# Artificial Mutants Generated by the Insertion of Random Oligonucleotides into the Putative Nucleoside Binding Site of the HSV-1 Thymidine Kinase Gene†

Dipak K. Dube,‡ Jay D. Parker, David C. French, David S. Cahill, Syamalima Dube,‡ Marshall S. Z. Horwitz, Khan M. Munir, and Lawrence A. Loeb*

*Department of Pathology, The Joseph Gottstein Memorial Cancer Research Laboratory, University of Washington, Seattle, Washington 98195*

*Received July 19, 1991; Revised Manuscript Received September 3, 1991*

ABSTRACT: We have obtained 42 active artificial mutants of HSV-1 thymidine kinase (ATP:thymidine 5′-phosphotransferase, EC 2.7.1.21) by replacing codons 166 and 167 with random nucleotide sequences. Codons 166 and 167 are within the putative nucleoside binding site in the HSV-1 *tk* gene. The spectrum of active mutations indicates that neither $Ile^{166}$ nor $Ala^{167}$ is absolutely required for thymidine kinase activity. Each of these amino acids can be replaced by some but not all of the 19 other amino acids. The active mutants can be classified as high activity or low activity on two bases: (1) growth of *Escherichia coli* KY895 (a strain lacking thymidine kinase activity) in the presence of thymidine and (2) uptake of thymidine by this strain, when harboring plasmids with the random insertions. *E. coli* KY895 harboring high-activity plasmids or wild-type plasmids can grow in the presence of low amounts of thymidine (<1 μg/mL), but are unable to grow in the presence of high amounts of thymidine. On the other hand, *E. coli* KY895 harboring low-activity plasmids can grow at a high concentration of thymidine (>50 μg/mL) in the media. The high-activity plasmids also have an enhanced [$^3$H]dT uptake. The amounts of thymidine kinase activity in vitro in unfractionated extracts do not correlate with either growth at low thymidine concentration or the rate of thymidine uptake. Heat inactivation studies indicate that the mutant enzymes are without exception more temperature-sensitive than the wild-type enzyme. This thermolability could account for the less than expected thymidine kinase activity in the extracts and suggests that amino acid substitutions at $Ile^{166}$ and $Ala^{167}$ have produced major changes in protein stability.

Generating mutants by inserting random nucleotide sequences into genes has proven to be a facile method for selecting and defining the repertoire of sequences that have the potential to code for functional genetic elements (Horwitz & Loeb, 1986; Oliphant & Struhl, 1987; Joyce, 1989; Tuerk & Gold, 1990; Dube & Loeb, 1989; Horwitz et al., 1989; Ellington & Szostak, 1990). This approach has been applied to promotor sequences (Horwitz & Loeb, 1986; Oliphant & Struhl, 1987), binding sites of transcriptional factors (Blackwell & Weintraub, 1990), and active sites in enzymes (Dube & Loeb, 1989). It provides a new methodology for analyzing the relationship of protein structure to function (Dube & Loeb, 1989; Horwitz et al., 1989; Blackwell & Weintraub, 1990; Oliphant & Struhl, 1989; Kaiser et al., 1987). In our initial studies, we have remodeled the promotor for the tetracycline resistance gene (Horwitz & Loeb, 1986) as well as the active site of the gene encoding β-lactamase, an enzyme that cleaves the β-lactam ring of penicillin antibiotics (Dube & Loeb, 1989). Our results and those of others indicate that a surprisingly large number of nucleotide sequences can substitute for those found in nature. In the case of promotors, many of these artificial sequences are more active than wild-type sequences (Horwitz & Loeb, 1988).

The effectiveness of random sequences for the generation of mutant enzymes is based on the hypothesis that multiple amino acid substitutions can be tolerated within the active site and that a large number of the substitutions may yield enzymes with altered or even new catalytic activities. The selection of biologically active molecules from large population(s) of random sequences allows exploration of the relationship of protein structure to function in the absence of extensive knowledge about the structure or catalytic function of a protein.

In this study, we present a general approach for the insertion of random nucleotide sequences into genes within plasmids, and we report on initial studies with thymidine kinase. Thymidine kinase from Herpes simplex virus type 1 (HSV-1) is a multifunctional enzyme exhibiting thymidine kinase, thymidylate kinase, and cytidine kinase activities. It also phosphorylates acycloguanosine, a guanine derivative employed clinically for antiviral activity (Fyfe et al., 1978). The HSV-1 thymidine kinase gene has been cloned, sequenced, and expressed in *Escherichia coli* (McKnight, 1980; Wagner et al., 1981; Kit et al., 1981; Garpin et al., 1981). From kinetic analyses, it has been proposed that HSV-1 thymidine kinase contains two substrate binding sites—one for ATP and one for thymidine. The ATP binding site has been identified by Liu and Summers using site-directed mutagenesis (Liu & Summers, 1988). On the basis of mutants that are resistant to antiviral drugs and those that exhibit alterations in the $K_m$ for thymidine, it has been proposed that amino acid residues 168–176 constitute a portion of the putative thymidine binding site (Darby et al., 1986). Our previous studies support this assignment; a mutant HSV-1 obtained by substituting $Ile^{166}$ and $Ala^{167}$ with $Glu^{166}$ and $Leu^{167}$ is unable to complement *E. coli* KY895 that is devoid of thymidine kinase activity (Dube et al., 1991). In order to further delineate the thymidine binding site, we have substituted random nucleotides for those in codons 166 and 167, and after transforming *E. coli* KY895,

† This research was supported by Outstanding Investigator Grant R-35-CA39903 from the National Cancer Institute.

* To whom correspondence should be addressed.

‡ Present address: Department of Medicine, SUNY—Upstate Medical Center, 750 E. Adams St., Syracuse, NY 13210.

0006-2960/91/0430-11760$02.50/0 

we obtained a large number of mutants that exhibit thymidine kinase activity. Each of the mutants has amino acid changes at positions 166 and/or 167. An analysis of the properties of these mutants provides insights on the function of thymidine kinase and on its requirements for catalytic activity.

## Materials and Methods

Plasmid pHETK2, *E. coli* KY895, and rabbit anti-TK antiserum were generous gifts of Dr. William Summers, School of Medicine, Yale University, New Haven, CT. Restriction endonucleases were obtained commercially and were used according to suppliers' instructions. Standard molecular cloning methods were employed (Maniatis et al., 1982). *E. coli* KY895 cells were rendered competent using the rubidium chloride method of Hanahan (1983).

Oligonucleotides were synthesized by Operon Technologies (San Pablo, CA) using phosphoramide chemistry. For the construction of the nonproducer plasmid pKTPD, we used 2 partially complementary oligomers of 17 nucleotides in length: (i) 5′-CCCCTCGAGCGCGGTAC-3′, (ii) 3′-TCGAGGGGAGCTCGCGC-5′. For the construction of random sequences at codons 166 and 167, two large oligonucleotides with 3′-complementary termini were used: (I) 5′-TGGGAGCTCACATGCCCCGCCCCCGGCCCTC-ACCCTCATCTTCGATCGCCAT-3′ (52-mer); (II) 5′-ATAAGGTACCGCGCCGCGGGGTAGCACAGGAG-GGCNNNNNNGGGATGGCGATCGAA-3′ (56-mer).

The stretches of random nucleotides within insert II are designated by Ns and were synthesized using mixtures containing equimolar amounts of all four nucleoside phosphoramidite derivatives.

*Construction of an Inactive Vector for the Insertion of Random Nucleotide Sequences.* The nonproducer strain pTKPD was obtained by substituting a small oligonucleotide for a portion of the putative nucleoside binding site located between two unique restriction sites in the HSV-1 thymidine kinase gene. The parent pHETK2 (Waldman et al., 1983) encodes the HSV-1 *tk* gene under the control of $P_L$ and $P_R$ promoters of λ phage. The expression of the HSV-1 gene from these promoters is under the control of a temperature-sensitive repressor of λ phage. The plasmid also contains the gene for β-lactamase. Both oligonucleotides i and ii (20 pmol of each) were kinased and annealed to form a double-stranded oligonucleotide with *Kpn*I- and *Sst*I-compatible ends and with an internal *Xho*I site. The plasmid pHETK2 was digested with *Sst*I and *Kpn*I restriction endonucleases, and the large fragment was isolated by agarose gel electrophoresis and subsequent electroelution (Figure 1, step I). Two picomoles of the large fragment was ligated with 6 pmol of the double-stranded oligonucleotides containing a unique *Xho*I site and flanking *Sst*I or *Kpn*I sites. The resultant double-stranded circular DNA product was used to transform competent *E. coli* KY895 cells (Igarishi et al., 1983). Clones containing the recombinant plasmid pTKPD grow on LB plates containing 50 μg/mL carbenicillin. The presence of recombinant plasmid DNA was verified by the cleavage at the *Xho*I site. The inability of pTKPD to support the growth of *E. coli* KY895 in the thymidine kinase selection medium suggests that it does not produce a functional thymidine kinase. The vector containing the disrupted *tk* gene was digested with *Sst*I and *Kpn*I, purified by gel electrophoresis and electroelution, and used for the insertion of oligonucleotides containing random sequences.

*Construction of Oligonucleotides Containing Random Nucleotide Sequences.* A large oligonucleotide containing stretches of random sequences was constructed and used to replace the oligonucleotides in the nonproducer strain. Oligonucleotides I and II containing 12 complementary base pairs at their 3′ ends were elongated by DNA polymerase, producing a 96-mer with a stretch of 6 random nucleotides at codons 166 and 167. Eighty-five picomoles each of oligonucleotides I and II was annealed in a 20-μL volume containing 50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, and 1 mM dithiothreitol at 65 °C for 10 min and then gradually cooled to 24 °C. Thereafter, all four dNTPs were added to a final concentration of 1 mM, and the noncomplementary segments were copied by 5 units of the large fragment of *E. coli* DNA polymerase I at 37 °C for 15 min. The resultant double-stranded oligomer was digested with *Sst*I and *Kpn*I to generate sticky ends for directional cloning into the pTKPD-derived vectors. The substitutions at these restriction sites code for three silent mutations and provide markers to indicate that the mutant thymidine kinase gene was indeed coded for by the random insert and did not arise by some other mutational process during transfection or selection.

*Selection of TK-Positive Clones.* Active *tk* genes were selected from plasmids containing random inserts in place of the nonproducer sequences. The nonproducer plasmid pTKPD was digested with *Sst*I and *Kpn*I, and the large fragment was isolated by agarose gel electrophoresis and subsequent electroelution. Ligation of 5 pmol of pTKPD-derived vectors with 15 pmol of the double-stranded oligonucleotides containing random sequences was carried out with T4 DNA ligase at 4 °C overnight. The ligation product was digested with *Xho*I to cleave any contaminant religated pKTDP and then used to transform competent *E. coli* KY895 cells. Aliquots of the transformation mixture were plated either on LB/carbenicillin plates, to determine the number of transformants, or on TK selection plates to identify clones expressing thymidine kinase. TK selection plates contained 50 μg/mL carbenicillin, 4 μM fluorodeoxyuridine, 0.8 μM thymidine, 40 μM uridine, 2% peptone (BRL), and 0.5% Gel-Rite (Scott Laboratories, Inc., Carson, CA) or 1.2% agar. For thymidine kinase selection, plates were incubated at 42 °C for 8 h to induce the expression and then shifted to 37 °C. Codon assignments for recombinant clones were made by sequencing plasmid miniprep DNA as previously described (Dube & Loeb, 1989; Sanger et al., 1977).

*Plate Assays for Measuring Growth as a Function of Thymidine Concentration.* A linear concentration gradient of thymidine (0–5 μg/mL, or stated otherwise) was an adaptation of the method of Schultz et al. (1987) used for plates with a linear gradient of β-lactam antibiotics. The medium contained 10–25 μg/mL fluorodeoxyuridine, 25 μg/mL uridine, 50 μg/mL carbenicillin, and 1.2% agar or 5 g/L Gel-Rite. In addition, we added X-gal (25 μg/mL) and IPTG (25 μg/mL) to enhance visualization of the recombinant clones that express β-galactosidase. The thymidine gradient was made by raising one end of a petri dish 5 mm and pouring into the dish L-agar containing the selection medium and appropriate concentrations of thymidine. The agar was permitted to solidify overnight in the tilted dishes and placed onto a horizontal surface, and an equal volume of the L-agar medium was added except that no thymidine was present. The plates were used immediately after solidification of the upper agar layer. An aliquot (30–50 μL) of a $1:10^5$ dilution of an overnight culture was used to inoculate the plates held at a slanting position.

*Analysis of Thymidine Kinase Activity. E. coli* (*tdk*$^-$) harboring plasmids containing nucleotide substitutions within the HSV-1 *tk* gene were grown in LB containing 50 μg/mL carbenicillin at 30 °C to log phase. Thereafter, the temperature was rapidly shifted to 42 °C, and incubation was con-

tinued with vigorous shaking for an additional 3 h. Cells were harvested by centrifugation at 4 °C, washed with 0.9% isotonic saline solution, suspended in ice-cold extraction buffer (50 mM Tris-HCl, 4 mM $MgCl_2$, 50 mM KCl, and 5 mM $\beta$-mercaptoethanol), and sonicated for 2 min at 4 °C. The extract was centrifuged at 10000*g* for 15 min at 4 °C, and the supernatant was used as an enzyme source. Thymidine kinase activity was determined by the method of Chen et al. (1979) in a total reaction volume of 50 $\mu$L containing 0.8 $\mu$M [$^3$H]dT (specific radioactivity 1000–1500 cpm/pmol), 3.0 mM ATP, 10 mM $MgCl_2$, 2 mM DTT, 50 mM Tris-HCl (pH 7.5), and 2.5% glycerol. The reaction was incubated at 35 °C, aliquots (5–10 $\mu$L) were removed at 0, 5, and 10 min, soaked onto DE81 paper, washed immediately with 95% ethanol, and dried and the amount of radioactivity rendered DEAE-adsorbable was determined.

For determinations of the thermolability of thymidine kinase, 200 $\mu$L of supernatant fraction was preincubated in a total volume of 350 $\mu$L containing 50 mM Tris-HCl (pH 7.5), 250 $\mu$g/mL bovine serum albumin, 10 $\mu$g of the proteinase inhibitor aprotinin, 0.05 $\mu$g of leupeptin, and 0.05 $\mu$g of pepstatin. After preincubation at 40 or 43 °C for the times indicated in the figures, 20-$\mu$L aliquots were removed, cooled to 4 °C, and assayed for residual thymidine kinase activity at 35 °C as detailed above.

*Uptake of [$^3$H]Thymidine.* Measurements of the rate of uptake of thymidine into *E. coli* tdk$^-$ harboring pHETK constructs with/without mutations in the HSV-1 *tk* gene were carried out as follows: aliquots (2 mL) of overnight cultures were diluted 1:100 with LB containing carbenicillin (50 $\mu$g/mL) and grown to log phase at 30 °C. To induce thymidine kinase, the temperature was shifted rapidly to 42 °C, and incubation was continued for 2 h with constant shaking. Thereafter, the culture was cooled to 25 °C, [$^3$H]dT (0.75 $\mu$M final concentration; specific activity 3000–4000 cpm/pmol) was added in a total volume of 2 mL, and after incubation for the times indicated, 0.05–0.2-mL aliquots were immediately filtered through a nitrocellulose filter (0.45 $\mu$m) and washed with ice-cold 0.9% saline. The transfer to nitrocellulose and washing were completed within a minute.

*Purification of HSV-1 Thymidine Kinase.* The purification of the enzyme was carried out using the affinity chromatographic technique as described by Cheng and Ostrander (1976).

*Gel Electrophoresis and Immunoblotting.* SDS–polyacrylamide gel electrophoresis was performed according to the method of Laemmli (1970). *E. coli* KY895 transformed with different plasmid constructs was grown in LB with 100 $\mu$g/mL carbenicillin to log phase at 30 °C. The temperature was shifted to 42 °C, and the cultures were maintained at the elevated temperature for 2 h with constant shaking. Cells from 1 mL of heat-induced *E. coli* KY895 were collected by centrifugation. The pellets were resuspended in 0.1 mL of SDS loading buffer and heated at 95 °C for 5 min. An aliquot of the sample was loaded directly on a 10% polyacrylamide gel containing SDS. The protein transfer and immunoblotting were carried out according to the method of Davis et al. (1986).

## Results

A general methodology for identifying infrequent active nucleotide sequences among a vast number of possible random permutations requires a series of procedures to prevent contamination with miniscule amounts of wild-type sequences as well as a highly sensitive method for selecting new sequences based on biological activity. We have designed a general

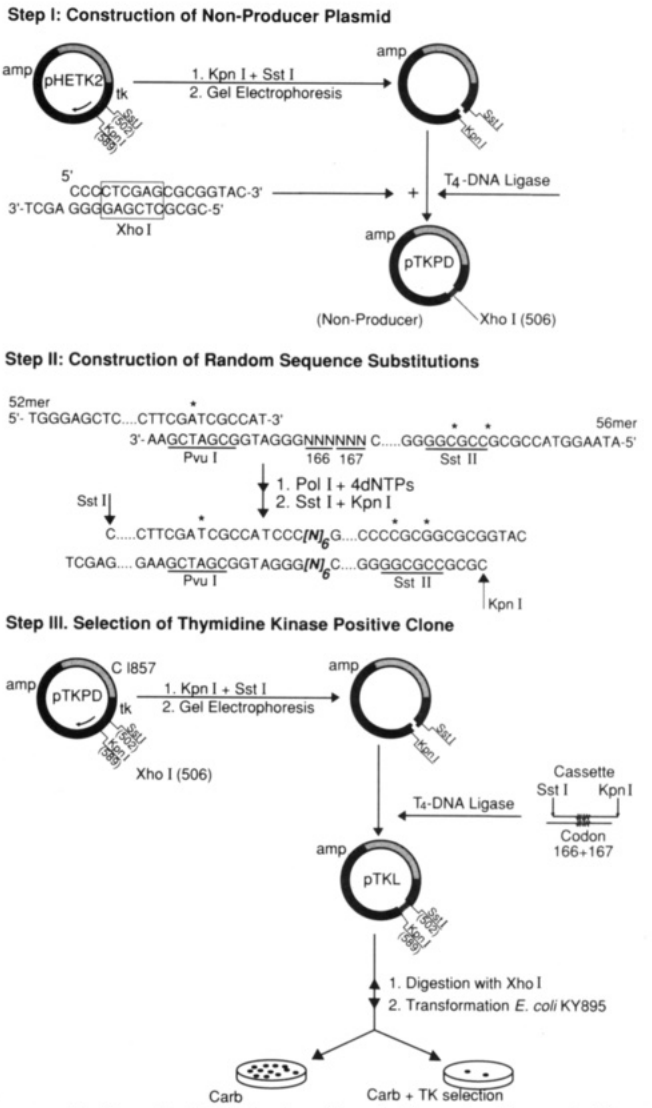


FIGURE 1: Overall scheme for insertion of oligonucleotides containing random DNA inserts into the HSV-1 *tk* gene. Step I: The construction of the nonproducer plasmid pTKPD. Step II: The synthesis of an oligonucleotide containing random sequences in place of codons 166 and 167. * denotes nucleotide substitution used to create silent mutations at these sites. Step III: Selection of TK-positive clones from plasmids with random substitutions.

strategy for the substitution of random sequences in a double-stranded DNA plasmid. The method focuses on the elimination of potential contamination by parental DNA molecules. In this procedure, an oligonucleotide containing an in-frame deletion is first substituted for the target sequence so as to create an inactive recombinant plasmid (a nonproducer strain) that serves as the starting vector for the insertion of random nucleotide sequences. Moreover, the inactive insert contains a unique restriction site so that any residual nonproducer vectors can be cleaved and thus eliminated from the pool of vectors prior to transformation. As an initial approach, we designed a method with the potential of inserting a long sequence of random nucleotides into the herpes thymidine kinase gene, but limited our studies to the replacement of codons 166 (Ile) and 167 (Ala) within the putative nucleoside binding site.

The general method for the insertion of oligonucleotides containing random inserts into double-stranded DNA is diagrammed in Figure 1. We first constructed a nonproducer strain containing a small oligonucleotide that replaced a portion of the thymidine kinase gene (step I). A large fragment

| Amino Acids Codon No 166 | 167 | Nucleotide sequence | Activity | Amino Acids Codon No 166 | 167 | Nucleotide sequence | Activity |
|---|---|---|---|---|---|---|---|
| Ala | Ala | GCC GCC | L | Phe | Ala | TTT GCT | L |
| | | GCG GCT | | Phe | Gly | TTC GGC | L |
| Ala | Cys | GCG TGC | L | Phe | Ser | TTC AGC | L |
| Ala | Gly | GCG GGC | L | | | | |
| | | GCG GGC | | Ser | Ala | TCT GCA | L |
| | | GCT GGC | | | | TCT GCG | |
| | | GCT GGG | | | | AGC GCC | |
| Ala | Leu | GCT TTA | - | Ser | Leu | TCT CTG | L |
| | | GCG CTA | | | | | |
| | | | | Ser | Ser | TCC TCG | L |
| | | | | | | AGC TCC | |
| Ala | Thr | GCA ACC | L | Ser | Val | AGC GTT | L |
| Arg | Gly | CGA GGC | - | | | | |
| Cys | Ala | TGT GCA | L | | | | |
| Cys | Thr | TGC ACG | L | Thr | Ala | ACC GCC | L |
| Ile | Gly | ATA GGC | L | | | ACA GCA | |
| | | | | | | ACC GCT | |
| Ile | Ser | ATC TCC | L | | | | |
| Ile | Thr | ATA ACT | H | Thr | Cys | ACC TGC | L |
| Leu | Ala | TTA GCG | H | Thr | Glu | ACG GAG | L |
| | | | | Thr | Gly | ACA GGT | L |
| | | CTC GCC | | Thr | Ser | ACC AGT | L |
| | | TTA GCC | | | | ACG AGT | L |
| | | CTA GCG | | | | ACG TCG | |
| | | CTA GCA | | | | ACG TCG | |
| | | CTT GCT | | Thr | Thr | ACC ACG | L |
| | | | | | | ACC ACC | |
| Leu | Cys | CTG TGC | - | Trp | Ala | TGG GCA | L |
| Leu | Gly | CTC GGG | H | Trp | Ser | TGG TCG | L |
| | | TTA GGT | | | | | |
| | | TTA GGA | | Tyr | Ala | TAC GCG | L |
| | | CTA GGC | | | | | |
| | | CTA GGA | | Val | Ala | GTA GCT | H |
| | | CTG GGG | | | | GTA GCA | |
| | | CTT GGG | | | | GTC GCA | |
| Leu | Ile | CTA ATC | L | Val | Gly | GTG GGG | H |
| | | | | Val | Ser | GTG AGC | - |
| | | | | Val | Thr | GTA ACG | H |
| | | | | | | GTG ACA | |
| Leu | Ser | CTG AGC | H | | | GTC ACA | |
| | | CTG TCT | | | | | |
| | | CTC TCA | | | | | |
| | | | | Wild Type | | | |
| Leu | Trp | CTG TGG | H | Ile | Ala | ATC GCC | |
| Leu | Val | TTG GTA | H | Mutant with no detectable activity | | | |
| Met | Ala | ATG GCT | L | Ala | Pro | GCC CCT | |
| | | ATG GCC | | Arg | Asp | AGG GAT | |
| | | | | Thr | Asn | ACG AAT | |
| Met | Ser | ATG AGT | - | Gln | His | CAA CAC | |

FIGURE 2: Amino acids at codons 166 and 167 inferred from nucleotide substitutions that yield active and inactive thymidine kinase mutants. The designations L and H refer to the low- and high-activity classes of mutants, respectively.

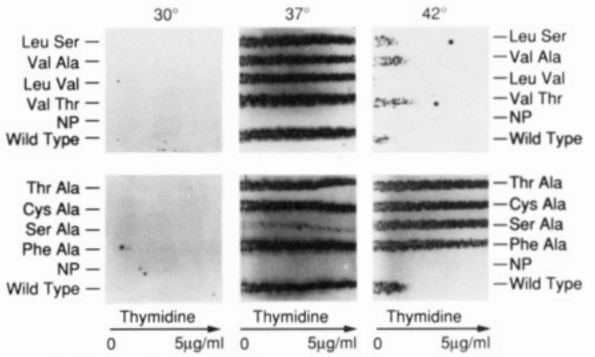


FIGURE 3: Determination of the permissive thymidine concentration for the growth of *E. coli* (devoid of thymidine kinase) harboring different pTKL constructs. Permissive thymidine concentration was determined from the growth of bacteria in plates containing a linear gradient of thymidine concentration from 0 to 5 μg/mL as described under Materials and Methods. NP refers to the nonproducer strain. The plates were incubated at three different temperatures: (A) 30 °C; (B) 37 °C; (C) 42 °C.

containing random nucleotide sequences was copied by *E. coli* polymerase I and used to replace the short oligonucleotide in the nonproducer strain (steps II and III). Sequences that coded for active thymidine kinase were identified on TK selection medium.

*Active Mutants Obtained by Selection.* Using random nucleotides in place of codons 166 and 167, we obtained a total of 3125 transformants that formed colonies on carbenicillin plates; of these, 135 also formed colonies on the TK selection medium. A compilation of nucleotide sequence changes and the inferred amino acid changes among the TK selectants is given in Figure 2. Of the 399 possible amino acid substitutions within the 2 codons, we have detected 40 that can grow on the TK selective medium. Among six colonies that grew on carbenicillin plates but not on TK selection media, four contained amino acid substitutions at codons 166 and 167 and two contained stop codons (Figure 2). The large number of different mutants that can grow on TK selection medium indicates that $Ile^{166}$ and $Ala^{167}$ are not absolutely required for the HSV-1 TK activity. However, not all amino acid substitutions at these positions produce active HSV-1 TK. Fifteen of the 40 amino acid substitutions were detected only once among the 135 TK-positive clones that were sequenced. Thus, it is not surprising that we have not as yet detected the wild-type sequence among the TK-positive clones.

*Growth of Mutants on Gradients of Thymidine.* Growth on thymidine was determined by streaking cultures of each of the mutants onto agar plates containing a gradient of thymidine concentration. Growth of each of the mutants was determined at three different temperatures, 30, 37, and 42 °C, on agar plates containing a linear gradient of 0–5 μg/mL thymidine. The results in Figure 3 show that none of the mutants tested formed visible colonies at 30 °C. At 37 °C, each of the transfectants containing a mutant plasmid, except for those harboring the nonproducer strain, NP, formed colonies throughout the thymidine gradient. At 42 °C, two types of TK-positive mutants can be observed: those that form colonies only in the presence of low amounts of thymidine and those that form colonies at all concentrations of thymidine, even as high as 250 μg/mL (results not shown). The wild type and the mutants that form colonies in the presence of low but not high concentrations of thymidine (Figure 3) have greater amounts of thymidine kinase activity in vitro and are designated as high-activity mutants. These mutants form colonies at thymidine concentrations lower than those of the wild type (Figure 4). Conversely, the mutants that form colonies at

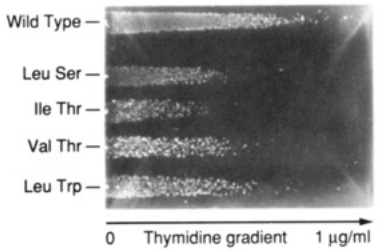


FIGURE 4: Determination of the permissive thymidine concentration of "high-activity" mutants and the wild type using a narrow gradient of thymidine concentration (0–1 μg/mL) at 42 °C.

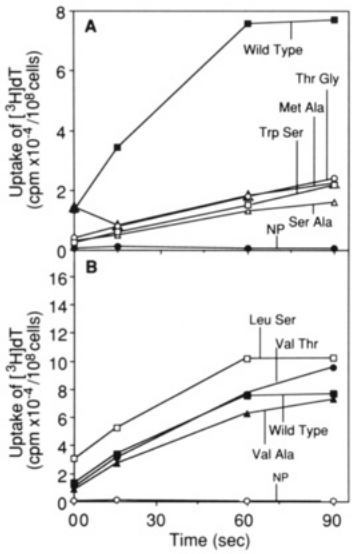


FIGURE 5: Uptake of [$^3$H]dT into *E. coli* harboring constructs with the indicated mutants in the HSV-1 thymidine kinase gene at codons 166 and 167. (A) "Low-activity" mutants; (B) "high-activity" mutants.

all thymidine concentrations tested have less thymidine kinase activity in vitro and are designated low-activity mutants.

*Thymidine Uptake.* We have recently established that thymidine kinase activity is required for the uptake of thymidine into *E. coli* (Dube et al., 1991). In the present investigation, we have measured the uptake of [$^3$H]dT into *E. coli* harboring plasmids containing wild type and selected mutations in the HSV-1 *tk* gene. The results (Figure 5A,B) show that there is essentially no detectable uptake of thymidine in *E. coli* harboring the plasmid with nonproducer *tk* gene. The rate of thymidine uptake is significantly less in the low-activity mutants compared to that of the wild type. The uptake of thymidine in the high-activity mutants is comparable to that of the wild type, and three of the mutants tested have greater activity than the wild type. Among the mutants so far studied, the rate of thymidine uptake parallels the growth rate with low amounts of thymidine (Table I).

*Thymidine Kinase Activity in Vitro.* Thymidine kinase activity was determined in cell-free extracts of *E. coli* harboring plasmids containing selected mutants. In these experiments, cell-free extracts were obtained from log-phase cultures that were induced for the production of thymidine kinase by growth at 42 °C for 12 h. The specific activity of the cell-free extracts containing the wild-type TK was 12.1 units/mg of protein (results not shown), which is in good agreement with published results (Darby et al., 1986). The specific activity of all of the *E. coli* harboring mutant TKs was lower than that of those harboring the wild-type gene. Of the mutants tested, the one with the single substitution of Val for Ile at position 166 exhibited the highest specific activity, 9.3 units/mg of protein (results not shown). Measurements of

Table I: Properties of Artificial HSV-I TK Mutants

| strain | permissive dT concn in plate assay (μg/mL) | uptake of [$^3$H]dT (% of wild type) | in vivo TK act. (% of wild type)[c] | in vitro TK act. (units/mg of protein) |
|---|---|---|---|---|
| wild type | <1.0 | *b* | *a* | 1.0 |
| nonproducer | | <0.8 | <2.9 | <0.001 |
| high activity | | | | |
| Ile-Thr | <1.0 | 111.7 | 127 | 0.08 |
| Leu-Ala | <1.0 | 65.3 | 75 | 0.74 |
| Leu-Ser | <1.0 | 103.7 | 145 | 0.008 |
| Val-Ala | <1.0 | 106.2 | 92 | 0.29 |
| Val-Thr | <1.0 | 170.7 | 169 | 0.01 |
| low activity | | | | |
| Ala-Ala | >5.0 | | | 0.011 |
| Ala-Cys | >50.0 | 9.1 | 13 | 0.026 |
| Ala-Gly | >40.0 | 13.6 | 15 | 0.003 |
| Cys-Ala | >50.0 | 30.5 | 46 | 0.148 |
| Cys-Thr | >50.0 | 23.8 | 23 | 0.000 |
| Ile-Gly | >5.0 | 37.8 | | 0.066 |
| Phe-Ala | >5.0 | 18.3 | 38 | 0.025 |
| Phe-Gly | >50.0 | 13.4 | 18 | 0.01 |
| Ser-Ala | >5.0 | 18.4 | | 0.159 |
| Ser-Leu | >5.0 | 39.5 | | 0.24 |
| Ser-Val | >50.0 | 18.0 | 18 | 0.0046 |
| Thr-Ala | >50.0 | 30.8 | | 0.11 |
| Thr-Cys | >200.0 | 31.4 | | 0.041 |
| Thr-Gly | >5.0 | 14.7 | 18 | <0.01 |
| Trp-Ala | >200.0 | 12.1 | 29 | 0.01 |
| Trp-Ser | >50.0 | 30.8 | 30 | <0.01 |

[a] In three different experiments, the mutants were grown concurrently and compared to wild-type control which was taken as 100%. The absolute TK activity of wild type was 63.0, 17.4, and 19.7 (pmol of [$^3$H]dTMP/$10^8$ cells/10 min). [b] In three separate experiments, the uptake of [$^3$H]dT was 71.65, 26.75, and 37.21 (pmol/$10^8$ cells/10 min). The percent uptake is that compared to the wild type assayed concurrently. [c] The in vivo TK activity reflects the fraction of [$^3$H]dT phosphorylated after being transported into the cell. The in vivo activity was determined as previously described (Dube et al., 1991).

thymidine kinase activity were carried out after incubation at 42 °C for 2, 6, and 12 h. Activity in wild-type extracts is highest after overnight induction, but for some mutants (e.g., $Ile^{166}$ $Thr^{167}$) it is highest after only 2–3 h. The amount of thymidine kinase activity after induction for 2 h was determined (Table I). The activity in cell extracts obtained from *E. coli* harboring the wild-type plasmid was approximately 1.0 unit/mg of protein. In all cases, extracts obtained from the mutants grown and harvested under identical conditions had a lower amount of TK activity. Thus, even though thymidine kinase activity in vivo, as determined on the basis of uptake or growth at low thymidine concentrations, is greater in some of the mutants than in the wild type, their TK activity in vitro is not correspondingly greater than that of the wild type.

*Analysis of Thymidine Kinase by Immunoblots.* The lower levels of TK activity in vitro from higher activity mutants could be the result of proteolytic degradation, thermal instability, or both. An analysis of proteins in these mutants with polyclonal antisera raised in rabbits against wild-type HSV-1 TK shows that all eight mutants tested and the wild-type express a predominant immunoreactive polypeptide of 46 000–48 000 Da (Figure 6), similar to that reported by others (Liu & Summers, 1988). This band corresponds to the major polypeptide obtained after extensive purification of HSV-1 TK by affinity chromatography. Extracts obtained from *E. coli* harboring the nonproducer plasmid (containing an inframe deletion) demonstrate a smaller immunoreactive polypeptide (Figure 6). The consistency in the amount of immunoreactive protein among the extracts obtained from the different mutants strongly suggests that there is no extensive degradation of any of the mutant enzymes studied. The data indicate that the

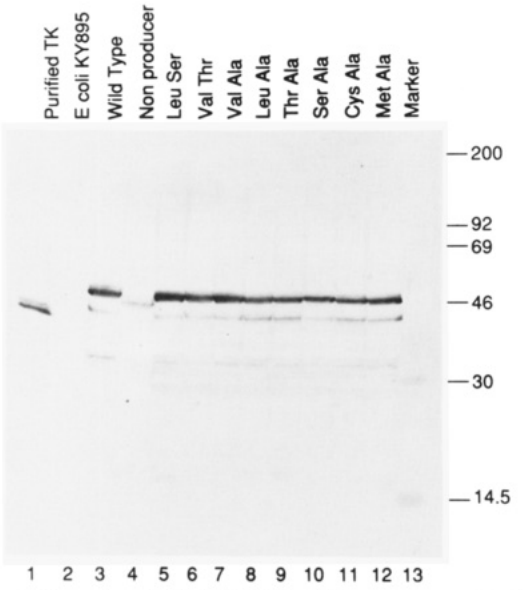


FIGURE 6: Expression of the HSV-1 thymidine kinase gene in *E. coli*. Lane 1, purified HSV-1 thymidine kinase enzyme; lane 2, extract from *E. coli* KY895; lane 3, *E. coli* KY895 transformed with pHETK2; lane 4, *E. coli* KY895 transformed with the nonproducer plasmid pKTPD; lanes 5–8, *E. coli* KY895 transformed with constructs containing high-activity mutants (Leu-Ser, Val-Thr, Val-Ala, or Leu-Ala); lanes 9–12, *E. coli* KY895 transformed with constructs containing low-activity mutants (Thr-Ala, Ser-Ala, Cys-Ala, or Met-Ala).

low activity in some of the mutants is not the result of enzyme proteolysis.

*Thermal Stability of Mutants of HSV-1 Thymidine Kinase.* The effects of preincubation at 40 °C on TK activity in extracts obtained from *E. coli* that harbored different HSV-1 *tk* mutants were determined. In total, extracts from 20 different mutants were tested, and without exception, each of them was less stable than those containing the wild-type enzyme. This thermolability was manifested in both the high-activity (Figure 7A) and the low-activity mutants (Figure 7B). Mutant Leu-Ser and mutant Val-Ala were found to be the two most thermostable at 40 °C (Figure 7). Yet, even they are heat-labile compared to the wild type at 43 °C (results not shown).

## Discussion

The insertion of random nucleotide sequences at defined sites within genes provides an opportunity to create artificial active molecules. The protocol we present has several important features that should be of general utility in screening large populations of random sequences. First, we have constructed a nonproducer strain and used it as an intermediate vector for the insertion of random sequences. This permits one to bypass the use of the wild-type strain for the insertion of random sequences and thus minimizes contamination. Second, the insert in the nonproducer strain contains a unique restriction site that allows us to quantitate the completeness of digestion by restriction enzymes used to cleave out the insert. Third, the random sequence insert is large enough to permit the placement of random sequences at different positions. Fourth, the random sequence insert contains two other unique restriction sites that can be used to measure the extent of ligation, to insert additional random sequences, as well as to serve as a marker for plasmids that contain random sequences. Fifth, the presence of the original single restriction site provides a mechanism for elimination of any nonproducer plasmid DNA prior to transfection. Lastly, functional artificial enzymes are identified by a positive selection assay rather than by screening

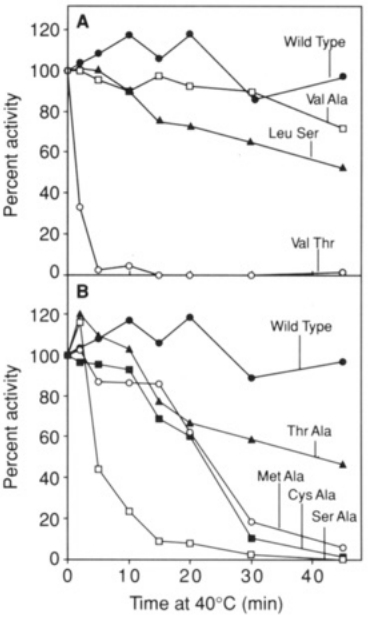


FIGURE 7: Heat inactivation of thymidine kinase at 40 °C using the cell-free extract as the enzyme source. (A) Comparison of the wild-type with the high-activity mutants. (B) Comparison of the wild-type with the low-activity mutants.

large numbers of recombinant bacterial clones. Even though the results presented are limited to replacements of only 2 codons, we have used a similar protocol for the insertion of 33 random nucleotides into the thymidine kinase gene with the recovery of active enzyme (unpublished results).

It has been proposed that amino acid residues 168–176 encompass a portion of the HSV-1 thymidine kinase gene that constitutes the putative nucleoside binding site. We have replaced the nucleotide sequences at codons 166 and 167 with a set of oligonucleotides containing all possible substitutions and selected from transformed colonies 40 clones that produce a functionally active thymidine kinase in *E. coli*. The sequencing of a limited number of recombinant clones that grow in carbenicillin but not in TK selection medium indicates that not all of the amino acid substitutions at these positions yield an active thymidine kinase. The frequency by which the same mutant clones were independently isolated indicates that we did not sample all of the possible amino acid replacements that can function at this site.

On the basis of the ability of *E. coli* harboring mutant plasmids to grow at low thymidine concentrations and on the rate of thymidine uptake, we divided the mutants into two groups—designated high activity and low activity. The growth experiment was carried out in plates containing concentration gradients of thymidine at three temperatures: 30, 37, and 42 °C. The lack of growth of nontransformed *E. coli* KY895 and of the nonproducer strain establishes the requirement of TK activity for growth. HSV-1 *tk* genes in pHETK2 or pTKL are under the control of the temperature-sensitive repressor of bacteriophage λ. At 30 °C, there is no expression of the HSV-1 *tk* gene. At 37 °C, both high-activity and low-activity mutants grow even at very high concentrations of thymidine (Figure 3) presumably due to leakiness of the CI857 repressor gene. For the low-activity mutants, growth is observed at 42 °C at all thymidine concentrations tested. For the wild-type and high-activity mutants, growth is observed only at low thymidine concentrations. We speculate that the lack of growth of the high-activity mutants at elevated concentrations of thymidine is due to an alteration of nucleotide pools in *E. coli*. At 42 °C, there is maximal expression of the HSV-1 *tk*

gene and enhanced thymidine uptake, resulting in an increase in thymidylate pools. And, indeed, there are reports that disruption of balanced pools of deoxyribonucleoside triphosphates has dramatic consequences in prokaryotic as well as eukaryotic cells (Bernstein et al., 1972; Barclay et al., 1981; Meuth, 1989).

The finding of multiple amino acid substitutions at these positions indicates that neither $Ile^{166}$ nor $Ala^{167}$ is absolutely required for TK activity. At least 11 other amino acids can replace $Ile^{166}$, and 9 others can replace $Ala^{167}$. For higher activity as judged from the plate assay, hydrophobic branched-chain amino acids, e.g., Ile, Leu, and Val, are found more frequently in codon 166. On the contrary, hydrophobic amino acids (Ala, Trp, and Val) or polar amino acids with an uncharged R group (Ser and Thr) are preferable for codon 167. A change of one amino acid at position 166 alters the TK activity; $Ile^{166}$ and $Gly^{167}$ together produce a TK with low activity; $Val^{166}$ and $Gly^{167}$ and $Leu^{166}$ and $Gly^{167}$ produce a TK with high activity. These observations, along with other reports, may argue that the catalytic property of an enzyme is determined by only a few amino acid residues and replacement of these by many others does not alter the conformational stability (Knowles, 1987; Reidhaar-Olson & Sauer, 1988).

The classification of the high-activity and low-activity mutants is in accord with the results on rates of [$^3$H]dT uptake. This association between thymidine kinase activity in vivo and thymidine uptake reinforces the experiments of McKeown et al. (1976) demonstrating that *E. coli* deficient in TK are also deficient in thymidine uptake. Thus, TK is likely to be involved in thymidine transport in *E. coli*, and HSV-1 TK can substitute for the endogenous *E. coli* TK in this process. The high-activity TK mutants also exhibit increased thymidine uptake (Figure 5 and Table I). However, we do not find a proportional increase in the amount of TK activity in vitro among the high-activity mutants (Table I).

Heat inactivation studies indicate that without exception each of the mutant TKs assayed is less stable than the wild-type enzyme. This thermolability could account in part for a reduction in thymidine kinase activity in vitro in the mutant enzymes and thus account for the apparent discrepancy between in vivo and in vitro activity. The thermolability of mutants obtained by selection from random libraries is not without precedent (Dube & Loeb, 1989; Oliphant & Struhl, 1989; Schultz & Richards, 1986) and could serve as a clue to mechanisms of selection that were operative during natural evolution. This argument can be substantiated by many different elegant studies (Low & Somero, 1976; Feeney & Osuga, 1976; Zuber, 1976; Argos et al., 1979; Hecht, 1984). In our initial studies on the insertion of completely random oligonucleotides into the active site of the $\beta$-lactamase gene, we obtained 7 active mutants from $2 \times 10^5$ *E. coli* clones harboring recombinant plasmids (Dube & Loeb, 1989). Each of the mutants exhibited a temperature-sensitive $\beta$-lactamase activity. Oliphant and Struhl (1989) inserted a partially random sequence in place of a sequence that codes for a 17 amino acid portion of the active site of $\beta$-lactamase and obtained a series of active mutations, many of which were also temperature-sensitive. The finding that many enzymes with substitutions within or near the active site are temperature-sensitive suggests that during evolution protein stability may have had a greater selective advantage than enhanced catalytic activity.

ACKNOWLEDGMENTS

We are grateful to Dr. W. C. Summers of Yale University for generous gifts of plasmid PHETK-2 and anti-TK antibody.

**Registry No.** Ile, 73-32-5; Ala, 56-41-7; thymidine kinase, 9002-06-6; thymidine, 50-89-5.

REFERENCES

Argos, P., Rossman, M. G., Grau, U. M., Zuber, H., Frank, G., & Tratschiw, J. D. (1979) *Biochemistry 18*, 5698–5703.
Barclay, B. J., & Little, J. G. (1981) *Mol. Gen. Genet. 181*, 305–316.
Bernstein, C., Bernstein, H., Mufti, S., & Strom, B. (1972) *Mutat. Res. 16*, 113–119.
Blackwell, T. K., & Weintraub, H. (1990) *Science 250*, 1104–1110.
Chen, M. S., Walker, J., & Prusoff, W. H. (1979) *J. Biol. Chem. 254*, 10747–10753.
Cheng, Y.-C., & Ostrander, M. (1976) *J. Biol. Chem. 251*, 2605–2610.
Darby, G., Larder, B. A., & Inglis, M. M. (1986) *J. Gen. Virol. 63*, 753–758.
Davis, L. G., Debner, D. M., & Battey, J. F. (1986) in *Basic Methods in Molecular Biology*, pp 311–314, Elsevier, New York.
Dube, D. K., & Loeb, L. A. (1989) *Biochemistry 28*, 5703–5707.
Dube, D. K., Horwitz, M. S. Z., & Loeb, L. A. (1991) *Gene 99*, 25–29.
Ellington, A. D., & Szostak, J. W. (1990) *Nature 346*, 818–822.
Feeney, R. E., & Osuga, D. T. (1976) *Comp. Biochem. Physiol., A: Comp. Physiol. 54A*, 281–286.
Fyfe, J. A., Keller, P. M., Furman, P. A., Miller, R. L., & Ellion, G. B. (1978) *J. Biol. Chem. 253*, 8721–8727.
Garapin, A. C., Colbere-Garapin, F., Cohen-Solal, M., Horadiniceanu, F., & Kouilsky, P. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 815–819.
Hanahan, D. (1985) *DNA Cloning* (Glover, D. M., Ed.) Vol. I, pp 109–135, IRL Press, Washington, DC.
Hecht, M. H., Sturtevant, J. M., & Stauer, R. T. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 5685–5689.
Horwitz, M. S. Z., & Loeb, L. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 7405–7409.
Horwitz, M. S. Z., & Loeb, L. A. (1988) *J. Biol. Chem. 263*, 14724–14731.
Horwitz, M. S. Z., Dube, D. K., & Loeb, L. A. (1989) *Genome 31*, 112–117.
Igarishi, K., Higara, S., & Yura, T. (1967) *Genetics 57*, 643–654.
Joyce, G. F. (1989) *Gene 82*, 83–87.
Kaiser, C. A., Preuss, D., Grisafi, P., & Botstein, D. (1987) *Science 235*, 312–317.
Kauffman, S. A. (1986) *J. Theor. Biol. 119*, 1–24.
Kit, S., Otsuka, H., Quavi, H., & Kit, S. (1981) *Gene 16*, 287–295.
Knowles, J. R. (1987) *Science 236*, 1252–1258.
Laemmli, U. K. (1970) *Nature (London) 227*, 680–685.
Liu, Q., & Summers, W. C. (1988) *Virology 163*, 638–642.
Low, P. S., & Somero, G. N. (1976) *J. Exp. Zool. 198*, 1–12.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
McKeown, M., Kahn, M., & Hanawalt, P. (1976) *J. Bacteriol. 126*, 814–822.
McKnight, S. L. (1980) *Nucleic Acids Res. 8*, 5949–5964.
Meuth, M. (1989) *Exp. Cell Res. 181*, 305–3116.
Oliphant, A. R., & Struhl, K. (1987) *Methods Enzymol. 155*, 568–582.

Oliphant, A. R., & Struhl, K. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 9094–9098.
Reidhaar-Olson, J. F., & Sauer, R. T. (1988) *Science 241*, 53–57.
Sanger, F., Nickler, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Schultz, J. C., Dalbadie-McFarland, G., Neitzel, J. J., & Richards, J. (1987) *Proteins: Struct., Funct., Genet. 2*, 290–297.
Schultz, S. C., & Richards, J. H. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 1198–1202.
Tuerk, C., & Gold, L. (1990) *Science 249*, 505–510.
Wagner, M. J., Sharp, J. A., & Summers, W. C. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 1441–1445.
Waldman, A. S., *Hausselein*, E., & Milman, G. (1983) *J. Biol. Chem. 258*, 11571–11575.
Zuber, H., Ed. (1976) *Enzymes and Proteins from Thermophilic Microorganisms*, Birkhauser Verlag, Basel.

# Arginine-395 Is Required for Efficient in Vivo and in Vitro Aminoacylation of tRNAs by *Escherichia coli* Methionyl-tRNA Synthetase†

Gourisankar Ghosh,‡,§ Hae Yeong Kim,‡ Jean-Philippe Demaret,∥,⊥ Simone Brunie,∥ and LaDonne H. Schulman*,‡
*Department of Developmental Biology and Cancer, Albert Einstein College of Medicine, Bronx, New York 10461, and Laboratoire de Biochimie (CNRS URA 240), Ecole Polytechnique, 91128 Palaiseau Cedex, France*
*Received July 10, 1991; Revised Manuscript Received September 19, 1991*

ABSTRACT: We have previously shown that the anticodon of methionine tRNAs contains the major recognition site required for aminoacylation of tRNAs by *Escherichia coli* methionyl-tRNA synthetase (MetRS) and have located part of the anticodon binding domain on the enzyme at a site close to Trp461 [Schulman, L. H., & Pelka, H. (1988) *Science 242*, 765–768; Ghosh, G., Pelka, H., & Schulman, L. H. (1990) *Biochemistry 29*, 2220–2225]. In order to gain information about other possible sites of contact between MetRS and its tRNA substrates, we have examined the effects of mutations at a series of positively charged residues on the surface of the C-terminal domain of the enzyme. Conversion of Arg356, Arg366, Arg380, or Arg453 to Gln had little or no effect on enzyme activity. Similarly, conversion of Lys402 or Lys439 to Asn failed to significantly alter aminoacylation activity. Conversion of Arg380 to Ala or Arg442 to Gln produced a 5-fold reduction in $k_{cat}/K_m$ for aminoacylation of $tRNA^{fMet}$, with no effect on methionine activation, indicating a possible minor role for these residues in interaction of the enzyme with the tRNA substrate. In contrast, mutation of a phylogenetically conserved residue, Arg395, to Gln increased the $K_m$ for aminoacylation of $tRNA^{fMet}$ about 30-fold and reduced $k_{cat}/K_m$ by 25 000-fold. The mutant enzyme was also shown to be highly defective by its inability to complement a strain of *E. coli* having an altered chromosomal MetRS gene. Examination of the kinetic parameters for ATP–$PP_i$ exchange catalyzed by the Gln395 enzyme showed little or no effect of the mutation on interaction of MetRS with methionine or ATP, indicating that the major role of Arg395 is in tRNA recognition. Mutation of a second conserved residue, Asn391, to Ala specifically increased the $K_m$ for aminoacylation of $tRNA^{fMet}$ 20-fold with little or no effect on the other kinetic parameters. Examination of the crystal structure of MetRS [Brunie, S., Zelwer, C., & Risler, J.-L. (1990) *J. Mol. Biol. 216*, 411–424] shows that Trp461 is located on a separate peptide, at a distance of 10–20 Å from Asn391 and Arg395. Extensive molecular dynamics simulation studies have revealed that the peptide loop containing Trp461 is the most flexible part of MetRS and that residues in this loop can move more than 5 Å with little energy cost. This suggests that tRNA binding may induce a conformational change in MetRS that allows simultaneous interaction of the anticodon loop of the tRNA with the two peptides containing Trp461 and Asn391–Arg395. Such a substrate-induced conformational change may be an important feature of the tRNA recognition mechanism.

Aminoacyl-tRNA synthetases catalyze the ATP-dependent activation of amino acids and their subsequent transfer to the 3′ terminus of appropriate cognate tRNAs. The mechanism of differentiation between cognate and noncognate tRNAs depends on sequence-specific contacts between protein–tRNA pairs and optimal fit of the surfaces of the two macromolecules. The overall accuracy of protein synthesis is dependent on the outcome of this selection process.

The details of only one synthetase–tRNA interaction are available to date. A high-resolution X-ray crystal structure of *Escherichia coli* glutaminyl-tRNA synthetase (GlnRS)[1] complexed with $tRNA^{Gln}$ and ATP has been solved at 2.5-Å

† Supported by Research Grant GM16995 from the National Institutes of Health to L.H.S. Partial salary support for L.H.S. was provided by National Cancer Institute Grant 5P30CA13330. Deoxyoligonucleotides were synthesized by a Core Facility of the Albert Einstein College of Medicine supported in part by National Cancer Institute Grant 5P30CA13330.
‡ Albert Einstein College of Medicine.
§ Present address: Department of Molecular Biophysics and Biochemistry and Howard Hughes Medical Institute, Yale University, New Haven, CT 06511.
∥ Ecole Polytechnique.
⊥ Present address: LPCB, Institut Curie et Université Paris VI, 11 rue Pierre et Marie Curie, 75231 Paris Cedex 05, France.

[1] Abbreviations: GlnRS, glutaminyl-tRNA synthetase; MetRS, methionyl-tRNA synthetase; MetRS547, monomeric MetRS truncated at amino acid residue 547; R395Q, MetRS547 with arginine-395 replaced by glutamine; other MetRS547 mutants are indicated in a similar fashion using the one-letter amino acid code; TyrRS, tyrosyl-tRNA synthetase; $tRNA^{fMet}$, initiator methionine tRNA; Met-AMP, methionyladenylate; MD, molecular dynamics.

0006-2960/91/0430-11767$02.50/0 