J. Mol. Biol. (1982) **158**, 573–597

Correlation Between the Abundance of Yeast Transfer RNAs and the Occurrence of the Respective Codons in Protein Genes

Differences in Synonymous Codon Choice Patterns of Yeast and *Escherichia coli* with Reference to the Abundance of Isoaccepting Transfer RNAs

TOSHIMICHI IKEMURA

Department of Biophysics, Faculty of Science
Kyoto University, Kyoto 606, Japan

(Received 14 January 1982)

There exists a similarity among the synonymous codon choice patterns of the yeast nuclear genes that have been sequenced thus far although these genes encode different types of protein molecules, and the patterns are significantly different from those of *Escherichia coli* genes. Based on constraints caused by the availability of *E. coli* transfer RNAs and the nature of their codon recognition related to the modified nucleotides at the anticodon wobble position, the characteristic patterns of synonymous codon choice commonly found for *E. coli* genes have been almost completely explained (Ikemura, 1981*a*,*b*). In the present paper, tRNAs of the yeast *Saccharomyces cerevisiae* were separated by two-dimensional polyacrylamide gel electrophoresis and the relative abundance of purified tRNA molecules was measured on the basis of molecular numbers in cells. A strong correlation between tRNA abundance and codon choice was found for each nuclear gene of yeast, but the correlation was less significant for 2μ plasmid genes. According to the criteria proposed for *E. coli* genes (Ikemura, 1981*b*) the order of codon preference in yeast nuclear genes was predicted based on the abundance of yeast isoaccepting tRNAs and on the nature of the modified nucleotides at their anticodons. Clear correlations between predictions and the actual preferences among synonymous codons were revealed, indicating that the codon choices in yeast genes are also constrained by a combination of tRNA availability and nature of its codon recognition. Then the difference in synonymous codon use between the two organisms can be attributed to the difference in these two factors.

1. Introduction

The choice among synonymous codons in both prokaryotic and eukaryotic genes is distinctly non-random (Fiers *et al.*, 1975; Air *et al.*, 1976; Efstratiadis *et al.*, 1977; Post *et al.*, 1979). Using a multivariate method to analyse codon usage in many messenger RNA sequences, Grantham *et al.* (1980,1981) have revealed clear similarities of synonymous codon usage between different protein genes of the same

0022–2836/82/200573–25 $03.00/0

or taxonomically related genomes and dissimilarities between genes of taxonomically distant genomes.

It has become increasingly clear that the non-random choice among synonymous codons in *Escherichia coli* genes can mostly be attributed to the availability of transfer RNAs within a cell (Post *et al.*, 1979; Post & Nomura, 1980; Ikemura 1980,1981*a*,*b*). On the basis of constraints caused by the abundance of *E. coli* tRNAs as well as by the nature of their codon recognition, five rules have been proposed which explain almost thoroughly the characteristic patterns of synonymous codon choice found for the *E. coli* genes sequenced thus far (Ikemura, 1981*a*,*b*). In order to know the applicability of these rules to understand the non-random codon choices found in other organisms and in order to find the mechanisms that cause species-specific codon choice patterns, tRNAs of the yeast *Saccharomyces cerevisiae* were separated and quantified using two-dimensional polyacrylamide gel electrophoresis. There were clear correlations between the amounts of cytoplasmic tRNAs and the occurrence of respective codons in the nuclear genes. The characteristic codon choice patterns of yeast nuclear genes could be accounted for almost completely by the five rules proposed for *E. coli* genes.

2. Materials and Methods

(a) *Yeast strains, growth conditions and preparation of tRNA*

The strain of *S. cerevisiae* used in most experiments was X2180–1B (a *mal gal2 SUC2 CUP1*). In one separate experiment, D13–1A (*a his3–532 trp1 mal gal2 SUC2 CUP1*) was used and the results are discussed in section (a) of the Discussion. Low phosphate YEPD medium (1% (w/v) Bacto yeast extract, 2% (w/v) Bacto peptone, 2% (w/v) glucose) was prepared following the procedure of Rubin (1975). The yeast strain was grown aerobically at 30°C in the medium. When the cell density had reached approx. 5×10^6 ml, carrier-free [^{32}P]orthophosphate (NEX–053) was added to the culture at a concentration of 0·05 mCi/ml (the total volume of the culture was usually 20 ml). Culturing was continued until the cell density reached approximately 1×10^8 ml, and cells were harvested by centrifugation. The RNA sample was prepared by a modified form of the method of Knapp *et al.* (1978). Harvested cells were washed once with 20 ml 0·01 M-EDTA, 0·1 M-sodium acetate (pH 5·0), and collected by centrifugation. They were then suspended in 2 ml 1% (w/v) sodium dodecyl sulphate, 0·02 M-EDTA, 0·1 M-sodium acetate (pH 5·0), and extracted with an equal volume of phenol containing 40 μg carrier RNA at 55°C for 30 min. In a separate experiment, the phenol extraction was performed at room temperature for 1 h (see Discussion, section (a)). After centrifugation, RNA was precipitated from the aqueous phase with 3 vol. ethanol and washed 3 more times by precipitation with ethanol as described by Ikemura & Dahlberg (1973). The RNA sample was divided into several portions and kept in 75% (v/v) ethanol at −20°C until individually electrophoresed to quantify the tRNAs. In a few cases the RNA sample was purified after the first precipitation with ethanol using a small DEAE-cellulose column as described by Knapp *et al.* (1978). The sample that was extracted from a 20 ml culture and so purified was applied to a single 2-dimensional gel. This second procedure was used to obtain a sufficient quantity of minor tRNAs for fingerprinting of RNAase T_1 digests, followed by oligonucleotide analyses after RNAase A secondary digestion as described below.

(b) *Gel electrophoresis and the quantification of individual tRNAs*

In order to purify tRNAs, 2-dimensional polyacrylamide gel electrophoresis was performed in 3 different ways (subsequently referred to as types A, B and C). Types A and C

consisted of modifications of the procedures of Ikemura & Nomura (1977) and Ikemura *et al.* (1975), respectively; the gel length of the first dimension was enlarged. Electrophoresis on a polyacrylamide gel slab (42 cm long, 13 cm wide and 3 mm thick) for the first dimension was carried out in a vertical cell EC480 (EC Apparatus, U.S.A.). For the A and C systems, first dimension electrophoresis on a 10% (w/v) polyacrylamide gel (acrylamide/bisacrylamide in a 19 : 1 ratio) was at a constant current of 85 mA (about 35 V/cm) at 15°C for 3·5 h, according to the procedure of Ikemura *et al.* (1975). In the B system, the electrophoresis on a 14% polyacrylamide gel was done at a constant current of 75 mA (about 38 V/cm) for 8·5 h. During electrophoresis a small amount of a bromophenol blue/xylene cyanol FF mixture was applied at 1 h intervals to the gel slot on which the RNA sample had been loaded. The trail of dyes indicated after electrophoresis which gel portions contained RNA molecules, without the use of autoradiography.

Second dimension electrophoresis on a gel slab (17 cm long, 23 cm wide and 3 mm thick) was done in an EC490 cell. After the first run, a 20 cm portion from the bottom of the first gel containing 4 S to 5 S RNAs was cut out and placed at the top of the electrophoresis cell at right angles to the first dimension. In the A and B systems, a 22% polyacrylamide gel containing 7 M-urea was polymerized at 15°C (Ikemura & Nomura, 1977). After polymerization electrophoresis buffer (Ikemura *et al.*, 1975) was poured into buffer reservoirs of the cell, and an extra portion of 22% polyacrylamide gel lying above the first gel was removed. Second dimension electrophoresis was performed at 15°C at a constant voltage of 20 V/cm for 1 h and 45 V/cm for 15 h. In the C system, second dimension electrophoresis was on a 20% acrylamide gel at 15°C at a constant current of 80 mA for 30 min and then 150 mA (about 55 V/cm at equilibrium) for 4 h (Ikemura *et al.*, 1975). Since the 20% and 22% gels stuck very tightly to the Plexiglass of the apparatus, it was necessary after electrophoresis to inject air between the gel and the Plexiglass by a syringe fitted with a needle before the apparatus was dismantled. The gel was then autoradiographed on X-ray film (Kodak XS-1). Individual radioactive tRNA spots were usually 3 to 5 mm in diameter. A hole, 6 mm diam., was punched in the X-ray film at the position of each tRNA spot, and a gel disk (6 mm diam.) was cut out using a plastic straw according to the hole in the film. The Čérenkov radiation of ^{32}P was measured as described by Ikemura & Ozeki (1977).

(c) *RNA fingerprint analysis of tRNAs*

Two-dimensional fingerprints, after RNAase T_1 digestion of the RNA, were made according to the standard method developed by Sanger and his colleagues (Sanger *et al.*, 1965; Barrell, 1971). Fractionated oligonucleotides in the T_1 fingerprint were further analysed by pancreatic RNAase A digestion followed by electrophoresis on DEAE-cellulose paper at pH 3·5 as described by Adams *et al.* (1969).

3. Results

(a) *Comparison of codon usage between yeast and* E. coli *genes*

Codon usages of yeast genes are listed in Table 1. Two genes of glyceraldehyde-3-phosphate dehydrogenase (GAPD; Holland & Holland, 1979,1980), enolase (Holland *et al.*, 1981) and histone H2B (Wallis *et al.*, 1980) are treated as collective genes, and five mitochondrial genes sequenced are also treated as one gene (Mit.; Bonitz *et al.*, 1980). They correspond to most, if not all, of the yeast protein genes that have been sequenced thus far. More than 30 *E. coli* protein genes have been completely sequenced. In Table 1, the codon usage of 25 of them is listed. Genes with similar functions (e.g. ribosomal protein genes) or those belonging to the same operon are treated usually as collective genes. (When they were analysed

TABLE 1

Codon usage observed for yeast and E. coli *genes*

Amino acid	Codon	Codon usage in yeast genes						Codon usage in *E. coli* genes							
		GAPD	Enolase	Actin	Histone H2B	*CYC1*	Mit.†	*tuf* AB	*ompA*	r-pro.	*tsf*	*rpo* BD	*otp*	*thr* AB	*trp* ABCE
Leu	UUA	0	5	2	2	1	164	0	1	1	0	2	2	14	21
	UUG	41*	73*	19*	10*	5*	2	0	1	2	0	8	7	23	24
	CUU	0	0	2	0	1	(2)	2	0	4	0	11	17	10	16
	CUC	0	0	0	0	0	(0)	1	0	3	0	18	11	18	16
	CUA	1	0	2	0	1	(12)	0	0	0	0	1	1	3	9
	CUG	0	0	0	0	0	(0)	53*	21*	79*	16*	141*	81*	55*	96*
Arg	CGU	0	2	5	0	0	0	41*	10*	48*	5*	89*	46*	24*	32*
	CGC	0	0	0	0	0	0	5*	3*	26*	4*	46*	14*	23*	51*
	CGA	0	0	0	0	0	0	0	0	0	0	1	3	6	4
	CGG	0	0	0	0	0	0	0	0	0	0	0	2	12	3
	AGA	22*	26*	13*	11*	3*	28	0	0	1	0	0	0	0	2
	AGG	0	0	0	1	0	0	0	0	0	0	0	0	2	1
Pro	CCU	1	1	6	2	1	26	0	1	3	0	9	5	6	11
	CCC	0	0	0	0	0	1	0	0	0	0	0	2	6	12
	CCA	22	27	13	8	3	24	2	3	4	2	11	7	5	16
	CCG	0	0	0	0	0	0	38*	15*	36*	3*	55*	22*	26*	40*
Gln	CAA	11	18	14	7	2	22	0	2	9	0	15	14	14	26
	CAG	0	0	0	1	0	3	16*	15*	33*	10*	73*	51*	32*	51*
Lys	AAA	3	10	6	15	6	23	35*	15*	90*	23*	77*	52*	24*	53*
	AAG	49*	62*	12*	24*	10*	1	11	4	24	5	37	19	21	11
Ala	GCU	19*	96*	15*	22*	3*	47	24*	22*	93*	17*	30*	42*	18*	31*
	GCC	16*	17*	11*	11*	4*	2	2	1	10	0	19	37	44	67
	GCA	0	0	0	1	0	32	11*	11*	45*	14*	30*	39*	20*	34*
	GCG	0	0	0	1	0	1	17*	3*	28*	8*	49*	39*	43*	60*
Val	GUU	45*	32*	14*	9*	1*	38	46*	17*	54*	20*	55*	42*	27*	28*
	GUC	27*	37*	12*	3*	0*	1	1	1	6	0	21	16	22	20
	GUA	0	0	0	0	0	54	21*	7*	40*	5*	34*	20*	8*	15*
	GUG	0	0	0	0	2	3	6*	2*	16*	4*	34*	22*	32*	44*
Gly	GGU	49*	72*	29*	8*	8*	61	38*	24*	40*	9*	78*	47*	31*	37*
	GGC	0*	0*	0*	0*	2*	0	41*	11*	34*	15*	47*	35*	36*	50*
	GGA	0	0	0	0	0	15	0	0	0	0	0	2	12	10
	GGG	0	0	0	0	2	3	2	0	0	0	5	4	14	14

Ser	UCU	24*	28*	14*	20*	2*	23	14	4	18	4	32	18	11	13
	UCC	26*	33*	12*	15*	0*	0	6	8	18	4	38	23	17	16
	UCA	0	0	3	1	1	46	0	0	1	0	2	3	7	10
	UCG	0	0	0	0	1	0	0	0	1	0	5	8	11	18
	AGU	0	0	2	0	0	12	0	0	1	0	3	6	7	10
	AGC	0	0	0	1	0	0	1	4	12	2	23	11	16	23
Thr	ACU	22*	17*	13*	15*	3*	27	25*	10*	36*	7*	19*	9*	5*	11*
	ACC	25*	23*	8*	5*	3*	1	31*	12*	26*	4*	63*	29*	21*	35*
	ACA	0	0	0	3	2	34	3	1	3	0	3	3	5	8
	ACG	0	0	0	0	0	0	1	1	0	0	13	4	8	18
Ile	AUU	16*	24*	14*	7*	2*	108	6	1	13	5	29	25	34	47
	AUC	23*	19*	16*	9*	2*	21	52*	15*	51*	14*	98*	57*	26*	37*
	AUA	0	0	0	0	0	6	0	0	0	0	0	1	2	1
Asn	AAU	0	2	0	0	2	49	0	1	3	0	4	9	28	25
	AAC	26*	38*	9*	6*	5*	4	14*	17*	42*	8*	66*	31*	24*	33*
Phe	UUU	0	3	2	2	2	34	2	2	10	1	15	10	16	31
	UUC	21*	28*	12*	2*	2*	38	26*	7*	23*	9*	44*	21*	23*	30*
Tyr	UAU	0	1	0	4	2	56	3	2	3	1	18	7	17	31
	UAC	20*	18*	14*	6*	3*	6	17*	15*	13*	2*	38*	22*	11*	18*
Glu	GAA	29*	53*	26*	15*	5*	31	60*	9*	61*	24*	147*	60*	52*	74*
	GAG	0	0	1	1	2	2	13	4	16	8	46	22	20	31
Cys	UGU	4	2	4	0	2	11	2	1	1	0	5	4	7	10
	UGC	0	0	0	0	1	1	4	1	6	2	5	3	16	17
His	CAU	0	0	3	2	2	36	3	2	3	1	5	3	11	20
	CAC	16	20	7	2	2	0	19	3	15	3	23	12	9	20
Asp	GAU	15	14	11	4	1	31	8	6	17	2	60	28	36	60
	GAC	30	47	9	2	3	1	41	16	45	12	85	36	18	34
Met	AUG	13	14	15	2	2	45	18	5	30	9	60	35	31	33
Trp	UGG	6	10	4	0	1	(26)	2	5	3	0	8	6	7	4

The number of methionine codons does not include the initiation codons. References for nucleotide sequences of individual genes are as follows: actin gene (Gallwitz & Sures, 1980); *CYC1* (Smith *et al.*, 1979); r-pro. (*rpl*AJKL+*rps*BJLT; Post *et al.*, 1979; Post & Nomura, 1980; An *et al.*, 1981; Mackie, 1981; Olins & Nomura, 1981); *tsf* (An *et al.*, 1981); *rpo*BD (Ovchinnikov *et al.*, 1981; Burton *et al.*, 1981); *atp* (genes for subunits c, b, δ, α and γ of the membrane-bound ATP synthetase: Gay & Walker, 1981*a,b*; Kanazawa *et al.*, 1981*a,b*; Nielsen *et al.*, 1981); *thrB* (Cossart *et al.*, 1981); *trp*E (Nichols *et al.*, 1981*a,b*); and other genes (see references cited by Ikemura, 1981*a,b*). The asterisks specify optimal codons of yeast or *E. coli* assigned in Table 6. CCG, but not CCA, was assigned to the Pro optimal codon of *E. coli* (see Ikemura, 1981*b*). One noticeable discrepancy between the optimal codon and those preferred in yeast genes was the rare use of GGU. This might relate to the prediction of Grosjean *et al.* (1978) that codons of the type (G/C)–(G/C)–pyrimidine, where the first and second letters give an intrinsically strong interaction with an anticodon, would support the optimal interaction strength when the third letter is U. However, the respective preference was not found for Gly codon usage in *E. coli* genes, and the prediction was not adopted as the criterion of both the previous and the present papers.

† Mit., mitochondrial genes (Bonitz *et al.*, 1980). Numbers of codons listed in parentheses at the position of Leu codons correspond to those of Thr codon usage and that at Trp to the use of the UGA codon.

separately, the conclusions reached were essentially the same as those of the present study (see Ikemura, 1981*a*,*b*: these reports have dealt with most of the *E. coli* protein genes that had been sequenced).) Table 1 shows that choices among synonymous codons in all yeast and *E. coli* genes are clearly non-random, and that the similarity of the choices exists between the different genes in each organism with the exception of yeast mitochondrial genes. For approximately half the amino acid families, synonymous choice patterns are clearly different between the two organisms. For example, the most preferred codon of Leu is UUG for all yeast nuclear genes but CUG for all *E. coli* genes, that of Arg is AGA for the former but CGU for the latter, and that of Gln is CAA for the former but CAG for the latter (Table 1).

In previous papers (Ikemura, 1981*a*,*b*) the relative quantities of *E. coli* tRNAs have been measured and the non-random choice among synonymous codons in *E. coli* genes has been attributed to the constraints due to the availability of tRNAs and to the nature of their codon recognition. Therefore it will be interesting to learn whether or not the same mechanism is responsible for the non-random patterns of yeast genes.

(b) *Relative amounts of yeast tRNAs*

Transfer RNAs can be separated with a high degree of resolution by two-dimensional polyacrylamide gel electrophoresis (Ikemura & Dahlberg, 1973; Fradin *et al.*, 1975). In an improvement on our previous methods used for *E. coli* tRNAs (Ikemura *et al.*, 1975; Ikemura & Nomura, 1977), systems that purify most yeast tRNAs were developed. Three separation patterns are shown in Figure 1. These represent the results from the three different electrophoretic systems (A, B and C; see Materials and Methods). Utilizing these systems, more than 40 tRNAs were purified. Each species separated on a gel was assigned to a known tRNA by RNA fingerprinting followed by oligonucleotide analyses. Six examples of ribonuclease T_1 fingerprints are presented in Figure 2. The results of gel spot-tRNA assignments are summarized in Table 2. Since ^{32}P-labelling was continued over a growth period of several generations, the radioactivity of each gel spot represents the amount of tRNA after correction for RNA chain length. The amount of each tRNA, expressed as a ratio to that of $tRNA_3^{Leu}$, is presented in Table 2. Combining the results obtained from the three systems, the relative abundance of most of the sequenced cytoplasmic tRNAs (21 out of 24 sequenced tRNAs) has been determined (Table 3). A qualitative estimate of these tRNAs based on published chromatographic separation patterns is also listed. The quantification data agree well with such previous knowledge. Codons that must be recognized by each tRNA are also listed in Table 3 (see the references of Sodd (1976) and those in the Table).

(c) *Correlations between the codon usage in yeast genes and the abundance of the cognate tRNAs*

In order to learn whether the choice of codon in yeast genes relates to tRNA abundance or not, codon frequency in yeast genes was examined in connection with the abundance of cognate tRNAs. The frequency of use of each tRNA (i.e.

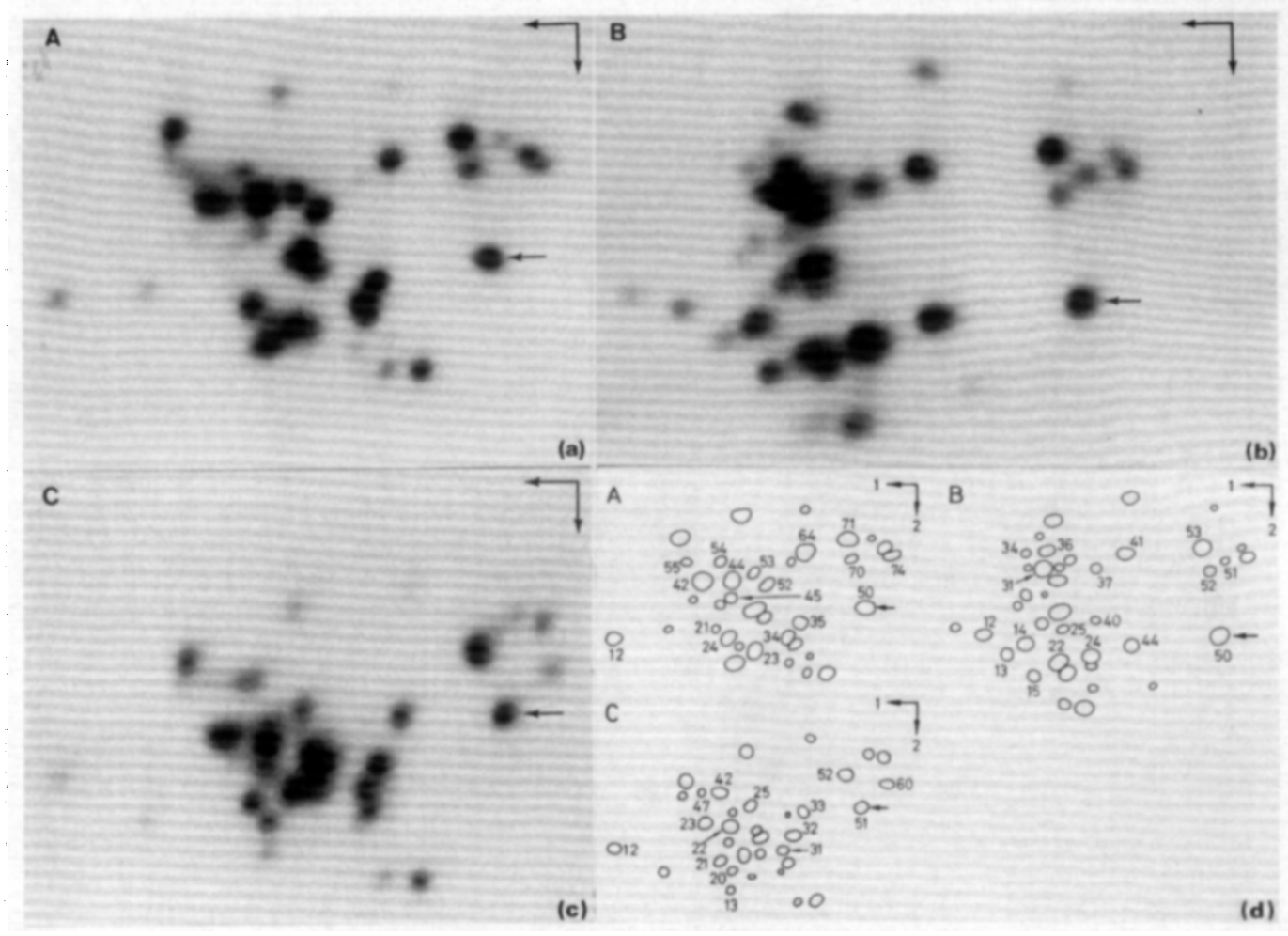

FIG. 1. Two-dimensional gel separation of ^{32}P-labelled yeast tRNAs. Electrophoresis in the 1st dimension was from right to left and that in the 2nd dimension was from top to bottom. (a), (b) and (c) Autoradiograms of A, B and C type gel runs, respectively. The arrow points to the spot of $tRNA_3^{Leu}$. (d) Schematic diagram of the autoradiograms. Spot numbers are those assigned to known tRNAs.

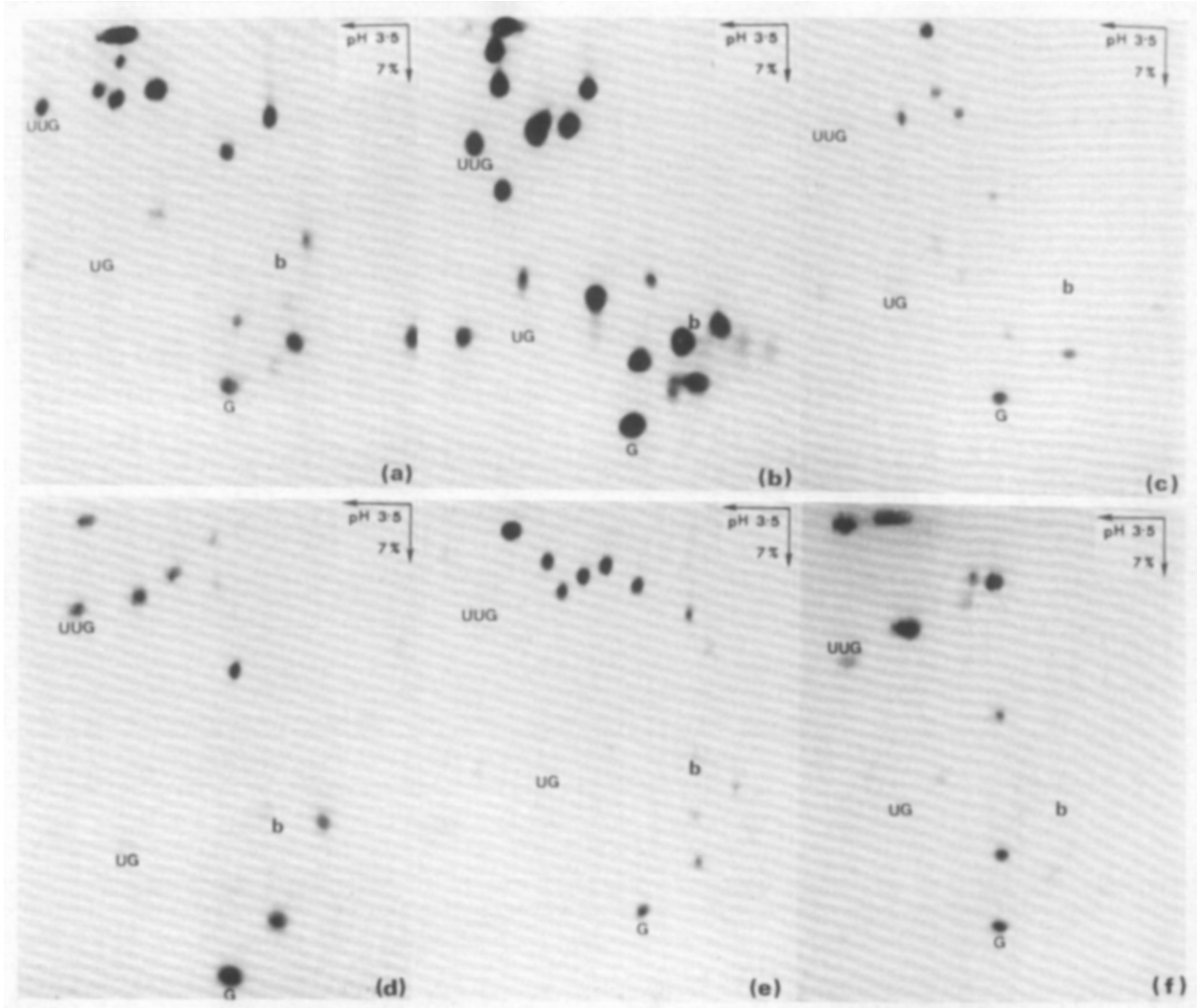

FIG. 2 Fingerprints of ribonuclease T_1 digests of tRNAs. Electrophoresis in the 1st dimension, on cellulose-acetate at pH 3·5 in 6 M-urea, 8 mM-EDTA, was from right to left; in the 2nd dimension, on DEAE-cellulose paper in 7% (v/v) formic acid, was from top to bottom. For reference, the positions of Gp, UGp and UUGp are indicated in every fingerprint, whether or not the corresponding spots were there. The position of a blue marker dye (xylene cyanol FF) is shown by b. (a) $tRNA_1^{Lys}$; (b) $tRNA_2^{Ser}$; (c) $tRNA_1^{Ala}$; (d) $tRNA^{Gly}$; (e) $tRNA_1^{Val}$; (f) $tRNA_7^{Lys}$.

TABLE 2

Assignment of gel spots to known tRNAs, and relative contents of tRNAs

	A		B		C	
tRNA	(Spot)	Content	(Spot)	Content	(Spot)	Content
Met_i	(12)	0·35	(12)	0·35	(12)	0·24
Arg2	(21)	0·22				
Thr1	(23)	1·03	(22)	0·78		
Arg3	(24)	1·00	(14)	0·77	(21)	0·92
Phe	(34)	0·84	(24)	0·70	(31)	0·74
Val1	(35)	0·94	(44)	0·99	(32)	1·10
Gly	(42)	1·46	(31)	1·54	(23)	1·25
Asp	(44)	1·43			(22)	1·21
Lys1	(45)	0·36	(13)	0·34	(13)	0·35
Leu3	(50)	1·00	(50)	1·00	(51)	1·00
Glu3	(52)	0·96				
Lys2	(53)	0·92	(37)	0·68	(25)	0·76
Ala1	(54)	0·58	(36)	0·69	(42)	0·72
Ala1†	(55)	0·31	(34)	0·33	(47)	0·23
Tyr	(64)	0·80	(41)	0·93	(33)	0·86
Leu1	(70)	0·48	(52)	0·46		
Ser2	(71)	1·22	(53)	1·12	(52)	1·17
SerUCA/G‡	(74)	0·45	(51)	0·39	(60)	0·31
Trp			(15)	0·67	(20)	0·54
Cys			(25)	0·39		
Val2b			(40)	0·15		

The nomenclature of tRNA is mainly according to Sodd (1976) and Sprinzl *et al.* (1980). Spot numbers of A, B or C type gels in Fig. 1 are listed in parentheses. Gel spots other than those listed correspond to tRNAs whose sequences have not been determined or to a mixture of tRNAs. The amount of $tRNA_3^{Leu}$ was normalized at 1·0 (underlined). Data for tRNA content listed in each column (A, B, C) are averages obtained in 3 gel runs analysing 2 RNA samples prepared separately. Deviations from the average were usually in the 20% range. As a background level of radioactivity, 3 gel disks with no detectable RNA spots were cut out, and the average of their counts (about 0·03 $tRNA_3^{Leu}$ units) was subtracted from the count of each tRNA spot.

† A partially modified form of $tRNA_1^{Ala}$ formed a gel spot separate from that of the mature $tRNA_1^{Ala}$.

‡The closely related molecules $tRNA_{UCA}^{Ser}$ and $tRNA_{UCG}^{Ser}$ (Olson *et al.*, 1981) were not separated by the present gel systems, and they were treated as a collective tRNA in the further analyses.

Mitochondrial tRNAs were not examined here because only a few nucleotide sequences, in which modified nucleotides are assigned, have been reported for them.

anticodon use) was deduced for individual genes according to the strategy used for *E. coli* genes (Ikemura, 1981*a*). The usage frequency of a tRNA that responds to a single codon was defined as the occurrence of the codon itself, and that of a tRNA that responds to multiple codons as the total of the occurrences of the codons (see also the legend to Table 3). In order to examine the correlation between tRNA content and frequency of tRNA usage calculated for each gene, the values of these two parameters are plotted in Figure 3. The results for yeast 2μ plasmid genes (Able, Baker, Charlie; Hartley & Donelson, 1980) are also shown here (they were treated as a collective gene; when they were analysed separately, essentially the same conclusion was reached). Data were then analysed by linear regression. Regression lines ($y = a + bx$) are drawn in each graph and the slope value (b) and y intercept (a) are listed in Table 4. High slope values were found for the genes for

TABLE 3

Codon recognition and relative content of tRNAs

tRNA	Codon	tRNA content	Estimate†	References for sequences, codon recognition and content of tRNA
Leu3	UUG	1·0	Major	Chang *et al.* (1973); Kowalski *et al.* (1971)
Leu1	CUA, CUG	0·47	Minor	Randerath *et al.* (1979)
Arg3	AGA, AGG‡	0·90	Major	Kuntzel *et al.* (1972); Keith & Dirheimer (1980)
Arg2	CGU, CGC, CGA	0·22	Minor	Weissenbach *et al.* (1975)
Lys2	AAG	0·79	Major	Madison & Boguslawski (1974); Smith *et al.* (1973)
Lys1	AAA, AAG‡	0·35	Minor	Smith *et al.* (1973)
Ser2	UCU, UCC, UCA§	1·17	Major	Zachau *et al.* (1966); Kruppa & Zachau (1972)
SerUCA/G	UCA, UCG	0·38	Minor	Etcheverry *et al.* (1979); Olson *et al.* (1981)
Val1	GUU, GUC, GUA	1·01	Major	Bonnet *et al.* (1974)
Val2b	GUG	0·15	Minor	Gorbulev *et al.* (1977)
Gly	GGU, GGC	1·41	Major	Yoshida (1973); Bergquist (1966)
Ala1	GCU, GCC, GCA	0·95	Major	Holley *et al.* (1965); Penswick *et al.* (1975)
Thr1	ACU, ACC, ACA	0·91	Major	Weissenbach *et al.* (1977)
Asp	GAU, GAC	1·32		Gangloff *et al.* (1972)
Glu3	GAA, GAG‡	0·96		Kobayashi *et al.* (1972)
Tyr	UAU, UAC	0·89		Madison & Kung (1967)
Phe	UUU, UUC	0·76		RajBhandary & Chang (1968); Söll & RajBhandary (1967)
Trp	UGG	0·61		Keith *et al.* (1972)
Cys	UGU, UGC	0·39		Holness & Attfield (1976)
Met_i	AUG	0·31		Simsek & RajBhandary (1972)

Codon recognition by individual tRNAs was according to Sodd (1976) and to the references listed.

† Estimate of content of individual tRNAs from the published works. The most abundant isoacceptor is designated as "major" and others as "minor".

‡ It has been well established that the introduction of 2-thiouridine and its derivatives (as well as 5-carboxymethyl uridine) at the anticodon wobble position produces a preference for using the A-terminated codon over the G-terminated codon (Nishimura, 1972,1978; Weissenbach & Dirheimer, 1978; see also Table 5). However, there seems to be some confusion about the coding properties of the respective tRNAs (e.g. see the references of Madison & Boguslawski, 1974). It has not yet been settled whether such tRNAs respond solely to the A-terminated codon or respond to the A-terminated codon much better than to the G-terminated codon. In cases where there are no tRNAs responding solely to the G-terminated codon, it now seems that the codon is most likely recognized by the tRNA having thiolated uridine or at least by the unmodified form of the tRNA (e.g. see the discussion by Chakrabortty *et al.*, 1975). Therefore, such G-terminated codons are included as those responded to by the respective tRNAs in this Table and in the further analyses made in Table 4 and in Figs 3 and 4. The analysis with or without this inclusion, however, does not produce any significant differences because the respective G-terminated codons are rarely used in yeast genes (see AGG and GAG uses in Table 1). In the case of the lysine family there is another, and the most abundant, isoacceptor solely responding to AAG (Lys2) and, therefore, this codon use is assigned to the use of this tRNA in the later analyses.

§ The anticodon, whose wobble base is I, seems to be poorly suited to the translation of the A-terminated codon (Olson *et al.*, 1981; Ikemura, 1981*b*; see also Table 5). Since there is one serine tRNA solely responding to UCA ($tRNA^{Ser}_{UCA}$), this codon use is assigned as the use of the tRNA in the further analyses made in Table 4 and in Figs 3 and 4. This assignment, however, does not affect the conclusions deduced there because this codon is rarely used in yeast genes (Table 1).

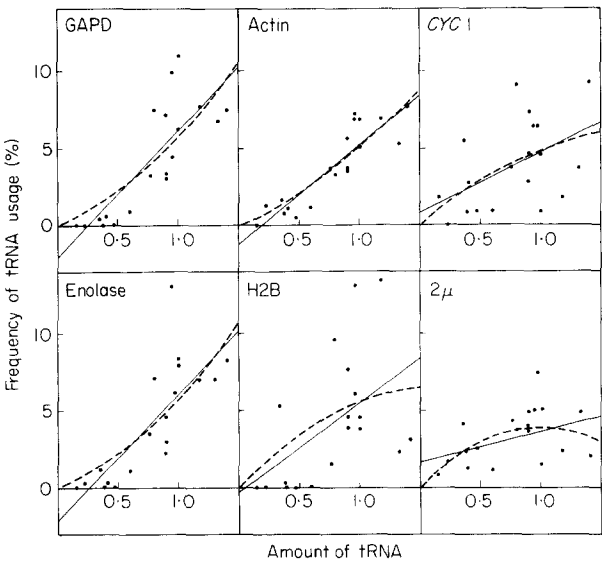

FIG. 3. The correlation between tRNA abundance and the frequency of its usage found for individual genes. Regression lines ($y = a + bx$) are shown by unbroken lines. Regression curves through the origin ($y = \alpha x^2 + \beta x$) were calculated according to Ikemura (1981*b*) and are shown by broken lines. GAPD, glyceraldehyde-3-phosphate dehydrogenase.

TABLE 4

Slopes and y *intercepts of regression lines and frequency of optimal codon use*

Gene	Slope	*y* intercept	Frequency of optimal codon use
GAPD	8·3	−2·1	0·99
Enolase	8·3	−2·1	0·97
Actin	6·5	−1·2	0·91
Histone H2B	5·9	−0·3	0·86
CYC1	3·8	+0·9	0·71
2μ	2·1	+1·8	0·41
Evenly used	0·7	+2·7	0·40

In regression analyses made in Fig. 3 and here, initiator $tRNA^{Met}$ was omitted from the calculation because its usage frequency is deduced solely from the gene's length. To learn the seeming correlation between tRNA content and its use in random base sequences, an idealized sequence in which all codons are evenly used was examined (Evenly used) and the value obtained was considered the basal level when actual genes were examined. GAPD, glyceraldehyde-3-phosphate dehydrogenase.

glyceraldehyde-3-phosphate dehydrogenase and enolase, both of which are glycolysis enzymes and are believed to be the most abundant protein species within yeast cells (Holland & Holland, 1979; Holland *et al.*, 1981). High slope values and negative *y* intercepts have previously been found for *E. coli* genes which encode abundant protein molecules (Ikemura 1981*a*,*b*). High slope values indicate the strong dependence of tRNA use on its abundance, and the negative *y* intercepts of such *E. coli* genes have been attributed to the "upward concavity" of the

representative function that was obtained by curvilinear regression. Curve-fitting was attempted for yeast genes using second-degree polynomial regression through the origin as had been done for *E. coli* genes, and the resulting functions are indicated by the broken lines in Figure 3. Upward concavity is shown for the glycolysis enzyme genes (and possibly for the actin gene). This concavity, however, is absent in other genes, whose data points seem to be represented by linear functions or upward convex curves. The slope value (b) of the 2μ plasmid gene was the lowest of all genes examined, indicating the weak dependence of its tRNA use on the abundance.

(d) *Synonymous codon usage*

(i) *Usage of isoaccepting tRNAs*

Let us now compare synonymous codon usage with the abundance of the cognate isoaccepting tRNAs. Because analyses of synonymous codon choice are not affected by the amino acid composition of the protein, it should be possible to estimate more directly the constraint on codon choice due to tRNA availability than was possible in Figure 3. Both the amount and the frequency of use of the most abundant tRNA of each amino acid family are normalized at 1·0. The uses of isoaccepting tRNAs for Arg, Leu, Ser, Val, and Lys are shown in Figure 4. The most abundant isoacceptor is almost always used in the highest frequency in yeast nuclear genes, as well as *E. coli* genes (Ikemura, 1980,1981*a*,*b*). It should be pointed out here that the codon to which the most abundant yeast isoacceptor responds often differs from the respective codon of *E. coli* (e.g. the major Leu isoacceptor of

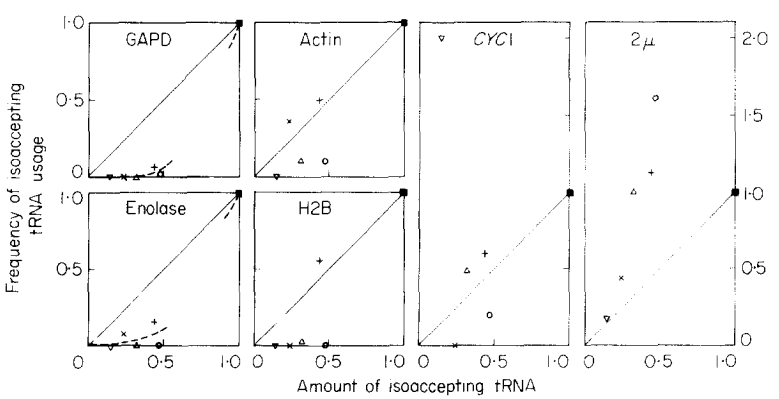

FIG. 4. The relationship between the content of isoaccepting tRNA and the frequency of its usage found for individual genes. Each line in the Figure can be predicted if the use of isoaccepting tRNA is proportional to its abundance. The most abundant isoacceptor of each amino acid is shown (■), and other isoacceptors are Arg (×), Leu (○), Ser (△), Val (▽) and Lys (+). A seemingly high frequency of the minor Val isoacceptor use (▽) in *CYC1* relates largely to the small size of the gene (108 codons) and to its low Val content (2 GUG and 1 GUU uses; see Table 1). The 2μ gene consists of 1089 codons and then the high frequency of minor tRNA codon usage in this gene is significant. GAPD, glyceraldehyde-3-phosphate dehydrogenase.

yeast responds to UUG but that of *E. coli* to CUG; see also Table 6). This should be one important causal factor of the differential patterns of synonymous choices found for the two organisms.

The line in each graph represents the correlation predicted if the frequency of usage of an isoaccepting tRNA is proportionate to its abundance. The data points of glyceraldehyde-3-phosphate dehydrogenase and enolase genes were clearly far from the line and are graphed as probable concave curves (broken line). This means that the dependence of tRNA use on tRNA content in these genes is much greater than that expected from a directly proportional relationship between the two variables. In other words, these genes selectively use codons responded to by the most abundant isoaccepting tRNA but almost completely avoid using codons responded to by other isoacceptors, showing the strong constraint on codon choice due to tRNA availability. The same has been found for *E. coli* genes coding for the very abundant proteins in *E. coli* cells (Ikemura, 1980,1981*a*,*b*).

In the case of the plasmid gene (2μ, in Fig. 4) a weak dependence of codon choice on isoacceptor abundance was revealed; minor isoacceptors were often used with high frequencies. The same has been found for the bacterial plasmid, transposon and phage genes, and possible explanations have been discussed (Ikemura, 1981*a*).

(ii) *Synonymous codons recognized by the same tRNA*

As in *E. coli* genes, codon choices in yeast nuclear genes appear to be constrained by tRNA availability. Constraints on choices among the codons to which the same tRNA responds have also been found for *E. coli* genes (Ikemura, 1981*a*,*b*). The following three constraints, presumably related to translational efficiency, have been proposed in the previous papers. (1) The thiolation of uridine in the wobble position of an anticodon produces a preference for using an A-terminated codon over a G-terminated codon (Sekiya *et al.*, 1969; Nishimura, 1972,1978) (formerly referred to as criterion 1); the same preference produced by an analogous modified nucleotide, 5-carboxymethyl uridine, has been pointed out by Weissenbach & Dirheimer (1978) and this is designated here as criterion 1′. (2) Grosjean *et al.* (1978) proposed that codons of the (A/U)-(A/U)-pyrimidine type would support an optimal interaction strength between a codon and an anticodon when the third letter is C (criterion 2). (3) Introduction of inosine, I, at the wobble position of an anticodon would produce a preference for U- and C-terminated codons over the A-terminated codon which leads to purine–purine wobble pairing (criterion 3).

The choice among codons recognized by a single tRNA was examined for yeast genes based on the same three criteria that have been proposed for *E. coli* genes (Table 5; refer to Table legend for Ile criterion 2). Clearly codon choices in yeast nuclear genes conform well to the same criteria as in *E. coli* genes, although this is not true in the plasmid gene. It should be noted here that the tRNA species which have the modified nucleotides at the anticodon wobble position often differ between the two organisms. For example, I at the wobble position has been found for only $tRNA^{Arg}_{major}$ among *E. coli* tRNAs but for various yeast tRNAs as listed in Table 5. This is also a mechanism responsible for the differential codon use patterns of the two organisms.

TABLE 5

Preferential usage observed between codons corresponding to the same tRNA

tRNA	Criterion	Codon	Use of codons in individual genes					
			GAPD	Enolase	Actin	H2B	*CYC1*	2μ
Glu3	1	GAA/GAG	29/0	53/0	26/1	15/1	5/2	51/31
Arg3	1′	AGA/AGG	22/0	26/0	13/0	11/1	3/0	21/23
Phe	2	UUC/UUU	21/0	28/3	12/2	2/2	2/2	20/28
Tyr	2	UAC/UAU	20/0	18/1	14/0	6/4	3/2	18/22
Asn	2	AAC/AAU	26/0	38/2	9/0	6/0	5/2	10/44
Ile	(2)	(AUC/AUU)	(23/16)	(19/24)	(16/14)	(9/7)	(2/2)	(15/32)
Ile	3	$\frac{\text{AUY}\dagger}{2}$/AUA	19·5/0	21·5/0	15/0	8/0	2/0	23·5/36
Val1	3	$\frac{\text{GUY}\dagger}{2}$/GUA	36/0	34·5/0	13/0	6/0	0·5/0	16/24
Arg2	3	$\frac{\text{CGY}\dagger}{2}$/CGA	0/0	1/0	2·5/0	0/0	0/0	6·5/6
Thr1	3	$\frac{\text{ACY}\dagger}{2}$/ACA	23·5/0	20/0	10·5/0	10/3	3/2	13/28
Ala1	3	$\frac{\text{GCY}\dagger}{2}$/GCA	32·5/0	56·5/0	13/0	16·5/0	3·5/0	15/25

An underline, if the ratio $\geq 1{\cdot}5$; a broken underline, if $1{\cdot}5 >$ the ratio $> 1{\cdot}0$.

The wobble position base of $tRNA_1^{Lys}$ is the 2-thiouridine derivative. However, the use of Lys codons in yeast genes should not be examined based on criterion 1 because AAG is also recognized by another and the major lysine isoacceptor (Lys2).

Criterion 2 is related to an optimal interaction strength between a codon and an anticodon; i.e. where both the first and the second letters of a codon give an intrinsically weak interaction with an anticodon (two A·U pairs), G·C pairing at the third letter position would be preferred over G·U wobble pairing. Then the proposal of Grosjean *et al.* (1978) is not directly applicable to the Ile codon usage in yeast genes because the wobble position base of yeast $tRNA^{Ile}$ is I. The respective codon uses, therefore, are listed in parentheses, and the criterion is not adopted in the latter study of Ile codon use made in Table 6.

† Half of the sum of U- and C-terminated codon use. Y, pyrimidine.

(e) *Proposals concerning synonymous codon choice optimal for the yeast translation system*

(i) *Assignment of optimal codon*

Constraints on synonymous codon choice based on the above three criteria, which are presumably related to translational efficiency, are well satisfied in yeast nuclear genes. This is especially evident for genes which encode abundant proteins (Table 5). When combined with the constraints due to tRNA availability, the preference of synonymous codons for most amino acids can be predicted. The predicted order of preference is listed in Table 6, along with that for *E. coli* genes (Ikemura, 1981*b*). The restriction due to tRNA availability is designated criterion 4. A few codons that can be recognized by two tRNAs are also considered here. In cases where one of the two tRNAs recognizes two codons without particular preference and the other isoacceptor of similar content recognizes only

TABLE 6

Anticipation of codon preference

Species	Amino acid	Criterion	Expected preference
Yeast	Phe	2	UUC > UUU
	Leu	4	UUG > CUY, CUR, UUA
	Ile	3	AUY > AUA
	Val	3	GUY > GUA
		4	GUY, GUA > GUG
		Sum	GUY > GUR
	Ser	3	UCY > UCA
		4	UCY, UCA > UCG, AGY
		Sum	UCY > UCR, AGY
	Thr	3	ACY > ACA
		4	ACY, ACA > ACG
		Sum	ACY > ACR
	Ala	3	GCY > GCA
		4	GCY, GCA > GCG
		Sum	GCY > GCR
	Tyr	2	UAC > UAU
	Asn	2	AAC > AAU
	Lys	4	AAG > AAA
	Glu	1	GAA > GAG
	Arg	1′	AGA > AGG
		3	CGY > CGA
		4	AGR > CGY, CGR
		Sum	AGA > AGG, CGY, CGR
	Gly	4	GGY > GGR
E. coli	Phe	2	UUC > UUU
	Leu	4	CUG > CUY, CUA, UUR
	Ile	2	AUC > AUU
		4	AUY > AUA
		Sum	AUC > AUU, AUA
	Val	4	GUR > GUC
		5	GUU > GUC
		Sum	GUR, GUU > GUC
	Pro	4	CCR > CCY
	Thr	4	ACY > ACR
	Ala	4	GCR > GCC
		5	GCU > GCC
		Sum	GCU, GCR > GCC
	Tyr	2	UAC > UAU
	Gln	4	CAG > CAA
	Asn	2	AAC > AAU
	Lys	1	AAA > AAG
	Glu	1	GAA > GAG
	Arg	3	CGY > CGA
		4	CGY, CGA > CGG, AGR
		Sum	CGY > CGR, AGR
	Gly	4	GGY > GGR
		5	GGG > GGA
		Sum	GGY > GGG > GGA

Y, pyrimidine; R, purine; sum, integration of preferences anticipated from individual criteria. Several amino acids are not listed because the content of their isoacceptors has not been clarified or because no criteria are applicable.

one of the two codons, that recognized by both tRNAs will be more frequently used than the other. This is criterion 5 applied for *E. coli* genes.

It is then proposed that the codon which is predicted in Table 6 to have the highest preference for each amino acid should be the codon optimal for translational efficiency of the yeast cytoplasmic protein-synthesizing system. The "optimal codon" thus predicted for yeast nuclear genes is specified by asterisks in Table 1. The spectrum of these asterisks conforms strikingly well to the spectrum of codons preferred by yeast nuclear genes. The optimal codons predicted previously for *E. coli* genes are also specified by asterisks at the positions of *E. coli* genes in Table 1. (After publication of the previous papers (Ikemura, 1981*a*,*b*), DNA sequences of about 15 new *E. coli* genes have been reported, and their preferred codons again agree well with the *E. coli* optimal codons previously proposed. The *atp* and *tsf* listed in Table 1 (as well as 3 ribosomal protein genes, *rpo*D, *thr*B and *trp*E: these are analysed as a part of the collective genes) correspond to the newly sequenced genes.) Based on the same criteria the optimal codons for each organism were deduced, and they conformed strikingly well to the preferred codons in the respective genes. This revealed that synonymous choice in genes of the two organisms is restricted by the common constraints, which are most probably related to the efficiency of translation process.

The difference between *E. coli* and yeast optimal codons can be attributed to differences in tRNA abundance as well as the modified nucleotides at the wobble position of the respective tRNAs. The spectrum of codons preferred by yeast mitochondrial genes (Table 1) clearly differs from that of yeast optimal codons which were deduced solely from information about cytoplasmic tRNAs. Because mRNA molecules of mitochondrial genes are translated using mitochondrial tRNAs, the availability of cytoplasmic tRNAs will not restrict codon choices of mitochondrial genes.

(ii) *Frequency of use of optimal codons*

Figure 5 shows the distribution of yeast optimal codons in four yeast genes. The optimal codon is indicated by O or o; the former when one codon is the optimal codon of the amino acid and the latter when multiple codons are assigned to it. The "non-optimal codon" is indicated by X. For the amino acids, methionine and tryptophan, a single codon corresponds and is indicated by Δ. To the amino acids Pro, Gln, Cys, His and Asp, no marks are added, either because their isoacceptors have not been quantified or because no criteria were applicable. Genes for glyceraldehyde-3-phosphate dehydrogenase and enolase, the most abundant proteins in yeast cells, use mostly the optimal codons. Optimal codon use decreases somewhat in the histone H2B gene and more drastically in the 2μ Able gene.

The frequency of use of optimal codon is defined as:

$$F_{op} = \frac{\text{sum of the number of O and o}}{\text{sum of the number of O, o and X}}.$$

The frequencies calculated for individual yeast genes are listed in Table 4. In order to examine the distribution of optimal and non-optimal codons in individual genes,

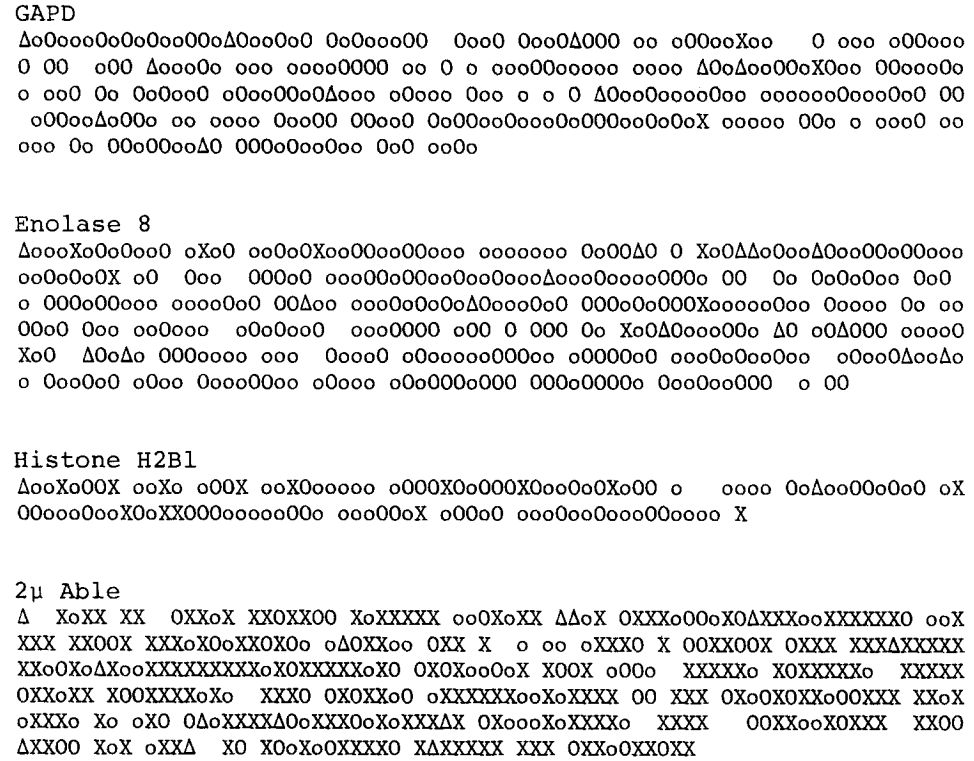

FIG. 5. Occurrences of optimal and non-optimal codons in yeast genes. To the amino acids His, Gln, Asp, Cys and Pro, no marks are added. In the DNA sequence of glyceraldehyde-3-phosphate dehydrogenase (GAPD) (*pgag491*), a few nucleotides have not been determined (Holland & Holland, 1980) and the respective codons are not marked here.

each gene was divided into segments of ten codons, and the frequency (F_{op}) in each segment was calculated and graphed using bar diagrams. Figure 6 shows that optimal codons (and therefore non-optimal codons) are usually well distributed within a gene. The frequency in the total portion of each gene (Table 4) is shown in the Figure by a horizontal broken line. The frequencies of glyceraldehyde-3-phosphate dehydrogenase and enolase were roughly similar and at the highest level of the yeast genes. The values for actin and histone H2B were lower than those of the above glycolysis enzymes but higher than that of *CYC1*. The order of these values appears to parallel the order of abundance of the respective proteins. (Precise correlations of this value and the level of gene expressivity has been revealed for *E. coli* genes (Ikemura, 1981*b*).) The value for the 2μ plasmid gene was the lowest of all genes examined.

4. Discussion

(a) *Measurement of tRNA content*

It is believed that most yeast tRNAs are encoded by multiple genes. For example, $tRNA_2^{Ser}$ seems to be encoded by 11 genes per yeast haploid genome (Page

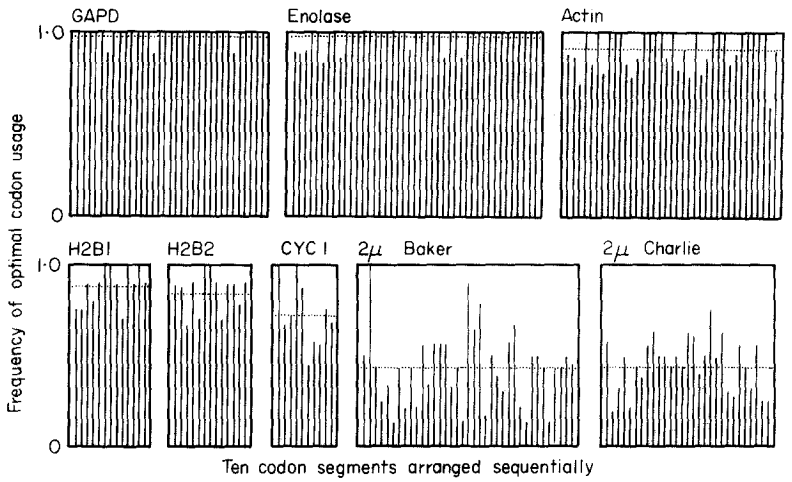

FIG. 6. Distribution of optimal codon use within individual genes. Each gene was divided into segments of 10 codons and frequencies of optimal codon usage in the segments were calculated. The residual codons at the last segment were included in the preceding segment and the frequency in the composite segment was calculated. Segments are sequentially arranged from the initiation codon side to the termination codon side. GAPD, glyceraldehyde-3-phosphate dehydrogenase.

& Hall, 1981). Therefore, a certain degree of heterogeneity in tRNA sequences can be expected among those with identical anticodons. For example, $tRNA_3^{Arg}$ consists of two different molecules that differ by only two nucleotides near the 5′ and 3′ ends in their total sequences (Keith & Dirheimer, 1980). The present study aims at quantifying anticodon abundance but not at measuring separately the abundance of tRNAs of such "micro-heterogeneity". It is conceivable that the micro-heterogeneous tRNAs may be separated from each other by the present gel electrophoreses. To minimize the possible artifacts of this potential separation (e.g. a certain tRNA that is possibly separated from the major population with a common anticodon might be missed being counted), three different systems of two-dimensional gel electrophoresis were used. They gave very different separation patterns, and each purified a large portion of yeast tRNA molecules. RNAs were eluted from all gel spots detected on autoradiograms and were analysed by RNA fingerprinting. It turned out that only in a few cases tRNAs with identical anticodons were separated into multiple spots (e.g. $tRNA_1^{Ala}$), and that the differences in those molecules separated were attributed to differences in nucleotide modification rather than nucleotide sequence. The present electrophoretic systems, therefore, usually do not separate tRNAs that display only small differences in nucleotide sequence.

In order to check the effects of the RNA preparation method on tRNA abundance, two different RNA extraction procedures were examined. In most experiments, the phenol extraction of cell suspensions was performed at 55°C according to Knapp *et al.* (1978) and a separate experiment at room temperature was done. They gave essentially the same result of relative abundance of tRNAs (data not shown). Systematic studies of the effects of changes in the strain and/or

the growth medium has not been made. When a second *S. cerevisiae* strain closely related to that used here was examined, similar tRNA abundance was found (see Materials and Methods). One noticeable difference, however, was that the relative ratio of initiator $tRNA^{Met}$ to other tRNAs was about 60% of the present value. Further studies of this discrepancy have not been done. The difference, if real, should not affect the present conclusions since initiation codon usage was not dealt with here.

Ikemura (1981*a*) has reported that the amounts of individual *E. coli* tRNAs are roughly proportional to their gene copy numbers, and therefore the gene number should be a major factor in determining the contents of *E. coli* tRNAs. Recently Olson *et al.* (1981) have summarized the gene numbers of yeast serine and tyrosine tRNAs and pointed out the relationship between the contents of $tRNA^{Ser}$ isoacceptors and their gene numbers; the numbers of genes for $tRNA^{Ser}_{UCG}$, $tRNA^{Ser}_{UCA}$, $tRNA^{Ser}_{2}$ and $tRNA^{Tyr}$ are 1, 3, presumably 11 and 8, respectively. Table 3 shows that the contents of $tRNA^{Ser}_{UCA+UCG}$, $tRNA^{Ser}_{2}$ and $tRNA^{Tyr}$ are 0·38, 1·17 and 0·89 $tRNA^{Leu}_{3}$ units, respectively, revealing a rough proportionality between tRNA contents and gene copy numbers. The gene numbers of other yeast tRNAs of high content (Leu3, Glu3, Phe), as far as estimated by hybridizing ^{32}P-labelled tRNA (or DNA fragment of tRNA gene) to restriction fragments of total yeast DNA which were separated on gels and transferred to membrane filters, seem to range from 10 to 14 (Venegas *et al.*, 1979; Feldmann *et al.*, 1981; Valenzuela *et al.*, 1978). These findings are consistent with the idea that the gene copy number is again a major factor in determining tRNA contents in yeast cells.

(b) *Codon choices in genes of other organisms*

The correlation between codon frequency and tRNA abundance could have resulted from an accumulation of a great number of mutations occurring in both protein and tRNA genes. The evident similarity of synonymous codon usage among the genes of each organism that encode a wide variety of different protein molecules indicates constraints which have restricted the synonymous choice during the course of evolution. The present study revealed that one of the most important factors is the constraint related to the availability of tRNAs and to their codon recognition. Since this conclusion was true for both a prokaryote (*E. coli*) and a eukaryote (yeast), this should be fairly common among a wide range of species. Yeast was chosen in the present study because it and *E. coli* are among those species whose DNA and tRNA sequences have been most intensively studied. The codon choices found for the genes of other organisms are discussed next. Protein genes listed in Table 7 should correspond to most, if not all, of the protein genes that have been completely sequenced for prokaryotic organisms other than *E. coli*. Genes of each organism were treated as a collective gene except for *nif* H of *Klebsiella pneumoniae*. It is clear that the synonymous codon choice pattern of *Salmonella typhimurium*, *Shigella dysenteriae*, *Serratia marcescens*, *Klebsiella pneumoniae* and *Erwinia amylovora*, all of which belong to the Enterobacteriaceae, are similar to that of *E. coli*. For this reference the *E. coli* optimal codons, which are based on *E. coli* tRNA abundance and the modified nucleotides at their anticodons, are indicated here by asterisks. The most preferred codon for each amino acid is

TABLE 7

Codon usage observed for prokaryote genes

Amino acid	Codon	*Salmonella trp*ABD	*Serratia lpp*+*trp*G	*Erwinia lpp*	*Shigella trp*D	*Klebsiella nif*D+*trp*A		*Anabaena nif*H
Leu	UUA	8	0	0	3	1	0	8″
	UUG	6	0	0	2	1	0	6
	CUU	7	2	0	2	2	1	1
	CUC	13	3	0	5	4	7	2
	CUA	9	0	0	0	0	0	4
	CUG*	47″	24″	9″	11″	32″	11″	2
Arg	CGU*	14	7	4″	3	1	4	12″
	CGC*	22″	9″	1	6″	20″	9″	5
	CGA	3	0	0	0	2	0	0
	CGG	3	0	0	0	2	0	0
	AGA	0	0	0	0	1	0	1
	AGG	1	0	0	0	0	0	1
Pro	CCU	8	2	0	0	1	0	4″
	CCC	11	1	0	0	3	3	2
	CCA	2	2	0	2	2	1	3
	CCG*	25″	5″	0	9″	25″	4″	0
Gln	CAA	10	5	0	1	2	2	9″
	CAG*	25″	13″	6″	8″	16″	8″	1
Lys	AAA*	25″	6″	4″	2″	12″	11″	6
	AAG	6	2	1	1	10	5	8″
Ala	GCU*	14	8	5″	3	2	1	16″
	GCC	34	10″	1	4	30″	14″	2
	GCA*	9	4	4	2	2	1	8
	GCG*	47″	10″	1	10″	22	13	1
Val	GUU*	13	3	2	1	0	2	7
	GUC	20″	6	0	2	13	8	0
	GUA*	4	2	3″	0	5	0	10″
	GUG*	11	9″	1	7″	16″	12″	0
Gly	GGU*	22	2	2″	9″	5	5	26″
	GGC*	42″	21″	1	6	22″	20″	3
	GGA	6	0	0	1	6	1	0
	GGG	12	0	0	3	7	1	0
Ser	UCU	5	3	3″	3	4	0	4
	UCC	17″	5″	0	0	6	6″	5″
	UCA	4	0	2	0	1	1	1
	UCG	7	4	0	2	6	2	0
	AGU	2	0	0	0	0	0	0
	AGC	8	4	2	6″	11″	1	2
Thr	ACU*	1	2	4″	1	0	0	1
	ACC*	19″	6″	1	5″	15″	12″	14″
	ACA	7	0	0	1	0	0	1
	ACG	10	0	0	2	6	4	0
Ile	AUU	26	2	0	13″	9	7	6
	AUC*	27″	13″	3″	2	23″	17″	15″
	AUA	2	0	0	0	0	0	1
Asn	AAU	12	3	1	5	4	1	2
	AAC*	20″	10″	5″	5	7″	11″	14″

TABLE 7 *(continued)*

Amino acid	Codon	*Salmonella trp*ABD	*Serratia lpp*+*trp*G	*Erwinia lpp*	*Shigella trp*D	*Klebsiella nif*D+*trp*A	*nif*H	*Anabaena nif*H
Phe	UUU	11	1	0	3″	7	2	0
	UUC*	19″	4″	0	2	14″	4″	6″
Tyr	UAU	10	1	0	3	6″	1	2
	UAC*	13″	4″	1″	3	4	8″	7″
Glu	GAA*	27″	7″	0	8″	14″	18″	22″
	GAG	24	4	0	0	13	11	3
Cys	UGU	5	1	0	1	0	2	2
	UGC	6″	3″	1″	2″	7″	7″	4″
His	CAU	20″	1	0	6″	4	0	0
	CAC	8	9″	1″	3	8″	2″	7″
Asp	GAU	28″	6	1	4″	15″	5	1
	GAC	11	10″	7″	3	10	11″	16″
Met	AUG	25	9	0	7	12	15	15
Trp	UGG	3	1	0	1	1	0	0

Initiation codon use was not counted. For Ser, Cys, His and Asp, *E. coli* optimal codons could not have been deduced because no criteria were applicable or because their isoacceptors were not quantified (Ikemura, 1981*b*). The symbol ″ was added to the most preferred codon of each amino acid in individual genes. References for nucleotide sequences of the genes are as follows: *trp*A of *Salmonella typhimurium* and *Klebsiella aerogenes* (Nichols *et al.*, 1981*a*,*b*); *Salmonella typhimurium trp*B (Crawford *et al.*, 1980); *Serratia marcescens trp*G and *Shigella dysenteria trp*D (Nichols *et al.*, 1980); *Serratia marcescens lpp* (Nakamura & Inouye, 1980); *Erwinia amylovora lpp* (Yamagata *et al.*, 1981); *Klebsiella pneumoniae nif*HD (Sundaresan & Ausubel, 1981; Scott *et al.*, 1981); and *Anabaena* 7120 *nif*H (Mevarech *et al.*, 1980).

noted (″). The spectra of most preferred codons in these Enterobacteriaceae genes agree well with the *E. coli* optimal codons (asterisks). This suggests that the tRNA abundance and the modified nucleotides at the anticodon wobble position have been well conserved in these Enterobacteriaceae. However, the codon choices of cyanobacterium *Anabaena* gene have several amino acids which differ from those of Enterobacteriaceae genes (e.g. Leu, Pro, Lys and Gln). It is worthwhile mentioning that codon choices by *nif*H (the gene encoding dinitrogenase reductase) of *Anabaena* (Mevarech *et al.*, 1980) and of *Klebsiella pneumoniae* (Sundaresan & Ausubel, 1981) clearly differ in these amino acids. In other words, the codon choices of *Klebsiella nif*H are more similar to those of most *E. coli* genes and of *Klebsiella trp*A or *nif*D than to those of *Anabaenae nif*H; *Klebsoella trp*A and *nif*D also have similar codon usage patterns (data not shown). This is one good example of the similarity of synonymous codon choice between different genes of the same or of taxonomically close genomes and the dissimilarity between genes of taxonomically distant genomes (even between those encoding the cognate protein). Recently the α-amylase gene of *Bacillus amyloliquefaciens* has been partially sequenced (Palva *et al.*, 1981). As far as can be judged by a partial sequence, the codon choice in the *Bacillus* gene is fairly different from that of *E. coli*. This may indicate that the abundance of isoaccepting tRNAs and the modified nucleotides at anticodons of

Bacillus and *Anabaena*, both of which are taxonomically distant from *E. coli*, differ significantly from those of *E. coli*.

Amino acid families in which *E. coli* and yeast exhibit similar codon choice patterns are considered next. These amino acids are listed in the lower portions of Tables 1 and 7. Good examples are the amino acids, Asn, Phe, Tyr and Ile, that are related to criterion 2 which states that codons of the (A/U)-(A/U)-pyrimidine type would support an optimal interaction strength with anticodon when the third letter is C. As long as one tRNA deals with these two codons, this criterion should be generally applicable regardless of the organism, except in cases where special modified nucleotides at the anticodon modulate the interaction strength between a codon and an anticodon (see the legend to Table 5). Actually the codon choices of the genes of yeast and of the six Enterobacteriaceae as well as *Anabaena* conform well to the criterion. It should be noted that codon choices even of mammalian genes conform well to it (e.g. see the codon usage table compiled by Grosjean (1980)).

Finally, the codon choice by genes of multicellular eukaryotes is discussed. In cells of certain highly specialized tissues, tRNA abundance is adjusted to codon frequency in the mRNA for the protein predominantly produced in the tissue (Garel *et al.*, 1970; Smith & McNamara, 1971). For example, in the silk gland of *Bombyx mori*, the amounts of several tRNA species are adjusted to the characteristic codon frequency in fibroin mRNA (Suzuki & Ikemura, 1973; Garel, 1976; Chavancy *et al.*, 1979; Chevallier & Garel, 1979). When correlation between tRNA abundance and the occurrence of respective codons in mRNAs of multicellular organisms is examined, it may be necessary to consider the tRNA abundance in the tissue in which the mRNA is actively translated. Although many mammalian protein genes have been sequenced, codon choices in such organisms are not dealt with in detail here because of this possible complication.

The author is very grateful to Dr H. Ozeki for valuable discussion and encouragement and to Dr M. Kimura for encouragement. This study was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan.

REFERENCES

Adams, J. M., Jeppesen, P. G. N., Sanger, F. & Barrell, B. G. (1969). *Nature (London)*, **223**, 1009–1014.

Air, G. M., Blackburn, E. H., Coulson, A. R., Galibert, F., Sanger, F., Sedat, J. W. & Ziff, E. B. (1976). *J. Mol. Biol.* **107**, 445–458.

An, G., Bendiak, D. S., Mamelak, L. A. & Friesen, J. D. (1981). *Nucl. Acids Res.* **9**, 4163–4172.

Barrell, B. G. (1971). In *Procedures in Nucleic Acid Research* (Cantoni, G. L. & Davies, D. R., eds), vol. 2, pp. 751–779. Harper and Row, New York.

Bergquist, P. L. (1966). *Cold Spring Harbor Symp. Quant. Biol.* **31**, 435–447.

Bonitz, S. G., Berlani, R., Coruzzi, G., Li, M., Macino, G., Nobrega, F. G., Nobrega, M. P., Thalenfeld, B. E. & Tzagoloff, A. (1980). *Proc. Nat. Acad. Sci., U.S.A.* **77**, 3167–3170.

Bonnet, J., Ebel, J. P., Dirheimer, G., Shershneva, L. P., Krutilina, A. I., Venkstern, T. V. & Bayev, A. A. (1974). *Biochimie*, **56**, 1211–1213.

Burton, Z., Burgess, R. R., Lin, J., Moore, D., Holder, S. & Gross, C. A. (1981). *Nucl. Acids Res.* **9**, 2889–2903.
Chakraburtty, K., Steinschneider, A., Case, R. V. & Mehler, A. H. (1975). *Nucl. Acids Res.* **2**, 2069–2075.
Chang, S. H., Kuo, S., Hawkins, E. & Miller, N. R. (1973). *Biochem. Biophys. Res. Commun.* **51**, 951–955.
Chavancy, G., Chevallier, A., Fournier, A. & Garel, J. P. (1979). *Biochimie*, **61**, 71–78.
Chevallier, A. & Garel, J. P. (1979). *Biochimie*, **61**, 245–262.
Cossart, P., Katinka, M. & Yaniv, M. (1981). *Nucl. Acids Res.* **9**, 339–347.
Crawford, I. P., Nichols, B. P. & Yanofsky, C. (1980). *J. Mol. Biol.* **142**, 489–502.
Efstratiadis, A., Kafatos, F. C. & Maniatis, T. (1977). *Cell*, **10**, 571–585.
Etcheverry, T., Colby, D. & Guthrie, C. (1979). *Cell*, **18**, 11–26.
Feldmann, H., Olah, J. & Friedenreich, H. (1981). *Nucl. Acids Res.* **9**, 2949–2959.
Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Merregaert, J., Min Jou, W., Raeymaekers, A., Volckaert, G., Ysebaert, M., Van de Kerckhove, J., Nolf, F. & Van Montagu, M. (1975). *Nature (London)*, **256**, 273–278.
Fradin, A., Grühl, H. & Feldmann, H. (1975). *FEBS Letters*, **50**, 185–189.
Gallwitz, D. & Sures, I. (1980). *Proc. Nat. Acad. Sci., U.S.A.* **77**, 2546–2550.
Gangloff, J., Keith, G., Ebel, J. P. & Dirheimer, G. (1972). *Biochim. Biophys. Acta*, **259**, 210–222.
Garel, J. P. (1976). *Nature (London)*, **260**, 805–806.
Garel, J. P., Mandel, P., Chavancy, G. & Daillie, J. (1970). *FEBS Letters*, **7**, 327–329.
Gay, N. J. & Walker, J. E. (1981*a*). *Nucl. Acids Res.* **9**, 2187–2194.
Gay, N. J. & Walker, J. E. (1981*b*). *Nucl. Acids Res.* **9**, 3919–3926.
Gorbulev, V. G., Axelrod, V. D. & Bayev, A. A. (1977). *Nucl. Acids Res.* **4**, 3239–3258.
Grantham, R., Gautier, C., Gouy, M., Mercier, R. & Pavé, A. (1980). *Nucl. Acids Res.* **8**, r49–r62.
Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. & Mercier, R. (1981). *Nucl. Acids Res.* **9**, r43–r74.
Grosjean, H. (1980). In *Transfer RNA: Biological Aspects* (Söll, D., Abelson, J. N. & Schimmel, P. R., eds), pp. 565–569, Cold Spring Harbor Laboratory, New York.
Grosjean, H., Sankoff, D., Min Jou, W., Fiers, W. & Cedergren, R. J. (1978). *J. Mol. Evol.* **12**, 113–119.
Hartley, J. L. & Donelson, J. E. (1980). *Nature (London)*, **286**, 860–865.
Holland, J. P. & Holland, M. J. (1979). *J. Biol. Chem.* **254**, 9839–9845.
Holland, J. P. & Holland, M. J. (1980). *J. Biol. Chem.* **255**, 2596–2605.
Holland, M. J., Holland, J. P., Thill, G. P. & Jackson, K. A. (1981). *J. Biol. Chem.* **256**, 1385–1395.
Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., Penswick, J. R. & Zamir, A. (1965). *Science*, **147**, 1462–1465.
Holness, N. J. & Atfield, G. (1976). *Biochem. J.* **153**, 447–454.
Ikemura, T. (1980). In *Genetics and Evolution of RNA polymerase, tRNA and Ribosomes* (Osawa, S., Ozeki, H., Uchida, H. & Yura, T., eds), pp. 519–253, University of Tokyo Press (Japan), Tokyo, and Elsevier/North-Holland Biomedical Press, Amsterdam.
Ikemura, T. (1981*a*). *J. Mol. Biol.* **146**, 1–21.
Ikemura, T. (1981*b*). *J. Mol. Biol.* **151**, 389–409.
Ikemura, T. & Dahlberg, J. E. (1973). *J. Biol. Chem.* **248**, 5024–5032.
Ikemura, T. & Nomura, M. (1977). *Cell*, **11**, 779–793.
Ikemura, T. & Ozeki, H. (1977). *J. Mol. Biol.* **117**, 419–446.
Ikemura, T., Shimura, Y., Sakano, H. & Ozeki, H. (1975). *J. Mol. Biol.* **96**, 69–86.
Kanazawa, H., Mabuchi, K., Kayano, T., Tamura, F. & Futai, M. (1981*a*). *Biochem. Biophys. Res. Commun.* **100**, 219–225.
Kanazawa, H., Kayano, T., Mabuchi, K. & Futai, M. (1981*b*). *Biochem. Biophys. Res. Commun.* **103**, 604–612.
Keith, G. & Dirheimer, G. (1980). *Biochem. Biophys. Res. Commun.* **92**, 116–119.

Keith, G., Roy, A., Ebel, J. P. & Dirheimer, G. (1972). *Biochimie*, **54**, 1417–1426.
Knapp, G., Beckmann, J. S., Johnson, P. F., Fuhrman, S. A. & Abelson, J. (1978). *Cell*, **14**, 221–236.
Kobayashi, T., Irie, T., Yoshida, M., Takeishi, K. & Ukita, T. (1972). *Biochim. Biophys. Acta*, **366**, 168–181.
Kowalski, S., Yamane, T. & Fresco, J. R. (1971). *Science*, **172**, 385–387.
Kruppa, J. & Zachau, H. G. (1972). *Biochim. Biophys. Acta*, **277**, 499–512.
Kuntzel, B., Weissenbach, J. & Dirheimer, G. (1972). *FEBS Letters*, **25**, 189–191.
Mackie, G. A. (1981). *J. Biol. Chem.* **256**, 8177–8182.
Madison, J. T. & Boguslawski, S. J. (1974). *Biochemistry*, **13**, 524–527.
Madison, J. T. & Kung, H. K. (1967). *J. Biol. Chem.* **242**, 1324–1330.
Mevarech, M., Rice, D. & Haselkorn, R. (1980). *Proc. Nat. Acad. Sci., U.S.A.* **77**, 6476–6480.
Nakamura, K. & Inouye, M. (1980). *Proc. Nat. Acad. Sci., U.S.A.* **77**, 1369–1373.
Nichols, B. P., Miozzari, G. F., van Cleemput, M., Bennet, G. N. & Yanofsky, C. (1980). *J. Mol. Biol.* **142**, 503–517.
Nichols, B. P., Blumenberg, M. & Yanofsky, C. (1981*a*). *Nucl. Acids. Res.* **9**, 1743–1755.
Nichols, B. P., van Cleemput, M. & Yanofsky, C. (1981*b*). *J. Mol. Biol.* **146**, 45–54.
Nielsen, J., Hansen, F. G., Hoppe, J., Friedl, P. & von Meyenberg, K. (1981). *Mol. Gen. Genet.* **184**, 33–39.
Nishimura, S. (1972). *Prog. Nucl. Acid Res. Mol. Biol.* **12**, 49–85.
Nishimura, S. (1978). In *Transfer RNA* (Altman, S., ed.). pp. 168–195. The MIT Press, Cambridge, Mass.
Olins, P. O. & Nomura, M. (1981). *Cell*, **26**, 205–211.
Olson, M. V., Page, G. S., Sentenac, A., Piper, P. W., Worthington, M., Weiss, R. B. & Hall, B. D. (1981). *Nature (London)*, **291**, 464–469.
Ovchinnikov, Y. A., Monastyrskaya, G. S., Gubanov, V. V., Guryev, S. O., Chertov, O. Y., Modyanov, N. N., Grinkevich, V. A., Makarova, I. A., Marchenko, T. V., Polovnikova, I. N., Lipkin, V. M. & Sverdlov, E. D. (1981). *Eur. J. Biochem.* **116**, 621–629.
Page, G. S. & Hall, B. D. (1981). *Nucl. Acids Res.* **9**, 921–934.
Palva, I., Pettesson, R. F., Kalkkinen, N., Lehtovaara, P., Sarvas, M., Söderlund, H., Takkinen, K. & Kääriäinen, L. (1981). *Gene*, **15**, 43–51.
Penswick, J. R., Martin, R. & Dirheimer, G. (1975). *FEBS Letters*, **50**, 28–31.
Post, L. E. & Nomura, M. (1980). *J. Biol. Chem.* **255**, 4660–4666.
Post, L. E., Strycharz, G. D., Nomura, M., Lewis, H. & Dennis, P. P. (1979). *Proc. Nat. Acad. Sci., U.S.A.* **76**, 1697–1701.
RajBhandary, U. L. & Chang, S. H. (1968). *J. Biol. Chem.* **243**, 598–608.
Randerath, E., Gupta, R. C., Chia, L. Y., Chang, S. H. & Randerath, K. (1979). *Eur. J. Biochem.* **93**, 79–94.
Rubin, G. M. (1975). In *Methods in Cell Biology* (Prescott, D. M., ed.), vol. 12, pp. 45–64. Academic Press, New York.
Sanger, F., Brownlee, G. G. & Barrell, B. G. (1965). *J. Mol. Biol.* **13**, 373–398.
Scott, K. F., Rolfe, B. G. & Shine, J. (1981). *J. Mol. Appl. Genet.* **1**, 71–81.
Sekiya, T., Takeishi, K. & Ukita, T. (1969). *Biochim. Biophys. Acta*, **182**, 411–426.
Simsek, M. & RajBhandary, U. L. (1972). *Biochem. Biophys. Res. Commun.* **49**, 508–515.
Smith, C. J., Teh, H. S., Ley, A. N. & D'Obrenan, P. (1973). *J. Biol. Chem.* **248**, 4475–4485.
Smith, D. W. E. & McNamara, A. L. (1971). *Science*, **171**, 577–579.
Smith, M., Leung, D. W., Gillam, S., Astell, C. R., Montgomery, D. L. & Hall, B. D. (1979). *Cell*, **16**, 753–761.
Sodd, M. A. (1976). In *Handbook of Biochemistry and Molecular Biology Nucleic Acids* (Fasman, G. D., ed.) 3rd edit., vol. 2, pp. 423–456, CRC Press, Cleveland.
Söll, D. & RajBhandary, U. L. (1967). *J. Mol. Biol.* **29**, 113–124.
Sprinzl, M., Grueter, F., Spelzhaus, A. & Gauss, D. H. (1980). *Nucl. Acids Res.* **8**, r1–r22.
Sundaresan, V. & Ausubel, F. M. (1981). *J. Biol. Chem.* **256**, 2808–2812.
Suzuki, Y. & Ikemura, T. (1973). *Yearb. Carnegie Instn.* **73**, 26–27.

Valenzuela, P., Venegas, A., Weinberg, F., Bishop, R. & Rutter, W. J. (1978). *Proc. Nat. Acad. Sci., U.S.A.* **75**, 190–194.
Venegas, A., Quiroga, M., Zaldivar, J., Rutter, W. J. & Valenzuela, P. (1979). *J. Biol. Chem.* **254**, 12306–12309.
Wallis, J. W., Hereford, L. & Grunstein, M. (1980). *Cell*, **22**, 799–805.
Weissenbach, J. & Dirheimer, G. (1978). *Biochim. Biophys. Acta*, **518**, 530–534.
Weissenbach, J., Martin, R. & Dirheimer, G. (1975). *Eur. J. Biochem.* **56**, 527–532.
Weissenbach, J., Kiraly, I. & Dirheimer, G. (1977). *Biochemie*, **59**, 381–391.
Yamagata, H., Nakamura, K. & Inouye, M. (1981). *J. Biol. Chem.* **256**, 2194–2198.
Yoshida, M. (1973). *Biochem. Biophys. Res. Commun.* **50**, 779–784.
Zachau, H. G., Dütting, D. & Feldmann, H. (1966). *Z. Physiol. Chem.* **347**, 212–235.

Edited by S. Brenner