The Valyl-tRNA Synthetase from *Bacillus stearothermophilus* Has Considerable Sequence Homology with the Isoleucyl-tRNA Synthetase from *Escherichia coli*†

Thor J. Borgford, Nigel J. Brand, Tamara E. Gray, and Alan R. Fersht*
Department of Chemistry, Imperial College of Science and Technology, London SW7 2AY, U.K.
Received October 2, 1986; Revised Manuscript Received December 18, 1986

ABSTRACT: We report the DNA sequence of the *valS* gene from *Bacillus stearothermophilus* and the predicted amino acid sequence of the valyl-tRNA synthetase encoded by the gene. The predicted primary structure is for a protein of 880 amino acids with a molecular mass of 102 036. The molecular mass and amino acid composition of the expressed enzyme are in close agreement with those values deduced from the DNA sequence. Comparison of the predicted protein sequence with known protein sequences revealed a considerable homology with the isoleucyl-tRNA synthetase of *Escherichia coli*. The two enzymes are identical in some 20–25% of their amino acid residues, and the homology is distributed approximately evenly from N-terminus to C-terminus. There are several regions which are highly conservative between the valyl- and isoleucyl-tRNA synthetases. In one of these regions, 15 of 20 amino acids are identical, and in another, 10 of 14 are identical. The valyl-tRNA synthetase also contains a region HLGH (His-Leu-Gly-His) near its N-terminus equivalent to the consensus HIGH (His-Ile-Gly-His) sequence known to participate in the binding of ATP in the tyrosyl-tRNA synthetase. This is the first example of extensive homology found between two different aminoacyl-tRNA synthetases.

The aminoacyl-tRNA synthetases are a "family" of enzymes responsible for the charging of specific tRNAs with their cognate amino acids. However, unlike many other families of enzymes that catalyze a common reaction, they are extremely diverse structurally, varying grossly in size, primary sequence, and subunit composition (Schimmel, 1979). For example, phenylalanyl-tRNA and glycyl-tRNA synthetases are $\alpha_2\beta_2$ tetramers of M_r 270000 (Bartmann et al., 1975; Fayat et al., 1983; Ostrem & Berg, 1974; Webster et al., 1983); tyrosyl-tRNA synthetase is an α_2 dimer of M_r 90000 (Winter et al., 1983); cysteinyl-tRNA synthetase is a monomer of M_r 54000 (Bruton & Cox, 1979); valyl-, leucyl-, and isoleucyl-tRNA synthetases are monomers of M_r 100000/110000 (Schimmel, 1979).

To date, the only significant sequence homology between several aminoacyl-tRNA synthetases consists of a sequence of four amino acids which have now been identified as participating in binding ATP (Jones et al., 1986; Leatherbarrow et al., 1985). Perhaps, the most likely candidates for more extensive homology would be the valyl-, leucyl-, and isoleucyl-tRNA synthetases because they are of similar size and all three possess a common second function—an editing mechanism for correcting errors of misactivation (Fersht, 1986). The sequence of the isoleucyl-tRNA synthetase from *Escherichia coli* has been determined (Webster et al., 1984), but there are no reports of other members of this class of ligase. We report herein the sequencing of the valyl-tRNA synthetase from *Bacillus stearothermophilus* and the identification of a considerable sequence homology between this enzyme and the isoleucyl-tRNA synthetase.

Experimental Procedures

Enzymes and Reagents. All restriction endonucleases, T4 DNA ligase, and the Klenow fragment of DNA polymerase I were obtained from Cambridge Biotechnology Limited. Calf intestinal phosphatase (CIP) was purchased from Boehringer. All oligonucleotides used in sequencing were synthesized on an Applied Biosystems Model 380B DNA synthesizer. *E. coli* 236c was a gift from Dr. K. Bohman, Department of Molecular Biology, Biomedicum, Uppsala, Sweden. [α-^{32}P]dATP (specific activity >400 Ci/mmol) was purchased from Amersham U.K.

Subcloning of the valS Gene. As reported elsewhere (Brand & Fersht, 1986), the *valS* gene was originally cloned as a 10.2-kilobase (10.2-kb) insert of the plasmid vector pAT153 (Twigg & Sherratt, 1980). The plasmid bearing the *valS* gene was designated pNB1. Subclones of pNB1 were produced by the "shotgun" cloning of restriction fragments into the plasmid vector pUC9 (Vieira & Messing, 1982). Conditions for the partial digest of the plasmid pNB1 with the restriction endonuclease *Sau*3AI were established empirically. Approximately 25 μg of pNB1 was digested to produce fragments roughly within the size range of 2.0–6.0 kb. The partially digested DNA was size fractionated by electrophoresis on a 0.8% agarose gel, and fragments in the size range 3.0–4.5 kb were isolated by electroelution into a trough which had been cut in the gel (Maniatis et al., 1982). After extraction with phenol/chloroform and precipitation in ethanol, the 3.0–4.5-kb fragments were ligated into pUC9 which had been cut with *Bam*HI and treated with calf intestinal phosphatase. Recombinant molecules containing the *valS* gene were selected for by the complementation of a temperature-sensitive defect in an *Escherichia coli* host cell line 236c, as before (Brand & Fersht, 1986). Defined restriction fragments of one plasmid subclone, pTB8, were themselves subcloned and used in complementation analysis with *E. coli* 236c.

Sequencing of the valS Gene. A 3.6-kb *Pst*I fragment of the subclone pTB8 was sequenced (Bankier & Barrell, 1983) by using the dideoxy chain termination method of Sanger et al. (1977).

Protein Purification. Crude lysates of pTB8-infected *E. coli* DH1 (Low, 1968) were prepared, heat treated, and purified essentially as described for the purification of the tyrosyl-tRNA synthetase (Wilkinson et al., 1983). An additional anion-exchange step was performed as described for tyrosyl-tRNA synthetase (Lowe et al., 1985) on a Pharmacia fast-protein liquid chromatograph (FPLC). Enzyme assays

† This work was funded by the MRC of the U.K.

0006-2960/87/0426-2480$01.50/0

Table I: Codon Usage of Aminoacyl-tRNA Synthetase Genes of *B. stearothermophilus*[a]

	U			C			A			G			
	Trp	Tyr	Val	Trp	Tyr	Val	Trp	Tyr	Val	Trp	Tyr	Val	
U	F 5	12	16	S 0	3	1	Y 8	4	9	C 0	0	0	U
	F 3	7	15	S 3	0	3	Y 6	10	25	C 3	2	4	C
	L 0	3	5	S 1	2	2	Term			Term			A
	L 7	12	21	S 7	6	14	Term			W 3	6	27	G
C	L 6	10	18	P 1	0	2	H 3	3	13	R 0	3	5	U
	L 5	10	23	P 1	0	3	H 4	4	11	R 7	19	33	C
	L 1	1	0	P 3	2	2	Q 12	12	16	R 2	1	2	A
	L 9	7	9	P 9	9	39	Q 6	5	9	R 11	5	17	G
A	I 11	13	23	T 0	0	0	N 2	4	5	S 0	1	1	U
	I 16	14	28	T 6	4	14	N 7	8	24	S 5	7	9	C
	I 0	1	0	T 1	2	5	K 17	18	40	R 0	0	0	A
	M 10	6	27	T 10	23	29	K 5	8	19	R 0	1	0	G
G	V 4	4	7	A 4	3	4	D 3	8	21	G 1	0	6	U
	V 12	10	27	A 10	9	26	D 13	15	34	G 11	19	29	C
	V 0	1	1	A 4	3	3	E 15	17	60	G 3	4	4	A
	V 13	5	23	A 10	19	28	E 14	22	24	G 5	13	15	G

[a] The first nucleotide of each codon appears in large type at the left of the table, the second nucleotide appears in large type at the top of the table, and the third nucleotide appears to the right of the table. Corresponding amino acids are indicated by a single-letter code. The column headings Trp, Tyr, and Val refer to the tryptophanyl-, tyrosyl-, and valyl-tRNA synthetases, respectively.

were performed as described previously (Fersht & Kaethner, 1976). Purification yielded a homogeneous protein preparation as determined by sodium dodecyl sulfate (SDS)–polyacrylamide gel electrophoresis. Protein concentrations were estimated from the total nitrogen content (Jaenicke, 1974) assuming that valyl-tRNA synthetase is 17% by weight nitrogen, and determinations were independently corroborated by the Folin phenol method (Lowry et al., 1951) with bovine serum albumin as standard.

Amino Acid Composition of Valyl-tRNA Synthetase. The relative amino acid composition of the valyl-tRNA synthetase was determined from the hydrolysates of known amounts of protein. Samples of valyl-tRNA synthetase (60 pmol) were hydrolyzed at 110 °C in 6N HCl in vacuo for 22, 46, 70, 93, and 111 h. Analysis were subsequently performed on a Beckman 121MB amino acid analyzer.

Results

Subcloning and Sequencing of valS Gene. The subcloning of pNB1 produced numerous plasmids capable of complementing the temperature-sensitive defect in *E. coli* 236c. Of these, 12 were characterized by restriction mapping, and 1 subclone, having a 4.5-kb insert, was chosen for further analysis. The restriction map of this plasmid, pTB8, appears in Figure 1. It was found by the complementation analysis of restriction fragments subcloned into various plasmid vectors that the structural portion of the *valS* gene was confined to a 3.6-kb *Pst*I fragment of pTB8 (Figure 1). Accordingly, only this 3.6-kb *Pst*I fragment was sequenced initially.

Greater than 90% of the 3.6-kb *Pst*I fragment was sequenced in both directions by overlapping individual sequences derived from fragments shotgun cloned into M13mp8 and M13mp9 (Messing & Vieira, 1982) (Figure 2). Sequences were aligned by computer employing the DBUTIL sequencing program (Staden, 1980). In addition, the entire *Pst*I fragment was cloned into M13mp8 in both orientations. A bank of 12 sequencing primers was generated and used to complete the sequencing of both strands and confirm the sequence of the structural portion of the *valS* gene in the intact fragment. The structural portion of the gene was then resequenced in its entirety with dITP replacing dGTP to minimize sequence compressions (Mills & Kramer, 1979). A large open reading frame was found, preceded by a Shine–Dalgarno sequence characteristic of *Bacillus* species (Band & Henner, 1984) and followed by a palindromic ρ-independent termination sequence (Rosenberg & Court, 1979) (Figure 3). Translation of this sequence would produce a protein which is 880 amino acids in length with M_r 102036. The calculated size of this protein is in good agreement with published values estimated from SDS–polyacrylamide electrophoresis and gel filtration of the valyl-tRNA synthetase against known standards (Koch et al., 1974). Sequences characteristic of RNA polymerase binding sites were not evident in the region 5′ to the initiation codon, up to the *Pst*I site, for which sequence was available. Subsequently, a sequence characteristic of an RNA polymerase binding site was found when the region immediately upstream of the *Pst*I site was sequenced (Figure 3). Hence, transcription of this gene begins 432 base pairs (bp) 5′ to the initiation codon.

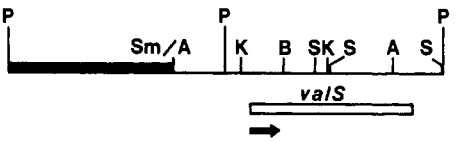

FIGURE 1: Restriction map of the plasmid pTB8. The bold line represents the vector portion of the plasmid (pUC9). The restriction sites illustrated are P (*Pst*I), S (*Sal*I), B (*Bam*HI), K (*Kpn*I), Sm (*Sma*I), and A (*Ava*I). The boxed region beneath the map shows the extent of the structural portion of the *valS* gene, the arrow indicating the direction of transcription.

Codon Usage of valS Gene. The codon usage of the *valS* gene is very similar to that of two other aminoacyl-tRNA synthetases of *B. stearothermophilus* (Table I) (Winter et al., 1983; Barstow et al., 1986) and is also typical of highly expressed genes of *E. coli* (Grosjean & Fiers, 1982).

Expression of valS Gene. SDS–polyacrylamide gel electrophoresis was performed on the heat-treated crude extracts of *E. coli* DH1 cultures which were both uninfected and infected with pTB8. High levels of expression of a thermostable protein having a molecular mass of approximately 100000

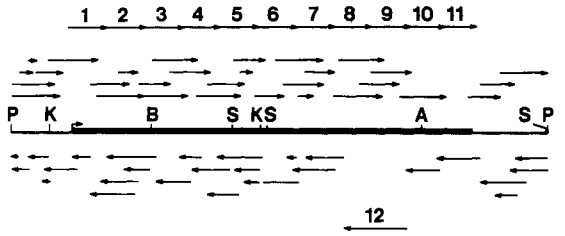

FIGURE 2: Sequencing strategy for the 3.6-kb *Pst*I fragment of pTB8. The unbroken line represents the *Pst*I fragment, and the bold segment of this line represents the structural region of the *valS* gene. Unnumbered arrows indicate the portion of the sequence derived from individual shotgun clones. Numbered arrors indicate the portion of the sequence derived from primers to the intact *Pst*I fragment. Arrows above the line are sequences from 5′ to 3′, and arrows below the line are sequences from 3′ to 5′ with respect to the gene.

were evident only in pTB8-infected cells (data not shown). The presence of valyl-tRNA synthetase in the crude extracts was assessed by enzyme assay (the ability to aminoacylate $tRNA^{Val}$), and it was found that only pTB8-infected cells expressed a thermostable valyl-tRNA synthetase. Ultimately, it was possible to purify 40 mg of valyl-tRNA synthetase per liter of cell culture of pTB8-infected DH1. However, since the promoter and structural portion of the *valS* gene are separated by a *Pst*I site, it was not possible to achieve such high levels of expression from plasmid vectors carrying the 3.6-kb *Pst*I fragment alone (Figures 1 and 3).

An extinction coefficient for the valyl-tRNA synthetase of 2.1 mg^{-1} mL^{-1} cm^{-1} was determined directly from the absorbance at 280 nm of a known concentration of protein as determined by nitrogen content. This value is in very close agreement with the extinction coefficient of 1.85 mg^{-1} mL^{-1} cm^{-1} based on the amino acid composition deduced from the DNA sequence of *valS*. The calculated extinction coefficient assumes a contribution of 5480 M^{-1} cm^{-1} for tryptophan, 1180 M^{-1} cm^{-1} for tyrosine, 120 M^{-1} cm^{-1} for cysteine (Mulvey et al., 1974), and a molecular mass of 102 036.

The relative tryptophan composition of the valyl-tRNA synthetase was also determined by a spectroscopic method (Edelhoch, 1967) according to

$$N_{Trp} = \epsilon_{288}/3103 - \epsilon_{280}/10318 \quad (1)$$

where N_{Trp} is the number of moles of tryptophan per mole of protein and the molecular mass of the protein is assumed to be 102 036. This method yielded a value of 26.7 tryptophans per molecule whereas the predicted value is 27.

Amino Acid Content of Valyl-tRNA Synthetase. The amino acid compositions of the valyl-tRNA synthetase deduced from the DNA sequence of *valS* and from hydrolysis of valyl-tRNA synthetase protein are presented in Table II.

Homologies with Known DNA and Protein Sequences. Homologies were sought for the DNA sequence of the *valS* gene with other genes and the derived protein sequence of the valyl-tRNA synthetase with other proteins. Data bases made available through the ICRF (Imperial Cancer Research Fund) were searched, and a significant homology was found between the valyl-tRNA synthetase and the isoleucyl-tRNA synthetase of *E. coli* (Webster et al., 1984).

A matrix analysis of the homology between the two enzymes was constructed by the DIAGON program (Staden, 1982, 1986) (Figure 4). Most of the identities between the two enzymes fall along roughly the same diagonal and are distributed evenly from N-terminus to C-terminus. The "best" alignments of the two enzymes (Figure 5) would indicate that they are identical in 20–25% of their amino acid residues. Variation in this estimate results from differences in the number of gaps allowed into one or the other of the two sequences during alignment.

Table II: Amino Acid Composition of Valyl-tRNA Synthetase[a]

amino acid	residues per molecule: from DNA sequences	from hydrolysates
Ala	61	66
Arg	57	52
Asn	29	89
Asp	55	
Cys	4	nd[b]
Gln	25	121
Glu	84	
Gly	54	57
His	24	25
Ile	51	38
Leu	76	67
Lys	59	64
Met	27	20
Phe	31	32
Pro	46	53
Ser	30	25
Thr	48	45
Trp	27	27*
Tyr	34	34
Val	58	63

[a] Tryptophans determined spectrophotometrically as described under Experimental Procedures. The relative amino acid composition is presented assuming that the valyl-tRNA synthetase is 880 residues in length. The amino acid composition determined from hydrolysates sums to 878 due to rounding. [b] Not determined.

There are several regions of the valyl- and isoleucyl-tRNA synthetases which are extremely homologous. The two most evident of these are found roughly midway between the N- and C-termini of the two enzymes in Figure 5. The region of the valyl-tRNA synthetase from residues 520 to 539 is identical in 15 out of 20 positions with the region extending from residues 597 to 616 of the isoleucyl-tRNA synthetase. Similarly, the region from 402 to 416 of valyl-tRNA synthetase is identical in 10 out of 14 residues with the region 458–471 of the isoleucyl-tRNA synthetase. In addition to these strong homologies, the valyl-tRNA synthetase also contains a region homologous with the so-called HIGH (His-Ile-Gly-His) region described by several authors (Webster et al., 1984; Jones et al., 1986;, Myers & Tzagoloff, 1985).

A number of homologies of doubtful significance were found between the sequence of the *valS* gene and DNA sequences in the ICRF data bases. However, homology was found between a region toward the 3′ end of the gene encoding *B. stearothermophilus* tyrosyl-tRNA synthetase and a region toward the 3′ end of the *valS* gene (Figure 6). Of 21 bases, 18 are identical between the 2 genes in these regions. The amino acid sequence encoded by the *valS* gene at this point (corresponding to residues 522–528) is in the center of the region of greatest homology with the isoleucyl-tRNA synthetase. Yet, the gene encoding the tyrosyl-tRNA synthetase is in a different reading frame at this point and, therefore, does not encode the same amino acid sequence. There is independent confirmation of the protein sequence in both the tyrosyl-tRNA synthetase (Winter et al., 1983) and the isoleucyl-tRNA synthetase (Webster et al., 1984). Therefore, the difference in reading frames is not likely to be due to an error in sequencing. Outside of this region of homology, the DNA sequences diverge considerably.

DISCUSSION

The DNA sequence of the *valS* gene was initially determined by shotgun sequencing (Bankier & Barrell, 1983). A routine and facile method of sequencing this gene is required, however, because we intend to conduct an extensive dissection

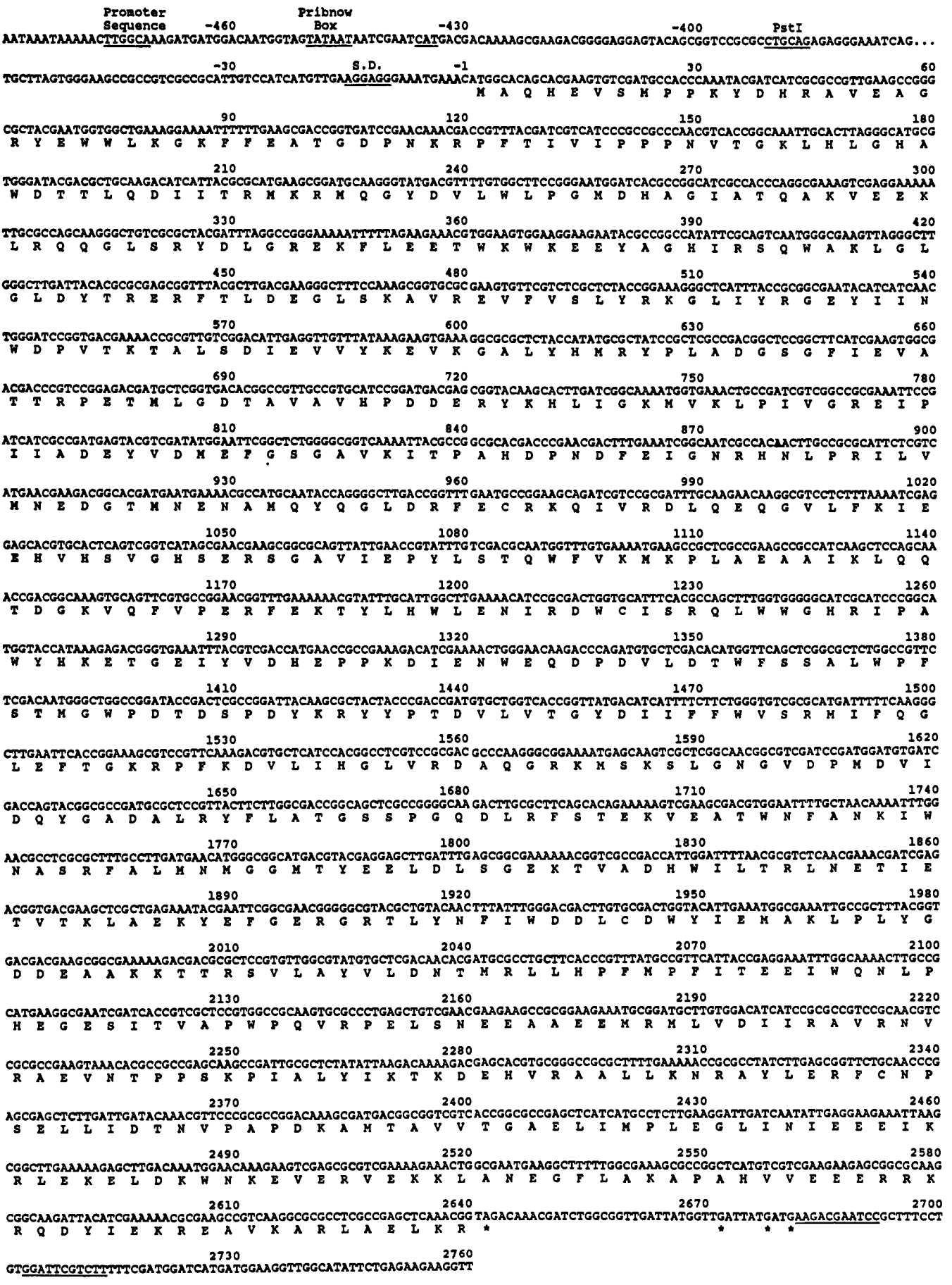

FIGURE 3: DNA sequence of the *valS* gene and corresponding amino acid sequence. Underlined segments between −490 and −430 indicate putative promoter regions. The underlined region beginning at −16 is the Shine–Dalgarno (S.D.) ribosome binding site. Termination codons are indicated by asterisks. The underlined segments between 2670 and 2730 are complementary sequences indicative of the stem structure of a ρ-independent terminator.

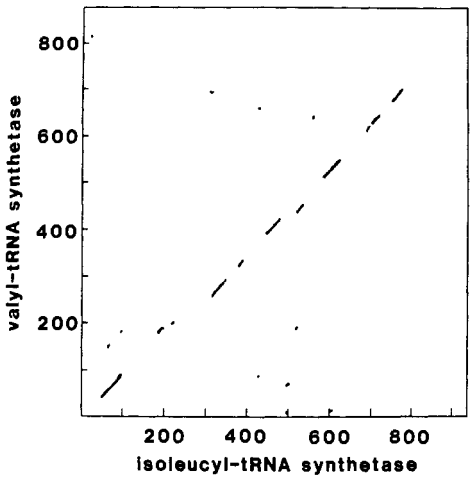

FIGURE 4: Matrix analysis of the homology found between the valyl-tRNA synthetase (y axis) and isoleucyl-tRNA synthetase (x axis). The matrix was generated by computer employing the DIAGON program where the program parameters are span length = 25 and identities = 8.

of the structure and activity of the *valS* gene product, by site-directed mutagenesis. Sequencing primers were constructed, designed to prime at intervals of 200–250 bases, spanning the entire structural portion of the gene. Consequently, we have been able to sequence the *valS* gene repeatedly, albeit only in the 5′ to 3′ direction. Primers have also made it possible for us to sequence the entire structural portion of the gene with ITP in place of GTP, largely eliminating sequence compressions.

The codon usage of the *valS* gene is presented in Table I along with the codon usage of two other aminoacyl-tRNA synthetase genes of *B. stearothermophilus*. There appear to be no inconsistencies in the codon usage of the three genes. All three genes commonly avoid the codons ACT (Thr), UGU (Cys), and AGA (Arg), and all three are typical of highly expressed genes of *E. coli* (Grosjean & Fiers, 1982). This is in keeping with the observation of high expression of *B. stearothermophilus* valyl-tRNA synthetase of pTB8-infected *E. coli*.

The amino acid composition deduced from translation of the DNA sequence of the *valS* gene agrees well with the amino acid composition of the valyl-tRNA synthetase (Table II). The observed and expected extinction coefficients for the valyl-tRNA synthetase, 2.1 and 1.85 mg^{-1} mL^{-1} cm^{-1}, respectively, are in very close agreement, as are the molecular masses. It is, therefore, unlikely that there are inaccuracies in the DNA sequence which have in turn led to an incorrect prediction of the protein sequence of the valyl-tRNA synthetase. Unfortunately, the N-terminus of the protein was found to be blocked, so precluding a simple determination of the N-terminal sequence. In a study based on the analysis of tryptic fragments, Koch et al. (1974) reported the existence of repeated sequences in the methionyl-, leucyl-, and valyl-tRNA synthetases of *B. stearothermophilus*. Our predicted primary structure of the valyl-tRNA synthetase shows no evidence of the repeating amino acid sequences. The previously reported sequence of the methionyl-tRNA synthetase of *E. coli* (Barker et al., 1982) is similarly without repeating sequences as is the isoleucyl-tRNA synthetase (Webster et al., 1984).

The amino acid sequences of some 11 of the 20 different aminoacyl-tRNA synthetases from a variety of sources have now been determined (Putney et al., 1981; Hoben et al., 1982; Winter et al., 1983; Fayat et al., 1983; Webster et al., 1983, 1984; Myers & Tzagoloff, 1985; Fasiolo et al., 1985; Pape et al., 1985; Freedman et al., 1985). Examination of the collected sequences indicates that there is very little homology of primary structure, between any two enzymes within this family. Further, the likelihood of finding strong homologies between different aminoacyl-tRNA synthetases is probably low because of their gross variations in molecular mass and subunit composition (Schimmel, 1979).

It is commonly believed that the valyl-, leucyl-, and isoleucyl-tRNA synthetases represent a subfamily of the aminoacyl-tRNA synthetases because they are monomeric, similar in molecular mass, and possess editing mechanisms. Additionally, the similarity of their amino acid substrates (short aliphatic side chains) suggests that evolution from a common progenator. Hence, the primary structures of these three enzymes may be expected to contain homologies better than average for the aminoacyl-tRNA synthetases.

The matrix analysis of homology between the valyl- and isoleucyl-tRNA synthetases is shown in Figure 4. Each point on the matrix represents the center of a region of identity or the center of a region of homology according to the scoring criteria of the DIAGON program (Staden, 1982, 1986). This method is based on the measured frequency of amino acid replacements in homologous proteins (Dayhoff, 1972). If these

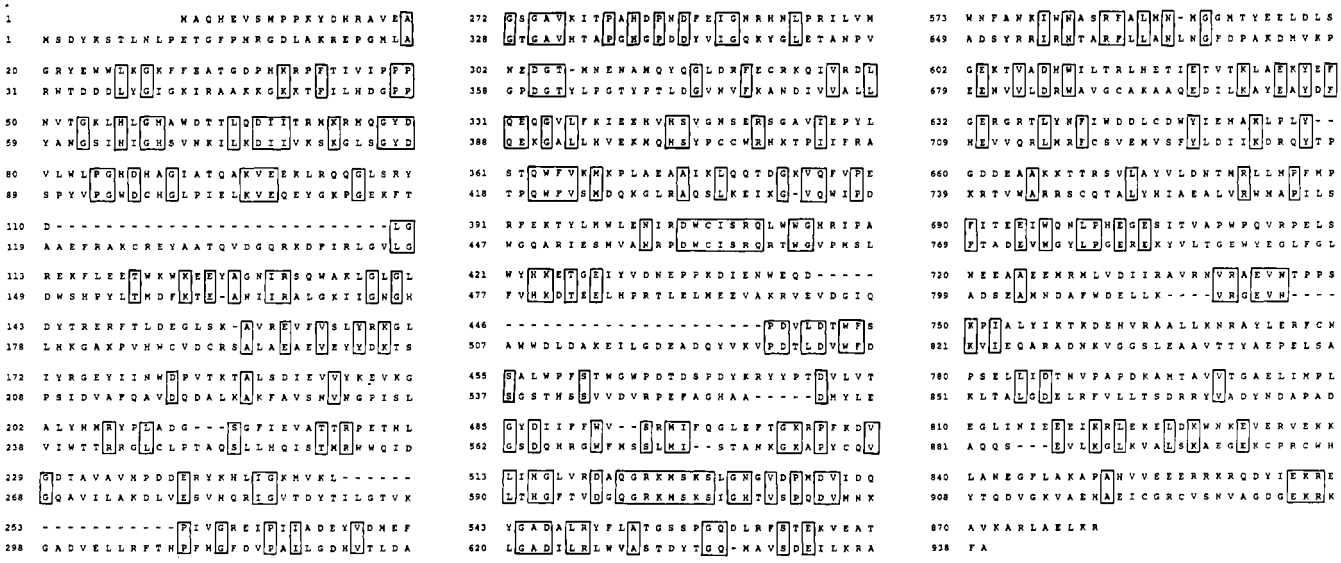

FIGURE 5: Alignment of the amino acid sequences of the valyl-tRNA synthetase (upper line) and isoleucyl-tRNA synthetase (lower line). Numbers at the left of each line indicate the position of the first amino acid in that line. Dashes indicate where gaps have been introduced.

```
          D  A  Q  G  R  K  M  S  K  S  L
1558  GACGCCCAAGGGCGGAAAATGAGCAAGTCGCTC  1590
        **  ********* *********
988     TTGACAGCGGCGGAAATTGAGCAAGGGTTCAAA  1020
          L  T  A  A  E  I  E  Q  G  F  K
```

FIGURE 6: Alignment of DNA and corresponding amino acid sequences of the *valS* gene (top) and the tyrosyl-tRNA synthetase gene (bottom) from *B. stearothermophilus*. Identities are indicated by asterisks. Numbers refer to the positions of the rightmost and leftmost nucleotides in each sequence.

two synthetases were not homologous, or if homology between the two was insignificant, then the distribution of diagonal lines in this plot would be random. However, they appear to share considerable homology, since the matrix plot produces what is essentially a single broken line along roughly the same diagonal. An alignment of the valyl- and isoleucyl-tRNA synthetase sequences also illustrates the homology between the two enzymes. Alignments are somewhat arbitrary, as the best agreement of sequences requires that gaps be introduced into one or the other of the two, and it is important to note that the isoleucyl-tRNA synthetase is 59 amino acids longer than the valyl-tRNA synthetase. Assuming that the number of gaps in either sequence is at a minimum and that the spacing of these gaps is not overly biased, alignment would indicate that they are identical in 20–25% of their residues. Although it is difficult to assign a precise value to this homology, for the reasons mentioned, it is possible to say unequivocally that the valyl- and isoleucyl-tRNA synthetases are more closely related than any other two aminoacyl-tRNA synthetases so far sequenced. This is particularly evident when comparing the sequence from positions 520–539 in the valyl-tRNA synthetase with the sequence from positions 597–616 in the isoleucyl-tRNA synthetase (Figure 5). In this section, the synthetases are identical in 15 out of 20 residues. Also, there are 10 out of 14 identities between the sequence of the valyl-tRNA synthetase extending from position 402 to 416 and the sequence of the isoleucyl-tRNA synthetase extending from position 458 to 471. Both of these regions demonstrate greater relatedness than sequences associated with the consensus HIGH region (Webster et al., 1984; Myers & Tzagoloff, 1985; Jones et al., 1986) found near the N-terminus of some aminoacyl-tRNA synthetases.

The HIGH (His-Ile-Gly-His) region is present in the tyrosyl-tRNA synthetase for which there is a refined crystal structure (Bhat et al., 1982; Blow & Brick, 1985), and it is known that this region represents a portion of the nucleotide binding fold in the tyrosyl-tRNA synthetase. An equivalence has been established between isoleucine and leucine at the second poisition of the HIGH sequence and an equivalence between histidine and asparagine at the fourth position (Jones et al., 1986). The valyl-tRNA synthetase contains the sequence HLGH beginning at position 56 which may therefore be considered to be equivalent to the consensus HIGH. Notably, the *E. coli* tyrosyl-tRNA synthetase contains the sequence HLGH.

Clearly, the primary structure of valyl-tRNA synthetase is very similar to that reported for the isoeleucyl-tRNA synthetase. It is noteworthy that this similarity exists even though these enzymes are derived from two entirely different genus of bacteria. The aminoacyl-tRNA synthetases of *B. stearothermophilus* are very similar in many aspects to their counterparts in *E. coli* apart from their thermostability. The tyrosyl-tRNA synthetases from *E. coli* and *B. stearothermophilus*, for example, are kinetically very similar enzymes. Yet, they are only 56% homologous in primary structure, emphasizing the significance of a 20–25% homology between the valyl- and isoleucyl-tRNA synthetases. A correlation has been demonstrated between the similarity of structure of protein core regions and the extent of their primary sequence homology (Chothia & Lesk, 1986). Although this correlation was derived from examining the sequence homology of *known* structures, it does suggest that highly homologous regions of proteins (>50% of residues identical) share a common general structure. The region of the valyl-tRNA synthetase from position 512 to 560 is 55% identical with the region of the isoleucyl-tRNA synthetase from position 589 to 637.

Chemically modified tRNAs have been used to identify regions in the methionyl-tRNA synthetase which are close to the 3′ end of the enzyme-bound $tRNA^{Met}$. By analogy, a similar region was identified in the isoleucyl-tRNA synthetase extending from position 599 to 609 (Hountondji et al., 1986). This region is precisely the one which shows the greatest homology with the valyl-tRNA synthetases. These findings suggest that this highly conserved region is close to the active site in the native enzymes. The function is unknown.

In an extensive study, Schimmel and co-workers have demonstrated a modular arrangement of "functional domains" in the tetrameric alanyl-tRNA synthetase (Jasin et al., 1983, 1985). Deletion mutagenesis indicated that a large portion of the C-terminal end of the protein is dispensible to catalysis and this noncatalytic portion of the molecule is responsible solely for oligomerization of the subunits. Deletion mutagenesis of the tyrosyl-tRNA synthetase has showed that the C-terminal domain of this enzyme is responsible primarily for binding $tRNA^{Tyr}$ and, similar to the alanyl-tRNA synthetase, catalytic functions are confined largely to a region close to the N-terminus (Waye et al., 1983). The consensus HIGH region associated with ATP binding has always been found in the N-terminal domain of those synthetases that possess it. As the highly conserved regions of the valyl- and isoleucyl-tRNA synthetases are located midway and toward the C-terminus, it is quite possible that they are involved in a function other than the formation of aminoacyl adenylates or the charging of tRNAs—possibly in the second sites responsible for editing errors of misactivation.

It is noteworthy that the gene sequence of the valyl-tRNA synthetase corresponding to the highly conserved region between positions 522 and 529 is itself highly homologous with a region of the gene encoding tyrosyl-tRNA synthetase (Figure 5) (Winter et al., 1983). However, the significance of this coincidence of homologies is questionable given that these genes are in two different reading frames at this position.

Subsequent to the completion of the sequence of the valyl-tRNA synthetase gene from *B. stearothermophilus*, M. Härtlein, D. Madern, and R. Leberman communicated to us the (unpublished) sequence of the corresponding gene from *E. coli*. A preliminary comparison of the protein sequences deduced from the two genes indicated that they are in excess of 40% identical. The regions of greatest homology with the isoleucyl-tRNA synthetase reported here are also evident in the valyl-tRNA synthetase from *E. coli*. It will now be possible to probe the structure and activity of the valyl-tRNA synthetase by site-directed mutagenesis, especially the functions of the highly conserved regions.

REFERENCES

Band, L., & Henner, D. J. (1984) *DNA 3*, 17–21.

Bankier, A. T., & Barrell, B. G. (1983) in *Techniques in the Life Sciences*, Vol. B5, Elsevier, Ireland.

Barker, D. G., Ebel, J.-P., Jakes, R., & Bruton, C. J. (1982) *Eur. J. Biochem. 127*, 449–457.

Barstow, D. A., Sharman, A. F., Atkinson, T., & Minton, N. P. (1986) *Gene 46*, 37–45.

Bartmann, P., Hanke, T., & Holler, E. (1975) *J. Biol. Chem. 250*, 7668–7674.
Bhat, T. N., Blow, D. M., Brick, P., & Nyborg, J. (1982) *J. Mol. Biol. 158*, 699–709.
Blow, D. M., & Brick, P. (1985) in *Biological Macromolecules and Assemblies–Nucleic Acids and Interactive Proteins* (Jurnak, F., & McPherson, A., Eds.) Vol. 2, pp 442–469, Wiley, New York.
Brand, N., & Fersht, A. R. (1986) *Gene 44*, 139–142.
Bruton, C. J., & Cox, L. A.-M. (1979) *Eur. J. Biochem. 100*, 301–307.
Chothia, C., & Lesk, A. M. (1986) *EMBO J. 5*, 823–826.
Dayhoff, M. D. (1972) in *Atlas of Protein Sequence and Structure 1972*, Vol. 5, National Biomedical Research Foundation, Washington, D.C.
Edelhoch, H. (1967) *Biochemistry 6*, 1948–1954.
Fasiolo, F., Gibson, B. W., Walter, D., Chatton, B., Biemann, K., & Boulanger, Y. (1985) *J. Biol. Chem. 260*, 15571–15576.
Fayat, G., Mayaux, J.-F., Sacerdot, C., Fromant, M., Springer, M., Grunberg-Manago, M., & Blanquet, S. (1983) *J. Mol. Biol. 171*, 239–261.
Fersht, A. R. (1986) in *Accuracy in Molecular Processes* (Kirkwood, T. B. L., Rosenberger, R. F., & Galas, D. J., Eds.) Chapter 4, Chapman and Hall, London and New York.
Fersht, A. R., & Kaethner, M. M. (1976) *Biochemistry 15*, 3342–3346.
Freedman, R., Gibson, B., Donovan, D., Biemann, K., Eisenbeis, S., Parker, J., & Schimmel, P. (1985) *J. Biol. Chem. 260*, 10063–10068.
Grosjean, H., & Fiers, W. (1982) *Gene 18*, 199–209.
Hoben, P., Royal, N., Cheung, A., Yamao, F., Biemann, K., & Soll, D. (1982) *J. Biol. Chem. 257*, 11644–11650.
Hountondji, C., Lederer, F., Dessen, P., & Blanquet, S. (1986) *Biochemistry 25*, 16–21.
Jaenicke, L. (1974) *Anal. Biochem. 61*, 623–627.
Jasin, M., Regan, L., & Schimmel, P. (1983) *Nature (London) 306*, 441–447.
Jasin, M., Regan, L., & Schimmel, P. (1985) *J. Biol. Chem. 260*, 2226–2230.
Jones, M. D., Lowe, D. M., Borgford, T. J., & Fersht, A. R. (1986) *Biochemistry 25*, 1887–1897.
Koch, G. L. E., Boulanger, Y., & Hartley, B. S. (1974) *Nature (London) 249*, 316–320.
Leatherbarrow, R. J., Fersht, A. R., & Winter, G. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 7840–7844.
Low, B. (1968) *Proc. Natl. Acad. Sci. U.S.A. 60*, 160–167.
Lowe, D. M., Fersht, A. R., Wilkinson, A. J., Carter, P., & Winter, G. (1985) *Biochemistry 24*, 5106–5109.
Lowry, O. H., Rosebrough, N. J., Farr, A. L., & Randall, R. J. (1951) *J. Biol. Chem. 193*, 265–275.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) in *Molecular Cloning: a Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Messing, J., & Vieira, J. (1982) *Gene 19*, 269–276.
Mills, D. R., & Kramer, F. R. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 2232–2235.
Mulvey, R. S., Gualtieri, R. J., & Beychok, S. (1974) *Biochemistry 13*, 782–787.
Myers, A. M., & Tzagoloff, A. (1985) *J. Biol. Chem. 260*, 15371–15377.
Ostrem, D. L., & Berg, P. (1974) *Biochemistry 13*, 1338–1348.
Pape, L. K., Koerner, T. J., & Tzagoloff, A. (1985) *J. Biol. Chem. 260*, 15362–15370.
Putney, S. D., Royal, N. J., Neuman de Vegvar, H., Herlihy, W. C., Biemann, K., & Schimmel, P. (1981) *Science (Washington, D.C.) 213*, 1497–1501.
Rosenberg, M., & Court, D. (1979) *Annu. Rev. Genet. 13*, 319–353.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 741*, 5463–5467.
Schimmel, P. R. (1979) *Annu. Rev. Biochem. 48*, 601–648.
Staden, R. (1980) *Nucleic Acids Res. 8*, 3673–3694.
Staden, R. (1982) *Nucleic Acids Res. 10*, 2951–2961.
Staden, R. (1986) *Nucleic Acids Res. 14*, 217–231.
Twigg, A. J., & Sherratt, D. (1980) *Nature (London) 283*, 216–218.
Waye, M. M. Y., Winter, G., Wilkinson, A. J., & Fersht, A. R. (1983) *EMBO J. 2*, 1827–1829.
Webster, T. A., Gibson, B. W., Keng, T., Biemann, K., & Schimmel, P. (1983) *J. Biol. Chem. 258*, 10637–10641.
Webster, T., Tsai, H., Kula, M., Mackie, G. A., & Schimmel, P. (1984) *Science (Washington, D.C.) 226*, 1315–1317.
Wilkinson, A. J., Fersht, A. R., Blow, D. M., & Winter, G. (1983) *Biochemistry 22*, 3581–3586.
Winter, G., Koch, G. L. E., Hartley, B. S., & Barker, D. G. (1983) *Eur. J. Biochem. 132*, 383–387.
Vieira, J., & Messing, J. (1982) *Gene 19*, 259–268.