Volume 16 Number 19 1988 Nucleic Acids Research

Codon usage in *Pseudomonas aeruginosa*

Susan E.H.West and Barbara H.Iglewski*

Department of Microbiology and Immunology, School of Medicine and Dentistry, University of Rochester, 601 Elmwood Avenue, Rochester, NY 14642, USA

Received May 25, 1988; Revised and Accepted September 8, 1988

ABSTRACT

We have generated a codon usage table for *Pseudomonas aeruginosa*. Codon usage in *P. aeruginosa* is extremely biased. In contrast to *E. coli* and yeast, *P. aeruginosa* preferentially uses those codons within a synonymous codon group with the strongest predicted codon-anticodon interaction. We were unable to correlate a particular codon usage pattern with predicted levels of mRNA expressivity. The choice of a third base reflects the high guanine plus cytosine content of the *P. aeruginosa* genome (67.2%) and cytosine is the preferred nucleotide for the third codon position.

INTRODUCTION

The advent of nucleotide sequencing has allowed a detailed analysis of codon usage in several different taxonomic groups. Grantham et al. (1, 2) demonstrated that the choice among synonymous codons is nonrandom and is characteristic of a particular species. These observations are the basis of their genome hypothesis for codon usage: genes within a species have a similar pattern of codon preference, but genes from different taxonomic groups have different patterns (1, 2). Nevertheless, considerable heterogeneity has been observed in codon usage patterns within a species. In highly expressed genes from *E. coli* and yeast, there is a greater preference for the nucleotide in the third codon position which yields codon-anticodon interactions of intermediate energy than there is in weakly expressed genes (3-7). In addition, highly expressed genes from *E. coli* and yeast contain predominantly codons translated by the most abundant tRNAs and rarely contain codons recognized by the less abundant tRNAs (8, 9). Weakly expressed genes, such as *dnaG* and *lacI*, contain many codons which are recognized by the less abundant tRNAs (10, 11). In prokaryotes, there is a correlation between the G+C content of the genome and the G+C content at synonymous sites in codons (12-15). For example, in organisms with a high G+C content, such as *Micrococcus luteus* (G+C, 70-75.5%), G and C are used almost exclusively at the third position (95%) (16) and in organisms with a low G+C content, such as *Mycoplasma capricolum* (G+C, 25%), the third position is very AT rich (G+C, 9.1%) (17). Trifonov (18) observed a three-base periodical pattern $(G\text{-non-}G\text{-}N)_n$ in protein coding sequences which is expressed as a strong preference for guanine in the first codon position and a lack of guanine in the second position and suggested that this periodicity may be responsible for monitoring the correct reading frame during translation.

These patterns of nonrandom codon usage are based on data from a large number of genes in *E. coli* and yeast, organisms with G+C contents of 48-52% and 40%, respectively (1-7). It has not yet been established whether these patterns of nonrandom codon usage apply to organisms with a high G+C content, such as *Pseudomonas aeruginosa* (G+C, 67.2%) (19). Nucleotide sequence data are now available for 28 chromosomal, plasmid, phage, and transposon encoded genes from *P. aeruginosa*. In this report, we discuss the choice of codons utilized by *P. aeruginosa*; compare the pattern of codon usage in *P. aeruginosa* to *E. coli*, yeast, and *M. luteus*; and present a reference codon usage table for *P. aeruginosa*.

METHODS

The program CodonFrequency from the University of Wisconsin Genetics Computer Group (UWGCG) (20) was utilized to generate codon usage tables from each of the *P. aeruginosa* genes in the GenBank (release 52.0) and EMBL (release 12.0) data bases (21, 22) and from the nucleotide sequences of additional genes not yet included in these data bases. Codon usage tables were generated for the following *P. aeruginosa* genes: *toxA*, the structural gene for exotoxin A (23); the pillin gene from *P. aeruginosa* strains PAO (24), PAK (25), and PA103 (26); *plcS*, the structural gene for phospholipase C (27); *trpA* and *trpB* which encode the α and β subunits of tryptophan synthase (28); *trpG*, the gene coding for the glutamine amidotransferase subunit of the anthranilate synthase genes (29); *pmi* which encodes phosphomannose isomerase (30); *algD*, the gene coding for GDP mannose dehydrogenase (31); *toxR*, the gene for a positive regulatory factor which enhances transcription of exotoxin A (32); *plcR*, a gene of unknown function in the *plcS* operon (33); the gene for azurin, a blue copper protein which serves as electron donor for nitrite oxidase (34); *amiE* which encodes aliphatic amidase (35); *recA* (36); the gene for porin protein F (37); *lasA*, a gene which encodes a membrane protein involved in processing or secretion of elastase (38); *lasB*, the structural gene for elastase (39); the PaeR7 restriction/modification system encoded on pMG7 (40); the *merA*, *merD*, *merP*, *merR*, *merT*, *tnpA*, and *tnpR* genes of transposon Tn501 (41-44); the DNA binding and major coat protein genes of bacteriophage Pf3 (45); and the DNA binding protein from bacteriophage Pf1 (46).

The UWGCG program Correspond was used to look for patterns of similar codon usage by comparison of codon frequency tables (47). The Correspond statistic D^2 is equal to the sum over all 64 codons of: $(\text{Frequency}_{(\text{codon table 1})} - \text{Frequency}_{(\text{codon table 2})})^2$. The frequencies compared are the number of incidents of the codon in question divided by the total number of codons specifiying that amino acid or terminator in the table. The correspond statistic gets smaller as the patterns of codon usage become more similar.

The relative synonymous codon usage value (RSCU) is the observed frequency of a codon divided by the frequency expected if there is uniform usage within synonymous codon groups (11). This value allows comparison between data sets of different sizes.

The P2 index (6) describes the proportion of codons conforming to the intermediate strength of codon-anticodon interaction rule of Grosjean and Fiers (5). P2 = (WWC + SSU) / (WWY + SSY), where W = A or U, S = G or C, Y = C or U, and WWC is the observed number of codons of that description. A P2 index of 0.5 indicates uniform codon usage.

The codon preference statistic, described by Gribskov et al. (47), was determined with the program CodonPreference from UWGCG (20). This statistic is based on the ratio of the likelihood of finding a particular codon in a highly expressed gene to the likelihood of finding that codon in a random sequence with the same base composition. The reference codon usage table for this analysis was compiled from codon usage data for *P. aeruginosa* chromosomal genes excluding pillin.

RESULTS AND DISCUSSION

Codon usage patterns among *P. aeruginosa* chromosomal, plasmid, phage, and transposon encoded genes

To compare the patterns of codon usage among the *P. aeruginosa* chromosomal, plasmid, phage, and transposon genes, we calculated the RSCU values for these groups (Table 1), compared composite codon usage tables for these groups by Correspond analysis (Table 2), and calculated codon bias indices for each gene (Table 3).

The greatest difference between patterns of codon usage occurred between the *P. aeruginosa* pillin genes and other chromosomal genes (Tables 1 and 2). *P. aeruginosa* pillin belongs to a class of similar pillin genes which are present in *Neisseria gonorrhoeae*, *Moraxella bovis*, *Moraxella nonliquefaciens*, and *Bacteroides nodosus* (48). The G+C content of the *P. aeruginosa* PAO, PAK, and PA103 pillin genes is 48.7%, 48.6%, and 46.6%, respectively, and is significantly different from the *P. aeruginosa* genome (G+C, 67.2%). Codons ending in T and G are the preferred codons in the *P. aeruginosa* pillin genes and there is an increased use of codons ending in A in contrast to other chromosomal genes from *P. aeruginosa* (Table 1). Sastry et al. (24) suggested that the pattern of codon usage in *P. aeruginosa* pillin is similar to the pattern observed in highly expressed genes of *E. coli*. However, Correspond analysis indicated that the pattern of codon usage in the pillin genes was more similar to *E. coli* weakly expressed genes (Table 2). In the pillin genes, there is a marked preference for U over C in all synonymous codon groups except proline. This preference is in contrast to *E. coli* where the preference is for U in GCY codons (Y=pyrimidine) and for C in AUY codons (6). The low G+C content of the pillin genes and their distinctly different pattern of codon usage suggest that *P. aeruginosa* obtained the pillin gene from an organism with a lower G+C content. The G+C contents of *N. gonorrhoeae*, *M. bovis*, *M. nonliquefaciens*, and *B. nodosus*, the other organisms which possess this unique class of pillin, are 49.5-53.3%, 41.0-44.5%, 40-44%, and 45%, respectively.

Table 1. Relative synonymous codon usage in *P. aeruginosa* chromosomal, transposon, plasmid, and phage genes[a].

codon	chromosome[b]		pillin[c]		Tn501[d]		plasmid[e]		phage			
									Pf1[f]		Pf3[g]	
GGG	31	0.24	3	0.29	19	0.53	12	0.79	1	0.36	3	1.33
GGA	18	0.14	3	0.29	10	0.28	4	0.26	0	0.00	0	0.00
GGT	58	0.46	27	2.57	16	0.44	14	0.92	1	0.36	5	2.22
GGC	401	3.16	9	0.86	100	2.76	31	2.03	9	3.27	1	0.44
GAG	210	1.24	11	0.92	81	1.31	31	1.29	2	0.57	1	0.50
GAA	130	0.76	13	1.08	43	0.69	17	0.71	5	1.43	3	1.50
GAT	46	0.27	16	1.78	29	0.52	24	1.07	1	0.29	5	2.00
GAC	294	1.73	2	0.22	83	1.48	21	0.93	6	1.71	0	0.00
GTG	170	1.68	6	0.92	79	2.12	21	1.79	4	1.45	6	2.18
GTA	28	0.28	7	1.08	6	0.16	1	0.09	0	0.00	0	0.00
GTT	18	0.18	9	1.38	11	0.30	7	0.60	0	0.00	3	1.09
GTC	188	1.86	4	0.62	53	1.42	18	1.53	7	2.55	2	0.73
GCG	183	1.23	18	1.29	93	1.42	20	0.87	2	0.44	3	1.09
GCA	19	0.13	12	0.86	14	0.21	15	0.65	2	0.44	0	0.00
GCT	43	0.29	22	1.57	16	0.24	23	1.00	0	0.00	5	1.82
GCC	350	2.35	4	0.29	139	2.12	34	1.48	14	3.11	3	1.09
AGG	12	0.21	2	1.09	4	0.14	6	0.51	0	0.00	0	0.00
AGA	3	0.05	0	0.00	2	0.07	1	0.08	0	0.00	0	0.00
CGG	49	0.85	0	0.00	49	1.75	19	1.61	1	1.00	0	0.00
CGA	11	0.19	0	0.00	3	0.11	6	0.51	0	0.00	0	0.00
CGT	34	0.59	9	4.92	15	0.54	14	1.18	3	3.00	2	2.40
CGC	235	4.10	0	0.00	95	3.39	25	2.11	2	2.00	3	3.60
AGT	12	0.21	4	0.89	7	0.36	3	0.46	0	0.00	0	0.00
AGC	135	2.42	4	0.89	49	2.51	9	1.38	2	2.00	0	0.00
TCG	89	1.59	10	2.22	25	1.28	17	2.62	2	2.00	1	1.20
TCA	3	0.05	0	0.00	4	0.21	2	0.31	0	0.00	0	0.00
TCT	3	0.05	8	1.78	2	0.10	3	0.46	1	1.00	1	1.20
TCC	93	1.67	1	0.22	30	1.54	5	0.77	1	1.00	3	3.60
AAG	196	1.77	12	0.73	70	1.65	22	1.63	7	1.75	2	0.80
AAA	25	0.23	21	1.27	15	0.35	5	0.37	1	0.25	3	1.20
AAT	18	0.16	9	1.06	6	0.26	8	0.80	1	0.50	2	0.80
AAC	202	1.84	8	0.94	40	1.74	12	1.20	3	1.50	3	1.20
ATG	119	1.00	9	1.00	31	1.00	11	1.00	3	1.00	3	1.00
ATA	2	0.03	1	0.07	0	0.00	1	0.08	0	0.00	0	0.00
ATT	13	0.17	20	1.40	12	0.41	10	0.81	0	0.00	9	2.45
ATC	219	2.81	22	1.54	76	2.59	26	2.11	7	3.00	2	0.54
ACG	33	0.50	2	0.15	30	0.92	13	1.37	0	0.00	1	0.89
ACA	4	0.06	5	0.38	11	0.34	8	0.84	0	0.00	2	0.38
ACT	9	0.14	28	2.15	15	0.46	2	0.21	0	0.00	3	1.33
ACC	218	3.30	17	1.31	75	2.29	15	1.58	12	4.00	3	1.33
TGG	67	1.00	4	1.00	24	1.00	14	1.00	1	1.00	1	1.00
TGT	4	0.11	3	1.00	1	0.10	1	0.33	0	0.00	0	0.00
TGC	72	1.89	3	1.00	20	1.90	5	1.67	1	2.00	1	2.00

Table 1. continued.

codon	chromosome[b]		pillin[c]		Tn501[d]		plasmid[e]		phage			
									Pf1[f]		Pf3[g]	
TAT	31	0.32	7	1.75	11	0.44	10	0.77	1	0.67	2	1.33
TAC	162	1.68	1	0.25	39	1.56	16	1.23	2	1.33	1	0.67
TTT	5	0.06	8	1.45	7	0.19	2	0.12	1	0.40	6	1.20
TTC	176	1.94	3	0.55	65	1.81	31	1.88	4	1.60	4	0.80
TTG	47	0.52	9	1.86	38	0.87	18	1.26	0	0.00	3	2.56
TTA	3	0.03	0	0.00	1	0.02	1	0.07	0	0.00	0	0.00
CTG	348	3.82	15	3.10	171	3.93	47	3.28	5	4.28	3	2.56
CTA	5	0.05	0	0.00	5	0.11	1	0.07	0	0.00	0	0.00
CTT	12	0.13	3	0.62	10	0.23	9	0.63	0	0.00	0	0.00
CTC	131	1.44	2	0.41	36	0.83	10	0.70	2	1.29	1	0.85
CAG	209	1.72	12	1.60	79	1.63	13	1.30	8	1.14	4	0.80
CAA	34	0.28	3	0.40	18	0.37	7	0.70	6	0.86	6	1.20
CAT	32	0.45	0	0.00	11	0.50	5	0.59	0	0.00	1	2.00
CAC	111	1.55	0	0.00	33	1.50	12	1.41	0	0.00	0	0.00
CCG	171	2.49	9	2.00	60	2.50	19	1.95	13	4.00	1	0.67
CCA	7	0.10	0	0.00	10	0.42	5	0.51	0	0.00	0	0.00
CCT	8	0.12	5	1.11	2	0.08	7	0.72	0	0.00	4	2.67
CCC	89	1.29	4	0.89	24	1.00	8	0.82	0	0.00	1	0.67

[a] The data given are the number of times a codon occurred in the examined sequences and the relative synonymous codon usage value (the observed frequency of a codon divided by the frequency expected if there is uniform usage within synonymous codon groups).
[b] Compiled from codon usage tables for 15 *P. aeruginosa* chromosomal genes excluding pillin.
[c] Compiled from codon usage tables for the pillin genes from *P. aeruginosa* strains PAO, PAK, and PA103.
[d] Compiled from codon usage tables for the *merA*, *merD*, *merP*, *merR*, *merT*, *tnpA* and *tnpR* genes of the *P. aeruginosa* transposon Tn501.
[e] Compiled from codon usage tables for the *Pae*R7 restriction/modification system encoded on pMG7 from *P. aeruginosa*.
[f] Compiled from the codon usage table for the DNA binding protein from bacteriophage Pf1.
[g] Compiled from codon usage tables for the DNA binding and major coat protein genes of bacteriophage Pf3.

The patterns of codon usage in the chromosomal genes excluding pillin and in the *P. aeruginosa* mercury resistance transposon Tn501 are similar (Table 1); this observation is confirmed by the low D^2 statistic, 0.354, for these two groups (Table 2). The values for rare codon frequency, codon preference statistic, P2 index, and nucleotide composition of the third codon position for the individual transposon and chromosomal genes are within the same ranges (Table 3). There is a slightly increased use of A and T and slightly decreased use of C in the third codon position in the transposon genes compared to the chromosomal genes excluding pillin (Table 3). The mean G+C content of the seven Tn501 genes is 64.9%.

The codon usage pattern present in the PaeR7 endonuclease and methylase genes encoded on pMG7 is less biased than the pattern observed in the *P. aeruginosa* chromosomal genes

Table 2. Correspond analysis of codon usage tables for *P. aeruginosa* chromosomal, phage, and transposon encoded genes and for the individual pillin genes from *P. aeruginosa* strains PAO, PAK, and PA103[a].

	chromosome[b]	Tn501[b]	plasmid[b]	phage[b] Pf1	phage[b] Pf3	pillin[b]	individual pillin genes			enteric[c]
							PAO	PAK	PA103	high
Tn501[b]	0.354									
plasmid[b]	1.938	1.322								
phage Pf1[b]	2.750	3.208	3.010							
phage Pf3[b]	9.370	7.927	5.128	8.493						
pillin[b]	9.530	8.235	4.631	7.567	3.517					
PAO pillin	11.274	*	*	*	*	1.338				
PAK pillin	10.041	*	*	*	*	0.906	3.948			
PA103 pillin	10.125	*	*	*	*	0.526	2.799	1.533		
enteric high[c]	4.184	3.645	3.548	4.146	7.024	3.552	3.810	5.224	4.343	
enteric low[c]	5.326	4.088	2.386	4.977	2.888	1.560	3.175	2.562	1.776	3.390

[a] Values reported are the D^2 statistic which was calculated with the UWGCG program Correspond (19).
[b] Composite codon usage tables for chromosome, Tn501, plasmid, phage, and pillin are as described in the legend to Table 1.
[c] Data from UWGCG GenDoc.Data file (19) taken from Grantham et al. (3).
* Not reported.

excluding pillin. The preference for G or C in the third codon position is equivalent. The P2 index values are 0.51 and 0.53, respectively, indicating no preference for codon-anticodon interactions of intermediate binding energy (Table 3). There is also an increased incidence of A and T in the third position compared to the chromosomal genes excluding pillin. The mean G+C content of the PaeR7 endonuclease and methylase genes is 60.8%.

We compared codon usage in the two *P. aeruginosa* phage for which nucleotide sequence data are available and found that the patterns were distinctly different from each other and in the case of Pf3 different from the *P. aeruginosa* chromosome (Tables 1, 2, and 3). Pf3 prefers T in the third codon position while Pf1 prefers C (Table 3). These differences can be attributed to the G+C content of these two phage: Pf1 is 65.2% G+C, which is very similar to the *P. aeruginosa* chromosome, and Pf3 is 46.9% G+C.

Reference codon usage table for *P. aeruginosa*

A reference codon usage table for *P. aeruginosa* (Table 4) was generated from the individual codon usage tables for the following chromosomal genes: *plcS*, *plcR*, *trpA*, *trpB*, *trpG*, *pmi*, *algD*, *toxA*, *toxR*, *amiE*, *recA*, *lasA*, *lasB*, the gene for azurin, and the gene for porin protein F. We excluded the pillin genes because of their extremely divergent G+C content. Plasmid, phage, and transposon encoded genes were excluded because of their presence on autonomous replicons or a mobile genetic element. However, the similarity of codon usage patterns between the Tn501 and chromosomal genes may warrant the inclusion of the Tn501 genes

Table 3. Codon bias indices for 28 *P. aeruginosa* chromosomal, transposon, plasmid and phage genes.

gene	no. of codons	rare codon frequency[a]	codon preference statistic[b]	P2 index[c]	nucleotide composition of third position (%)			
					G	A	T	C
chromosome								
algD	438	63.9	1.392/0.644	0.50	33.9	4.6	5.5	56.0
amiE	347	43.2	1.328/0.651	0.45	37.3	3.5	6.1	53.2
azurin	149	60.4	1.415/0.631	0.54	36.5	3.4	5.4	54.7
lasA	378	89.9	1.170/0.694	0.42	40.1	6.1	6.6	47.2
lasB	499	64.1	1.303/0.657	0.49	32.3	4.2	6.2	57.2
plcR	208	105.8	1.089/0.678	0.33	38.6	9.7	7.2	44.4
plcS	731	69.8	1.227/0.677	0.42	31.6	5.2	6.0	57.1
pmi	481	37.4	1.338/0.684	0.40	34.4	4.6	2.3	58.8
porin F	351	122.5	1.277/0.612	0.67	23.1	6.6	15.1	55.1
recA	347	46.1	1.433/0.649	0.44	33.8	6.1	5.8	54.3
toxR	225	151.1	1.043/0.692	0.42	33.3	9.3	11.6	45.8
toxA	639	32.9	1.297/0.713	0.35	36.1	3.9	4.4	55.6
trpA	268	52.2	1.316/0.713	0.35	34.5	6.4	2.2	56.9
trpB	402	27.4	1.369/0.690	0.41	33.2	3.9	2.9	59.8
trpG	201	119.4	1.129/0.681	0.43	43.5	4.5	11.0	41.0
pillin								
PAO	150	326.7	0.625/0.468	0.64	26.8	12.8	38.9	21.5
PAK	150	406.7	0.511/0.462	0.66	26.7	15.3	40.7	17.3
PA103	150	400.0	0.540/0.441	0.55	28.0	15.3	39.3	17.3
transposon Tn501								
merA	562	122.8	1.099/0.665	0.40	37.7	8.7	9.8	43.8
merD	122	139.3	1.057/0.698	0.41	46.7	8.2	9.0	36.1
merP	92	119.6	1.095/0.638	0.15	29.3	9.8	9.8	51.1
merR	61	229.5	0.925/0.653	0.29	36.1	11.5	8.2	44.3
merT	117	247.9	0.865/0.608	0.48	33.3	10.3	13.7	42.7
tnpA	989	70.8	1.252/0.678	0.41	42.2	7.0	6.4	44.4
tnpR	187	96.3	1.220/0.660	0.45	42.2	7.0	6.4	44.4
plasmid pMG7								
endonuclease	247	206.5	0.937/0.595	0.51	36.0	12.6	14.6	36.8
methylase	532	219.9	0.887/0.618	0.53	36.5	8.4	19.9	35.2
phage								
Pf3 CP	45	333.3	0.568/0.438	0.50	24.4	17.8	40.0	17.8
Pf3 DNA BP	79	354.4	0.544/0.437	0.57	26.6	10.1	38.0	25.3
Pf1 DNA BP	145	62.1	1.294/0.665	0.42	33.8	10.3	6.2	49.7

[a] Rare codons are those codons which represent less than 10% of the codons chosen from a specific synonymous codon group. The rare codon frequency is the number of codons observed in a gene adjusted to the number of rare codons which would occur per 1000 codons.

[b] The codon usage data file for calculating the codon preference statistic (46) was the *P. aeruginosa* reference codon usage table. Values shown are the codon preference statistic for the coding sequence of the examined gene and the codon preference statistic for a random nucleotide sequence of the same composition.

[c] P2 index was calculated as described by Gouy and Gautier (6).

in this reference codon usage table. Nevertheless, we chose a conservative approach until additional *P. aeruginosa* genes are sequenced and a more detailed study of codon usage by this organism can be made.

Table 4. Codon usage in *Pseudomonas aeruginosa*[a].

amino acid	codon	no.[b]	/1000[c]	fraction[d]	amino acid	codon	no.	/1000	fraction
Gly	GGG	31	5.47	0.06	Trp	TGG	67	11.83	1.00
Gly	GGA	18	3.18	0.04	End	TGA	12	2.12	0.80
Gly	GGT	58	10.24	0.11	Cys	TGT	4	0.71	0.05
Gly	GGC	401	70.81	0.79	Cys	TGC	72	12.71	0.95
Glu	GAG	210	37.08	0.62	End	TAG	1	0.18	0.07
Glu	GAA	130	22.96	0.38	End	TAA	2	0.35	0.13
Asp	GAT	46	8.12	0.14	Tyr	TAT	31	5.47	0.16
Asp	GAC	294	51.92	0.86	Tyr	TAC	162	28.61	0.84
Val	GTG	170	30.02	0.42	Leu	TTG	47	8.30	0.09
Val	GTA	28	4.94	0.07	Leu	TTA	3	0.53	0.01
Val	GTT	18	3.18	0.04	Phe	TTT	5	0.88	0.03
Val	GTC	188	33.20	0.47	Phe	TTC	176	31.08	0.97
Ala	GCG	183	32.32	0.31	Ser	TCG	89	15.72	0.27
Ala	GCA	19	3.36	0.03	Ser	TCA	3	0.53	0.01
Ala	GCT	43	7.59	0.07	Ser	TCT	3	0.53	0.01
Ala	GCC	350	61.80	0.59	Ser	TCC	93	16.42	0.28
Arg	AGG	12	2.12	0.03	Arg	CGG	49	8.65	0.14
Arg	AGA	3	0.53	0.01	Arg	CGA	11	1.94	0.03
Ser	AGT	12	2.12	0.04	Arg	CGT	34	6.00	0.10
Ser	AGC	135	23.84	0.40	Arg	CGC	235	41.50	0.68
Lys	AAG	196	34.61	0.89	Gln	CAG	209	36.91	0.86
Lys	AAA	25	4.41	0.11	Gln	CAA	34	6.00	0.14
Asn	AAT	18	3.18	0.08	His	CAT	32	5.65	0.22
Asn	AAC	202	35.67	0.92	His	CAC	111	19.60	0.78
Met	ATG	119	21.01	1.00	Leu	CTG	348	61.45	0.64
Ile	ATA	2	0.35	0.01	Leu	CTA	5	0.88	0.01
Ile	ATT	13	2.30	0.06	Leu	CTT	12	2.12	0.02
Ile	ATC	219	38.67	0.94	Leu	CTC	131	23.13	0.24
Thr	ACG	33	5.83	0.13	Pro	CCG	171	30.20	0.62
Thr	ACA	4	0.71	0.02	Pro	CCA	7	1.24	0.03
Thr	ACT	9	1.59	0.03	Pro	CCT	8	1.41	0.03
Thr	ACC	218	38.50	0.83	Pro	CCC	89	15.72	0.32

[a] Total number of codons = 5663.
[b] The number of times a codon occurred in the examined sequences.
[c] The number of times a specific codon would occur per 1000 codons.
[d] The ratio of the number of occurrences of a specific codon to the number of occurrences of all codons in the same synonymous codon group.

Analysis of codon usage in 15 *P. aeruginosa* chromosomal genes

The mean DNA base composition of the sequences included in the reference codon usage table was 65.2% G+C and the values for the individual genes ranged from 60.1% for porin protein F to 69.5% for *toxA*. These values are in agreement with the published value of 67.2% G+C for the *P. aeruginosa* genome (19). The G+C content of nucleotides 1, 2, and 3 of the coding triplets was 64.8%, 43.5%, and 88.7%, respectively. These values are similar to those

obtained for other organisms with a high G+C content such as *M. luteus* (16), *Thermus thermophilus* (49), and *Streptomyces* (13, 50).

In agreement with the translation framing code hypothesis of Trifonov (18), *P. aeruginosa* preferrentially uses G in the first codon position and nucleotides other than G in the second codon position. GNN codons are used 38.6% of the time while NGN codons are used 20.4% of the time. We analyzed the first 60 nucleotides of the 15 *P. aeruginosa* chromosomal genes for the G-non-G-N periodicity described by Trifonov (18) and found that this pattern was not present in 8 of the first 20 codons from these genes (data not shown). This observation may be due to the small number of genes examined.

P. aeruginosa, as *M. luteus* (16), exhibits a strong bias for cytosine over guanine in the wobble position. NNC codons are used 54.5% of the time, NNG codons 34.2%, NNT codons 6.0%, and NNA codons 5.3%. Cytosine is the preferred nucleotide for the third position in all synonymous coding groups. Exceptions to this empirical rule are leucine CTG and proline CCG and the equivalent use of valine GTG and GTC and serine TCG and TCC (Table 4).

In synonymous codon groups in which there are six choices (arginine, serine, and leucine), the preferred codons contain a G or C at two of the three positions. These preferred codons are CGC and CGG for arginine, AGC, TCC, and TCG for serine, and CTC and CTG for leucine (Table 4). The other choices which contain an A or T at two of the three positions are used less than 10% of the time (Table 4).

The following 26 codons are used rarely if at all by *P. aeruginosa* and represent less than 10% of the codons chosen from a specific synonymous codon group.

Gly	GGG	Arg	CGT	Cys	TGT
Gly	GGA	Ser	AGT	Leu	TTG
Val	GTA	Ser	TCA	Leu	TTA
Val	GTT	Ser	TCT	Leu	CTA
Ala	GCA	Asn	AAT	Leu	CTT
Ala	GCT	Ile	ATA	Phe	TTT
Arg	AGG	Ile	ATT	Pro	CCA
Arg	AGA	Thr	ACA	Pro	CCT
Arg	CGA	Thr	ACT		

With the exceptions of glycine GGG, arginine AGG, and leucine TGG these codons all end in A or T and include 5 of the 8 codons containing only A or T. The other three exclusively AT containing codons, lysine AAA, tyrosine TAT, and the termination codon TAA are infrequently used (Table 4). In the *M. luteus* streptomycin operon, these same codons are also rarely or never used (16). In *E. coli*, rare codons are recognized by the least abundant tRNAs (8-9). tRNA levels are unknown in *P. aeruginosa*; therefore, it is not known if these rarely used codons are the ones recognized by the less abundant tRNAs. However, the expression of pillin, phage Pf3, and several *E. coli* antibiotic resistance genes indicates that *P. aeruginosa* possesses a full complement of tRNAs.

Table 5. Correspond analysis of codon usage tables from 15 *P. aeruginosa* chromosomal genes[a].

	trpB	*pmi*	*amiE*	*plcS*	*trpA*	*lasA*	*recA*	azurin	*plcR*	*toxR*	*trpG*	porin F	*algD*	*lasB*
pmi	0.358													
amiE	0.399	0.637												
plcS	0.561	0.526	0.609											
trpA	0.624	0.632	0.971	0.721										
lasA	0.799	1.361	0.848	0.825	0.777									
recA	0.809	1.206	0.809	1.248	1.102	1.291								
azurin	0.843	0.926	0.744	1.094	1.286	1.614	1.285							
plcR	1.116	1.225	1.097	1.008	1.543	1.368	1.870	1.577						
toxR	1.715	2.262	1.749	1.260	1.944	1.267	2.289	2.451	1.208					
trpG	1.850	2.108	1.443	1.609	1.516	1.225	1.725	2.077	1.641	1.478				
porin F	1.565	2.154	2.059	2.143	1.638	1.815	1.536	1.275	3.007	2.779	2.518			
algD	2.270	2.407	2.263	2.414	2.519	2.788	2.654	2.697	3.049	3.585	3.340	3.498		
lasB	2.328	2.560	2.463	2.531	2.584	2.499	3.011	3.048	3.214	3.223	3.507	3.578	2.376	
toxA	2.476	2.791	2.653	2.424	2.735	2.632	3.288	3.457	3.347	3.122	3.752	4.065	2.564	0.256

[a] The values reported are the D^2 statistic which was calculated using the UWGCG program Correspond (19).

The preferred termination codon is TGA (Table 1). This is the same termination codon preferred by *M. luteus*, whereas *E. coli* and yeast prefer TAA (7, 16). The preferred initiation codon in *P. aeruginosa* is ATG as it is in *E. coli*, yeast, and *Bacillus*. This is in contrast to *M. luteus* and *Streptomyces* which frequently use GTG as an initiation codon (16, 50).

Correspond analysis of codon usage in 15 *P. aeruginosa* chromosomal genes

We used the program Correspond (19) to assess the degree of similarity in codon usage among the 16 *P. aeruginosa* chromosomal genes (Table 5). The structural genes for two iron-regulated secreted proteins, *toxA* and *lasB*, are most similar in codon usage with a D^2 value of 0.256. Correspond analysis identified a group of 5 genes (*trpB*, *pmi*, *amiE*, *plcS* and *trpA*) with very similar codon usage patterns; the Correspond statistic D^2 ranges from 0.358 to 0.971. When the individual codon usage tables for these genes were compared to a composite table for only these genes, the D^2 statistic is 0.44 or less (data not shown). The pattern of codon usage in the *P. aeruginosa* PAO pillin gene is the most dissimilar from other *P. aeruginosa* chromosomal genes with D^2 values ranging from 10.269 to 13.457 (Table 5). It should be noted that the D^2 value comparing the codon usage tables for high and low expressed genes from enteric bacteria is 3.390.

Lack of correlation between mRNA expressivity and codon usage biases in *P. aeruginosa*

Several codon bias indices have been used to predict the level of mRNA expressivity in *E. coli* and yeast (3-7, 47). To identify the differences in codon usage patterns among *P. aeruginosa* chromosomal genes, in addition to Correspond analysis we calculated the P2 index (6), the codon preference statistic (47), the rare codon frequency, and the nucleotide composition of the third codon position. The results are given in Table 3. C is the preferred nucleotide for the third codon position in all genes except *trpG* where C and G are used at equal frequencies in the third codon position. The P2 index for *P. aeruginosa* chromosomal genes (except pillin and porin protein F) does not show a preference for codons which form codon-anticodon interac-

tions of intermediate binding energy. Porin protein F, which is the major outer membrane protein and is expressed at high levels (51), has the highest frequency of rare codons. Codon usage in this gene shows a slight preference for T in the third position as indicated by a P2 index of 0.67 (Table 3). This appears to be the result of a reduction in usage of G in the third codon position rather than a preference for T over C. Data on the number of molecules of a particular protein per cell are not available for *P. aeruginosa*. However, the genes for porin protein F, RecA protein, and secreted proteins such as elastase, exotoxin A, and phospholipase C are expected to be highly expressed. Correspond analysis, rare codon frequency, codon preference statistic, and P2 index did not separate these genes into a group separate from genes expected to be expressed at lower levels such as *trpA*, *trpB*, and *trpG* (Tables 3 and 5). In *E. coli*, *trpA*, *trpB*, and *trpG* are expressed at low to intermediate levels and are easily separated from highly expressed genes by Correspondence analysis and by P2 index values (3, 6). Thus, the statistical indexes based on codon usage which predict mRNA expressivity in *E. coli* do not appear to predict mRNA expressivity in *P. aeruginosa* (Table 3).

Comparison of codon usage patterns in *P. aeruginosa*, *E. coli*, yeast, and *Bacillus*.

The pattern of codon usage in *P. aeruginosa* differs from that observed in *E. coli*, yeast and *Bacillus*. *E. coli* and yeast preferentially use those codons which form codon-anticodon interactions of intermediate binding energy and which also correspond to the most abundant tRNAs (4-5, 7-9). In *Bacillus*, a minimal codon bias is seen only in highly expressed genes (52). In contrast, *P. aeruginosa* preferentially uses those codons within a synonymous codon group with the strongest predicted codon-anticodon interaction. The preference for a particular codon correlates with the predicted binding energy of the codon-anticodon pair. This observation agrees with the pattern of codon usage observed in microorganisms with a high G+C content such as *M. luteus* (16), *Streptomyces* (13, 50), and *T. thermophilus* (49). *P. aeruginosa* preferentially uses cytosine for the third codon position.

Identification of protein coding regions

Codon usage tables are useful for identifying protein coding regions in a nucleotide sequence (47, 53), locating DNA sequencing errors (47, 53), and constructing oligonucleotide probes from amino acid sequences. For DNA of very high G+C content, sequencing ambiguities due to base stacking which result in band compression occur frequently and are often difficult to resolve. This table should be valuable for locating such ambiguities in other *P. aeruginosa* genes. An earlier version of the *P. aeruginosa* codon usage table was helpful in locating protein coding regions and DNA sequencing ambiguities in the *lasA* and *lasB* genes (37, 38). This reference codon usage table for *P. aeruginosa* (Table 4) is presented in a format that can be used as the data file for the UWGCG program CodonPreference (20, 47).

ACKNOWLEDGEMENTS

This work was partially supported by Public Health Service grants AI24423 and AI25669 from the National Institutes of Health.

*To whom correspondence should be addressed

REFERENCES

1. Grantham, R., Gautier, C., Gouy, M., Mercier,, R., and Pave, A. (1980) Nucl. Acids Res. **8**, r49-r62.
2. Grantham, R., Gautier, C., and Gouy, M. (1980) Nucl. Acids Res. **8**, 1893-1912.
3. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M., and Mercier, R. (1981) Nucl. Acids Res. **9**, r43-r74.
4. Bennetzen, J.L., and Hall, B.D. (1982) J. Biol. Chem. **257**, 3026-3031.
5. Grosjean, H., and Fiers, W. (1982) Gene **18**, 199-209.
6. Gouy, M., and Gautier, C. (1982) Nucl. Acids Res. **10**, 7055-7074.
7. Sharp, P.M., Tuohy, T.M.F., and Mosurski, K.R. (1986) Nucl. Acids Res. **14**, 5125-5143
8. Ikemura, T. (1981) J. Mol. Biol. **146**, 1-21.
9. Ikemura, T. (1982) J. Mol. Biol. **158**, 573-598.
10. Konigsberg, W., and Godson, G.N. (1983) Proc. Natl. Acad. Sci. **80**, 687-691.
11. Sharp, P. M., and Li, W.-H. (1986) Nucl. Acids Res. **14**, 7737-7749.
12. Bibb, M.J., Findlay, P.R., and Johnson, M.W. (1984) Gene **30**, 157-166.
13. Bibb, M.J., Bibb, M.J., Ward, J.M., and Cohen, S.N. (1985) Mol. Gen. Genet. **199**, 26-36.
14. Bernardi, G., and Bernardi, G. (1985) J. Mol. Evol. **22**, 363-365.
15. Muto. A., and Osawa, S. (1987) Proc. Natl. Acad. Sci. **84**, 166-169.
16. Ohama, T., Yamao, F., Muto, A., and Osawa, S. (1987) J. Bacteriol. **169**, 4770-4777.
17. Muto, A., Kawauchi, Y., Yamao, F., and Osawa, S. (1984) Nucl. Acids Res. **12**, 8209-8217.
18. Trifonov, E.N. (1987) J. Mol. Biol. **194**, 643-652.
19. Palleroni, N.J. (1984) In Kreig, N.R., and Holt, J.G., (eds) (1984) Bergey's Manual of Systematic Bacteriology, Williams and Wilkins, Baltimore, Vol. 1, p. 165.
20. Devereux, J., Haeberli, P., and Smithies, O. (1984) Nucl. Acids Res. **12**, 387-395.
21. Bilofsky, H. S., and Burks, C. (1988) Nucl. Acids Res. **16**, 1861-1863.
22. Cameron, G.N. (1988) Nucl. Acids Res. **16**, 1865-1867.
23. Gray, G.L. Smith, D.H., Baldridge, J.S., Harkins, R.N., Vasil, M.L., Chen, E.V., and Heyneker, H.L. (1984). Proc. Natl. Acad. Sci. **81**, 2645-2649.
24. Sastry, P.A, Finlay, B.B., Pasloske, B.L., Paranchych, W., Pearlstone, J.R., and Smillie, L.B. (1985) J. Bacteriol. **164**, 571-577.
25. Pasloske, B.L., Finlay, B.B., and Paranchych, W. (1985) FEBS Lett. **183**, 408-412.
26. Johnson, K., Parker, M.L., and Lory, S. (1986) J. Biol. Chem. **261**, 15703-15708.
27. Pritchard, A.E. and Vasil, M.L. (1986) J. Bacteriol. **167**, 291-298.
28. Hadero, A. and Crawford, I.P. (1986) Mol. Biol. Evol. **3**, 191-204.
29. Crawford, I.P., and Eberly, L. (1986) Mol. Biol. Evol. **3**, 436-448.
30. Darzins, A., Frantz, B. Vanags, R.I., and Chakrabarty, A.M. (1986) Gene. **42**, 293-302.
31. Deretic, V., Gill, J.F., and Chakrabarty, A.M. (1987) Nucl. Acids Res. **15**, 4567-4581.
32. Wozniak, D.J., Cram, D.C., Daniels, C.J., and Galloway, D.R. (1987) Nucl. Acids Res. **15**, 2123-2135.
33. Shen, B.-F., Tai, P.C., Pritchard, A.E., and Vasil, M.L. (1987) J. Bacteriol. **169**, 4602-4607.
34. Canters, G.W. (1987) FEBS Lett. **212**, 168-172.
35. Bramar, W.J., Charles, I.G., Matfield, M., Cheng-Pin, L., Drew, R.E., and P.H. Clarke. (1987) FEBS Lett. **215**, 291-294.
36. Sano, Y., and Kageyama, M. (1987) Mol. Gen. Genet. **208**, 412-419.
37. Duchene, M., Schweizer, A., Lottspeich, F., Krauss, G., Marget, M., Vogel, K., vonSpecht, B.-U., and Domdey, H. (1988) J. Bacteriol. **170**, 155-162.
38. Schad, P.A., and Iglewski, B.H. (1988) J. Bacteriol. **170**, 2784-2789.
39. Bever, R.A, and Iglewski, B.H. (1988). J. Bacteriol. **170**, 4309-4314.
40. Theriault, G., Roy, P.H., Howard, K.A., Benner, J.S., Brooks, J.E., Waters, A.F., and Gingeras, T.R. (1985) Nucl. Acids Res. **23**, 8441-8461.
41. Brown, N.L., Ford, S.J., Pridmore, R.D., and Fritzinger, D.C. (1983) Biochem. **22**, 4089-4095.
42. Diver, W.P., Grinsted, J., Fritzinger, D.C., Brown, N.L. Altenbuchner, J., Rogowsky, P., and Schmitt, R. (1983) Mol. Gen. Genet. **191**, 189-193.

43. Misra, T.K., Brown, N.L., Fritzinger, D.C., Pridmore, R.D., Barnes, W.M., Haberstroh, L., and Silver, S. (1984) Proc. Natl. Acad. Sci. **81**, 5975-5979.
44. Brown, N.L., Winnie, J.N., Fritzinger, D., and Pridmore, R.D. (1985) Nucl. Acids Res. **15**, 5657-5669.
45. Putterman, D.G., Casadevall, A., Boyle, P.D., Yang, H.-L., Frangione, B., and Day, L.A. (1984) **81**, 699-703.
46. Maeda, K., Kneale, G.G., Tsugita, A., Short, N.J., Perham, R.N., Hill, D.F., and Petersen, G.B. (1982) EMBO J. 1, 255-261.
47. Gribskov, M., Devereux, J., and Burgess, R.R. (1984) Nucl. Acids Res. **12**, 539-549.
48. Marrs, C.F., Schoolnik, G., Koomey, J.M., Hardy, J., Rothbard, J., and Falkow, S. (1985) J. Bacteriol. **163**, 132-139.
49. Kagawa, Y., Nojima, H., Nukiwa, N., Ishizuka, M., Nakajima, T., Yasuhara, T., Tanaka, T., and Oshima, T. (1984) J. Biol. Chem. **259**, 2956-2960.
50. Zalacain, M., Gonzalez, A., Guerrero, M.C., Mattaliano, R.J., Malpartida, F., and Jimenez, A. (1986) Nucl. Acids Res. **14**, 1565-1581.
51. Nicas, T.I., and Hancock, R.E.W. (1983) J. Bacteriol. **153**, 281-285.
52. Shields, D.C. and Sharp, P. M. (1987). Nucl. Acids Res. **15**, 8023-8040.
53. Statden, R., and McLachan, A.D. (1982) Nucl. Acids Res. **10**, 141-156.