leader-attenuator region was isolated from the plasmid pRL410 (Landick et al., 1990) and ligated into the unique BamHI site of the PstI- and EcoRI-plasmid pHX1 to yield the plasmid pH8-1. Plasmid constructions were verified by DNA sequencing.

**Bacterial Strains and Growth Conditions**—The E. coli K12 strains used in the translational initiation assays were created by homologous recombination of plasmids pDH78, pDH35, pDH12, pDH53, and pDH1718 into the chromosomal copy of the lacZ gene of strain N03434 (Cole et al., 1987) as described previously (Guterson and Koshland, 1983; Wold and Hatfield, 1986) to yield strains IH78, IH35, IH12, IH53, and IH1718, respectively. The implicit absence of an encoded promoter and lacZ translation initiation regions in front of the functional chromosomal copy of the lac operon containing the downstream permease and thiogalactoside transacetylase genes, and places the chromosomally encoded lac promoter in front of the truncated nonfunctional lacZ gene. Bacteriophage P1-mediated transduction was used to transfer the metE::Tn10 locus of strain CAG18491 (obtained from C. Gross) into strain CY15078 (W3110, tnaA2, ΔtrpEA2, trpR; obtained from C. Yanofsky). A tetracycline-resistant methionine auxotrophic transductant was isolated and designated strain IH15078. The polA mutation from strain N03434, which is tightly linked to the metE locus, was transferred into strain IH15078 by selection for methionine prototrophy to produce strain IH910 (W3110, tnaA2, ΔtrpEA2, trpR, polA). The E. coli K12 strains used for the bacterial attenuation assays were created by the homologous recombination of the plasmids pBH121, pBH278, pBH256, pBH121, pBH1279, pBH1257, pBH1869, and pBH18610 into the chromosomal copy of the lacZ gene of strain IH910 to yield the strains IH211, IH278, IH256, IH212, IH279, IH257, IH869, and IH8610, respectively.

All cultures were grown on LB agar or in Luria broth (Miller, 1972). Ampicillin was added to the medium at a final concentration of 100 μg/ml for the growth of strains containing plasmids and 50 μg/ml for the growth of strains containing plasmids integrated into the bacterial chromosome.

**β-Galactosidase and Transacetylase Assays**—Cells were grown in logarithmic phase to a culture density of 0.5–0.7 OD{sub}600 harvested by centrifugation, washed, and resuspended in one-tenth the original culture volume in 50 mM Tris-HCl (pH 7.6) and 10 mM EDTA. Cell extracts were prepared by sonication of the resuspended cells with four 10-s bursts using a Tekmar Sonic Disrupter, model TM 2508, at a setting of 4.5. The sonicated extracts were clarified by centrifugation, and β-galactosidase activity was assayed by measuring the rate of ONPG hydrolysis according to the method of Miller (1972). β-Galactosidase activities were measured at four time points and two extract concentrations under conditions where the assay was linear with respect to time and enzyme concentration. Rates of ONPG formation were determined by a linear regression analysis of an ONPG versus time plot. DTNB was used to measure the thiogalactoside transacetylase-catalyzed formation of free coenzyme A liberated by the transfer of the acetyl group of acetyl-coA to the thiogalactoside acceptor IPTG according to the method of Miller (1972). Rates of coA formation were determined by a linear regression analysis of a coA versus time plot. Again, all rates were determined at, at least, two extract concentrations to ascertain that coA formation was measured in a linear range of enzyme activity. Protein concentrations were determined by the method of Bradford (1976).

**RESULTS**

Effect of Codon Pair Bias on Translation Initiation of lacZ mRNA—A ribosome pausing at a site near the beginning of an mRNA coding sequence can inhibit translation initiation by physically interfering with the attachment of a new ribosome to the message (Liljenstrom and von Heijne, 1987; Bergmann and Lodish, 1979). Therefore, to determine the translational efficiencies of selected codon pairs we constructed a trc-lacZ transcriptional fusion plasmid containing the trc promoter and the lacZ translation initiation regions with unique NcoI and BamHI sites in the beginning of the lacZ coding region of the lac operon (Fig. 1A). Codon pair substitutions into the beginning of the lacZ gene were facilitated by the insertion of synthetic, double-stranded DNA oligonucleotides containing selected codon pairs into these unique restriction sites. The placement of a slowly translated codon pair in the beginning of the lacZ coding region is expected to inhibit the translation of the resultant lacZ mRNA. The effects of over- and under-represented codon pairs on the translational initiation of the lacZ gene were measured relative to a background of randomly utilized (non-biased) codon pairs. To generate this background sequence, both strands of a double-stranded DNA oligonucleotide encoding eight randomly utilized codon pairs and appropriate restriction site ends were synthesized, annealed, and ligated into the unique NcoI and BamHI sites at the first and ninth codons of the lacZ coding sequence in pDH78 containing in E. coli strain IH78 (Fig. 1A and Table I).

When the slowly under-represented Ala-Leu codon pair (GCC CUU, CHISQ3 = −5.7) at positions 3 and 4 of this lacZ mRNA sequence was changed to the more highly under-represented Thr-Leu (ACC CUG, CHISQ3 = −27.3) codon pair, the
Codon Pair Bias and Translational Elongation

**A**

The **arc** promoter is labeled as follows:

\[
\begin{align*}
\text{arc promoter} & \quad \text{lac} \\
TGACATTATGATCGCTGATTGTAATGATGCGATACACATACACACG & \quad \text{lac} \\
\end{align*}
\]

The locations of the **lac** repressor binding site are identified and labeled **lac**.

**B**

![Diagram of codon pairs and translation initiation](https://example.com/diagram.png)

**Fig. 1.** A, the DNA sequence of the ptr: **lac**Z transcription and translation initiation region. The -35 and the -10 RNA polymerase recognition sequences of the **trc** promoter are underlined. The transcriptional start site of the **lac** mRNA is identified as +1. The palindromic sequences of the **lac** repressor binding site are identified and labeled **lac**. The Shine-Dalgarno region of the **lac**Z gene, important for ribosome attachment and translation initiation, is underlined and labeled **lac** S.D. The first 10 codons of the **lac**Z gene and the first 10 amino acids of β-galactosidase are shown. The locations of the Ncol and BamHI restriction endonuclease sites used to replace the first nine codons of the **lac**Z coding region with synthetic double-stranded DNA oligonucleotides are shown. Insertion of the oligonucleotides destroys the Ncol site by changing the 3' G to an A as shown. B, the **trp** leader polypeptide coding region. The DNA sequence of the region of the wild-type **trp** leader containing the 14 amino acid **trp** leader polypeptide coding sequence and the dyad symmetry encoding the stem-loop 1:2 structure (denoted by the arrows) of the leader RNA is shown in the upper sequence. The DNA sequence of the region of the **trp**Lap leader containing the 15 amino acid leader polypeptide coding sequence of the **trp**Lap leader with PstI and EcoRI sites, used to replace codons 8 through 15 with codons encoded in double-stranded DNA oligonucleotides, is shown in the lower sequence. The three bases encoding a glutamine codon, CAG, that were inserted into the leader to create the **highly over-represented** Thr-Leu (ACG CUG, **CHISQ3**: 27.3) codon pair, the translational activity of the **trp** leader is decreased nearly 10-fold (Table I, compare strains **IH35** and **IH12**). The fact that the amino acid sequence of β-galactosidase produced by the **lac** Z mRNA sequences in strains **IH35** and **IH12** is unchanged is the result of different translation initiation rates.

When a single nucleotide change was made in the highly under-represented Thr-Val (ACC CUG, **CHISQ3** = −27.3) codon pair to create the highly over-represented Thr-Leu (ACG CUG, **CHISQ3** = 78.9) codon pair, the translational activity of the **lac**Z gene decreased nearly 10-fold (Table I, compare strain **IH35** and **IH12**). The fact that the amino acid sequence of β-galactosidase produced by the **lac** Z mRNA sequences in strains **IH35** and **IH12** is unaltered argues against any intrinsic differences in β-galactosidase activities in this experiment. Also, the observation that the level of **lac** expression remains the same in these strains argues against any changes in meso-

| Strain | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Specific activity* |
|-------|---|---|---|---|---|---|---|---|---|------------------|
| IH78  | Met | Thr | Ala | Leu | Leu | Arg | Val | Asp | Pro | \(4120 \pm 989^b\) |
| I12   | AUG | ACA | GCC | CUU | CUA | CGA | GUG | GUA | CCC | 290 ± 39 |
| IH35  | AUG | ACA | ACC | CUG | CUA | CGA | GUG | GUA | CCC | 9800 ± 599 |
| IH12  | AUG | ACA | ACC | CUG | CUA | CGA | GUG | GUA | CCC | 1030 ± 115 |
| IH53  | AUG | ACA | CUG | ACC | CUA | CGA | GUG | GUA | CCC | 3900 ± 969 |
| IH1718| AUG | ACA | GCC | CUG | AUA | CGA | GUG | GUA | CCC | 2040 ± 65.5 |

* nmol/min/mg protein.

**Table I**

Effect of codon pair bias on translation initiation of β-galactosidase mRNA

steady state rate of β-galactosidase synthesis increased 2-fold (Table I, compare strains IH78 and IH35). The further observation that the expression of the downstream lacA gene of the polyhistronic lacZYA operon, which encodes a thiogalactoside transacetylase, remained the same in both constructs suggested that these nucleotide changes did not affect the transcriptional initiation rate or the stability of the **lac** mRNA. In fact, this is true for all of the codon pair substitutions reported in Table I. It should also be noted that none of the altered codon pairs described in Table I significantly alter the bias of the flanking codon pairs at positions two and three or four and five. These results suggest, therefore, that the highly under-represented codon pair at positions three and four is translated faster than the moderately under-represented pair at the same position and that the difference in the steady state levels of β-galactosidase produced from mRNAs containing these codon pairs is the result of different translation initiation rates.
sage stability or ρ-induced transcription termination.

Since our previous statistical analyses (Gutman and Hatfield, 1989; Hatfield and Gutman, 1992) showed that there is no correlation between the directionality of a codon pair and its bias, i.e. codon pair A-B versus B-A, we also wished to determine if there is a lack of correlation between the translational efficiencies of a codon pair in a forward and a reverse orientation. If under-represented codon pairs are translated faster than over-represented codon pairs and codon pair usage (codon context) is related to translational efficiency, as suggested by the above results, then the under-represented ACC CUG codon pair might be expected to be translated faster than the randomly utilized CUG ACC codon pair. The data in Table I show that this is, indeed, the case (compare strains IH35 and IH33). In fact, the mRNA sequence in strain IH35 is translated at about the same efficiency as the mRNA sequence in strain IH78 which is also composed of randomly used codon pairs (Table I). Furthermore, the fact that both ACC and CUG are frequently used codons suggests that the translational efficiency of these codons is not simply related to the frequency of usage of the individual codons.

Since the above results suggested that a highly over-represented codon pair is translated more slowly than an under-represented pair, we sought to extend these observations by determining how the translational efficiency of a modestly over-represented codon pair compares to the translational efficiencies of more highly over- and under-represented codon pairs. The data in Table I show that the modestly over-represented Ala-Leu codon pair (GCC CUG, CHISQ3 = 12.3) in strain IH1718 is translated only half as efficiently as the slightly under-represented Ala-Leu codon pair (GCC CUU, CHISQ3 = -5.4) in strain IH78 and 5-fold less efficiency than the highly under-represented Thr-Leu codon pair (ACC CUG, CHISQ3 = -27.3) in strain IH35. However, the modestly over-represented codon pair in strain IH1718 is not translated as slowly as the highly over-represented codon pair in strain IH12. Thus, these results suggest a relationship between the degree of codon pair utilization bias and translational efficiency.

All of the codon pair substitutions reported in Table I are at codon positions three and four of the lacZ coding sequence. Since it was possible that this region of the message might be important for translational initiation in a manner unrelated to codon context (Gold and Stormo, 1987), we examined the translational efficiency of the over- and under-represented codon pairs shown in Table I placed farther downstream at codon pair positions six and seven. In these cases, we observed the same results as shown in Table I (data not shown). Thus, there is no significant positional effect on the placement of these codon pairs early in the coding sequence of this gene. This suggests that the sequence changes we have made do not influence the translational initiation mechanism in a trivial way such as by facilitating base pairing with upstream sequences and interfering with the attachment of a ribosome to the lacZ message.

Effect of Codon Pair Bias on Transcription through the trp Attenuator—The data obtained from the translation initiation experiments suggest a relationship between codon pair bias and translational efficiency. In order to confirm this relationship, we used an independent in vivo assay also capable of determining the relative step times across selected codon pairs in a growing polypeptide chain. We placed the same over- and under-represented codon pairs examined in the translation initiation assay into the trpLep leader polypeptide coding region shown in Fig. 1B. Rapid translation of an under-represented codon pair through the trpLep leader polypeptide coding region and ribosome release at the stop codon results in the frequent formation of the stem-loop 1:2 structure in the leader RNA and a low level of transcription through the downstream attenuator region into the structural genes of the lac operon and, therefore, low β-galactosidase expression. On the other hand, a very slowly translated codon pair that stalls a ribosome at a position where the stem-loop 1:2 region of the trp leader RNA is disrupted for a longer period of time will cause increased transcription through the attenuator into the structural genes of the lac operon and increased β-galactosidase expression (Landick and Yanofsky, 1987). Thus, codon pair substitutions were placed at codon positions nine and ten immediately preceding the tandem tryptophan codons located at the beginning of the stem 1 region of the trp leader RNA (Fig. 1B). Nucleotide changes in codons nine and ten do not affect the base pairings in the stem-loop 1:2 region (Kolter and Yanofsky, 1983).

The data in Table II show that, as predicted, the replacement of the non-biased Lys-Gly codon pair AAA GGU (CHISQ3 = -0.8) with the under-represented (rapidly translated) Thr-Leu codon pair, ACC CUG (CHISQ3 = -27.3), does not significantly affect the basal level of transcription through the trp attenuator (compare strains IH211 and IH278). However, the replacement of this same codon pair with the highly over-represented (slowly translated) Thr-Leu codon pair, ACG CUG (CHISQ3 = 78.9), results in a 2-fold increase in transcription through the trp attenuator (Table II; compare strains IH211 and IH256).

This level of deattenuation is less than expected (5-6-fold) if transcription into the lacZ gene were fully deattenuated, which might have been expected with a codon pair that severely inhibits translation initiation (Table I). One explanation for this low level of transcription through the attenuator might be that the insertion of the over-represented codon pair into the leader alters the secondary structure of the leader RNA. This is unlikely, however, since the data in Table II show that comparable basal levels of transcription through the trp attenuator are observed when the translation initiation codon of the leader polypeptide coding region in strains IH211, IH278, and IH256 are changed from AUG to AUU (strains IH212, IH279, and IH257, respectively). While these codon changes abolish translation of the leader polypeptide coding region, the superattenuated basal levels of transcription through the attenuator of these three strains measured by the production of β-galactosidase are the same. Thus, the 2-fold increase in transcription into the lac structural genes observed in strain IH256 must be due to translation through the leader polypeptide coding region of the trp leader RNA and not due to an alteration of the intrinsic secondary structures. Also, the fact that the β-galactosidase to transacetylase activity ratios vary less than 2-fold for all of the strains shown in Table II suggests that the nucleotide changes we have introduced into codons nine and ten of the trp leader RNA do not significantly affect the translational initiation of the lacZ gene.

Another explanation for the low level of deattenuation observed with the highly over-represented ACG CUG codon pair might be that the stalling of a ribosome on the over-represented codon pair only partially disrupts the base pairings in the stem 1:2 region. This possibility was tested by pausing a ribosome at this same position by an independent mechanism. In this case, we placed a UGA translational stop codon at either base pair position nine in strain IH869 or ten in strain IH8610 of the trpLep leader polypeptide coding sequence (Table II). It has been demonstrated previously that the substitution of a Trp codon with a stop codon in the trp leader causes full deattenuation due to the slow release of the ribosome from the leader RNA (Landick, 1990). The data in Table II show that the
stalling of a ribosome at codons eight and nine with codon eight in the P-site and the stop codon at position nine in the A-site does not cause deattenuation. However, ribosome stalling at codons nine and ten in strain IH10 causes a 2-fold deattenuation, the same level observed when a ribosome is stalled at the over-represented codon pair at positions nine and ten in strain IH256. Therefore, the stalling of a ribosome at the codon pair immediately preceding the tandem Trp codon pair does, indeed, lead to only a partial deattenuation.

In summary, our interpretation of the results obtained with the attenuation assay are the same as our interpretations of the translation initiation assay; that is, the over-represented Thr-Leu codon pair ACG CUG (CHISQ3 = 78.9) is translated slower than the under-represented Thr-Leu codon pair ACC CUG (CHISQ3 = −27.3).

**DISCUSSION**

The data reported here suggest a correlation between the biased use of an individual codon pair and the translational efficiency of that pair. We have demonstrated that an over-represented codon pair is translated slower than an under-represented codon pair, and that the more over-represented a codon pair is, the slower it is translated. These effects of codon pair bias on translation are consistent with the facts that codon pair biases in E. coli are directional and limited to nearest neighbors (there is very little correlation between the $\chi^2$ values of any given codon pair and its reverse counterpart, and more than 95% of the codon pair utilization bias is removed when codon pairs separated by two or three intervening codons are examined (Gutman and Hatfield, 1989; Hatfield and Gutman, 1992)). For example, the non-biased Leu-Thr codon pair (CUG ACC, CHISQ3 = 0.0) is translated more than two times slower than the highly under-represented Thr-Leu codon pair (ACC CUG, CHISQ3 = −27.3; Table I). In addition to supporting the correlation between codon pair bias and translational efficiency, this observation also shows that translational efficiency is more closely related to codon context (codon pair bias) than it is to the utilization frequency of individual codons. This is because both ACC and CUG are frequently used codons in E. coli, but the ACC CUG and CUG ACC pairs are translated at markedly different rates in a context where the biases of the flanking codon pairs are not significantly altered. If the differing translation rates were due primarily to the frequency of usage of one or the other of these codons then the translation rates in both orientations would be expected to be the same.

The data presented in Table I also show that two codon pairs with nearly equal codon pair bias values, but encoding different amino acid pairs and differing at all six nucleotide positions, can exhibit the same translational efficiency (compare strains IH78 and IH53). This observation suggests a close relationship between codon pair $\chi^2$ values and translational efficiency. However, this close relationship might not be observed in every case. For example, it is known that identical codons located at different positions in an mRNA can be read by different isoacceptor tRNAs (Holmes et al., 1977; Goldman et al., 1979). Therefore, if translational efficiency is the consequence of the compatability of adjacent tRNA molecules on a translating ribosome, then the translational step time across a given codon pair could differ depending on the identity of the tRNA molecules decoding these codons. In this case, other factors that affect isoacceptor tRNA selection, such as III-I dinucleotide biases and codon-anticodon stacking energies, could also affect translational efficiency. These sorts of effects on the results presented in Table I cannot be excluded.

To confirm the conclusions drawn from the results of the translation initiation assay, we employed an independent trp attenuator-based assay to examine the translational efficiency of the same highly over-represented and under-represented codon pairs. With this assay we demonstrated that the highly under-represented Leu-Thr codon pair (ACC CUG, CHISQ3 = −27.3) that severely inhibits translation initiation also restricts translation of the polypeptide coding sequence in the trp leader attenuator region and causes increased transcription through the trp attenuator.

The observations reported here suggest that the discontinuities in the translation rates of genes are “hard wired” into the sequence of each gene. If this is so, then it is reasonable to
assume that the use of one codon next to another has co-evolved with the structure and abundance of tRNA isoacceptors in order to control the rates of translational step times without imposing additional constraints on amino acid sequences or protein structures. This hypothesis offers a simple explanation for the large, seemingly excessive, number of tRNA isoacceptor molecules found in all living cells. It implies that, for any given amino acid sequence, appropriately biased codon pairs can be employed to set the translational step times for the addition of amino acids to the growing polypeptide chain. In this manner, translational pauses important for the folding and other functions of nascent polypeptide chains can be incorporated into the DNA coding sequence of a gene. However, since the relative translational efficiencies of only a small number of codon pairs have been studied, it is not yet possible to ascertain how consistent the relationship between translational step times and codon pair bias values will be. As more codon pairs are examined, it will be interesting to determine if it is possible to use these values in a way that will be predictive of relative translational step times for the identification of translational pause sites.

In summary, we have employed two independent assays to demonstrate a close relationship between the translational efficiency of a codon pair and its degree of bias in protein coding sequences of *E. coli*. We have demonstrated that at least some codon pairs that are observed in protein coding sequences more frequently than predicted by the frequency of the individual codons in that pair are translated slower than codon pairs that are found less frequently than expected. Additionally, we have demonstrated a general relationship between the utilization bias of a codon pair and its translational bias; the more over-represented these codon pairs are the slower they are translated, and the more under-represented they are the faster they are translated. We have also shown that the translational efficiency of a given codon pair is correlated with its codon pair utilization bias and not with the utilization frequency of the individual codons of the pair. If we conclude, therefore, that codon pair bias is related to translational step times which are mechanistically related to the compatibility of adjacent tRNA molecules on the surface of a translating ribosome, then each of these observations is predicted by the results of our statistical analyses of the codon pair utilization patterns in protein coding sequences of *E. coli* (Gutman and Hatfield, 1989; Hatfield and Gutman, 1992).

Acknowledgments—We are grateful to Professor Charles Yanofsky for his advice and generous gift of strains and to George Gutman for helpful discussions. We also thank our laboratory colleagues, Elaine Ito, Bhavin Parekh, Kyu Rhee, Steve Sheridan, and She-Pin Hung, for helpful comments and discussions.

REFERENCES
Bergmann, J. E., and Lodish, H. F. (1979) J. Biol. Chem. 254, 11397–11407
Bosi, L., and Roth, J. R. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 123–127
Bradford, M. M. (1976) Anal. Biochem. 72, 248–254
Colby, D. S., Schell, P., and Guthrie, C. (1976) Cell 9, 449–463
Cole, J. R., Olsén, C. L., Hershey, J. W., Gruenberg-Manago, M., and Nomura, M. (1987) J. Mol. Biol. 196, 383–392
Feinstein, S. I., and Altman, S. (1977) J. Mol. Biol. 112, 453–470
Fluck, M., Salsar, W., and Epstein, R. H. (1977) Mol. & Gen. Genet. 151, 137–149
Furiste, J. P., Pansegraue W., Frank, R., Blocker, H., Schütz, P., Bagdassarian, M., and Lanka, E. (1986) Gene (Amst.) 48, 119–131
Gold, L., and Stormo, G. (1987) in Escherichia coli and Salmonella typhimurium Cellular and Molecular Biology (Neidhardt, F. C., ed) Vol. 2, pp. 1302–1307, American Society of Microbiology, Washington D.C.
Goldman, E., Holmes, W. M., and Hatfield, G. W. (1979) J. Mol. Biol. 129, 567–585
Gutman, G. A., and Hatfield, G. W. (1989) Proc. Natl. Acad. Sci. U.S.A. 86, 3699–3703
Guttenson, N. I., and Koshland, D. E. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 4894–4896
Hatfield, G. W., and Gutman, G. A. (1992) in Transfer RNA in Protein Synthesis (Hatfield, D. L., Lee, B. J., and Pirtle, R. M., eds.) pp. 157–190, CRC Press, Boca Raton, FL
Holmes, W. M., Goldman, M., and Miner, T. A., and Hatfield, G. W. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 1393–1397
Ikenura, T. (1992) in Transfer RNA in Protein Synthesis (Hatfield, D. L., Lee, B. J., and Pirtle, R. M., eds) pp. 113–124, CRC Press, Boca Raton, FL
Koehler, A. S., and Reddy, B. V. B. (1986) J. Biomol. Struct. Dynam. 3, 725–728
Kolter, R., and Yanofsky, C. (1983) J. Mol. Biol. 175, 299–312
Landick, R., and Yanofsky, C. (1987) in Escherichia coli and Salmonella typhimurium (Neidhardt, F. C., ed) Vol. 2, pp. 1276–1301, American Society for Microbiology, Washington, D.C.
Landick, R., Yanofsky, C., Choo, K., and Phung, L. (1990) J. Mol. Biol. 216, 25–37
Liljestrom, H., and von Heine, G. (1987) J. Theor. Biol. 124, 43–55
Miller, J. H. (1972) in Experiments in Molecular Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
Sharp, P. M., and Li, W. H. (1986) Nucleic Acids Res. 14, 7377–7394
Simons, R. W., Hounam, F., and Kiecker, N. (1987) Gene (Amst.) 58, 85–96
Smith, D., and Yarus, M. (1989) Proc. Natl. Acad. Sci. U.S.A. 86, 4397–4401
Sorensen, M., Kurland, C. G., and Pedersen, S. (1989) J. Mol. Biol. 207, 365–377
Wek, R. C., and Hatfield, G. W. (1988) J. Mol. Biol. 203, 643–663
Yarus, M., and Curran, J. (1992) in Transfer RNA in Protein Synthesis (Hatfield, D. L., Lee, B. J., and Pirtle, R. M., eds) pp. 319–365, CRC Press, Boca Raton, FL
