CHEM-527   Introductory Biochemistry
Codon Utilization

1. The degeneracy of the genetic code has been viewed as a protective device in that mutations to synonymous codons result in no change in the amino acid sequence of a coded protein. Before nucleotide sequencing became possible, it was commonly hypothesized that there would be no preference among synonymous codons.   Now that a large number of genes have been sequenced it is possible to test the hypothesis.   Table 1 displays the codon frequencies for the glyceraldehyde-3-phosphate dehydrogenase gene from baker's yeast.   In Table 2, the nucleotide sequence for the yeast phosphoglycerate kinase mRNA in yeast is given along with its amino acid sequence.

   a)   Tabulate the codon utilization frequency for the phosphoglycerate kinase. (Go to http://www.kazusa.or.jp/codon/ and play with the program rather than do this by hand.). In comparison with the GAPDH mRNA, what patterns in codon utilization do you see? What about other glycolytic enzymes?

   b)   Speculate on the significance of the codon utilization patterns in yeast.   What would you predict for other genes in yeast?   (Consider that glyceraldehyde 3-phosphate dehydrogenase may constitute several percent of the total soluble protein in a glycolyzing yeast cell.)

   c)   For the mathematically inclined: Generate a mathematical expression which will reflect the degree of non-randomness in codon utilization.   Ideally when codons are used randomly in a gene the value would be 0 and would range up to 1.0 for gene where only one of several codons was used for each amino acid.   Calculate the values of this expression for the genes you looked at..

Table 1.   Codon Utilization Table for the Yeast (*Saccharomyces cerevisiae*) Gene Coding for Glyceraldehyde-3-Phosphate Dehydrogenase

				Codon Position 2									
			U			C			A			G		
		AA	Cdn	No	AA	Cdn	No	AA	Cdn	No	AA	Cdn	No	
Codon Position 1	U	PHE	UUU	0	SER	UCU	13	TYR	UAU	0	CYS	UGU	2	U
		PHE	UUC	10	SER	UCC	12	TYR	UAC	11	CYS	UGC	0	C
		LEU	UUA	0	SER	UCA	0	STOP	UAA	(1)	STOP	UGA	0	A
		LEU	UUG	21	SER	UCG	0	STOP	UAG	0	TRP	UGG	3	G
	C	LEU	CUU	0	PRO	CCU	0	HIS	CAU	0	ARG	CGU	0	U
		LEU	CUC	0	PRO	CCC	0	HIS	CAC	8	ARG	CGC	0	C
		LEU	CUA	0	PRO	CCA	12	GLN	CAA	5	ARG	CGA	0	A
		LEU	CUG	0	PRO	CCG	0	GLN	CAG	0	ARG	CGG	0	G
	A	ILE	AUU	9	THR	ACU	12	ASN	AAU	0	SER	AGU	0	U
		ILE	AUC	11	THR	ACC	12	ASN	AAC	12	SER	AGC	0	C
		ILE	AUA	0	THR	ACA	0	LYS	AAA	1	ARG	AGA	11	A
		MET	AUG	7	THR	ACG	0	LYS	AAG	25	ARG	AGG	0	G
	G	VAL	GUU	22	ALA	GCU	25	ASP	GAU	9	GLY	GGU	25	U
		VAL	GUC	15	ALA	GCC	7	ASP	GAC	16	GLY	GGC	0	C
		VAL	GUA	0	ALA	GCA	0	GLU	GAA	12	GLY	GGA	1	A
		VAL	GUG	0	ALA	GCG	0	GLU	GAG	2	GLY	GGG	0	G

Codon Position 3

Table 2. The Amino Acid Sequence for Yeast Phosphoglycerate Kinase and the Coding (mRNA) Sequence for its mRNA

```
  M   S   L   S   S   K   L   S   V   Q   D   L   D   L   K   D   K
AUG-UCU-UUA-UCU-UCA-AAG-UUG-UCU-GUC-CAA-GAU-UUG-GAC-UUG-AAG-GAC-AAG-

  R   V   F   I   R   V   D   F   N   V   P   L   D   G   K   K   I
CGU-GUC-UUC-AUC-AGA-GUU-GAC-UUC-AAC-GUC-CCA-UUG-GAC-GGU-AAG-AAG-AUC-

  T   S   N   Q   R   I   V   A   A   L   P   T   I   K   Y   V   L
ACU-UCU-AAC-CAA-AGA-AUU-GUU-GCU-GCU-UUG-CCA-ACC-AUC-AAG-UAC-GUU-UUG-

  E   H   H   P   R   Y   V   V   L   A   S   H   L   G   R   P   N
GAA-CAC-CAC-CCA-AGA-UAC-GUU-GUC-UUG-GCU-UCU-CAC-UUG-GGU-AGA-CCA-AAC-

  G   E   R   N   E   K   Y   S   L   A   P   V   A   K   E   L   Q
GGU-GAA-AGA-AAC-GAA-AAA-UAC-UCU-UUG-GCU-CCA-GUU-GCU-AAG-GAA-UUG-CAA-

  S   L   L   G   K   D   V   T   F   L   N   D   C   V   G   P   E
UCA-UUG-UUG-GGU-AAG-GAU-GUC-ACC-UUC-UUG-AAC-GAC-UGU-GUC-GGU-CCA-GAA-

  V   E   A   A   V   K   A   S   A   P   G   S   V   I   L   L   E
GUU-GAA-GCC-GCU-GUC-AAG-GCU-UCU-GCC-CCA-GGU-UCC-GUU-AUU-UUG-UUG-GAA-

  N   L   R   Y   H   I   E   E   E   G   S   R   K   V   D   G   Q
AAC-UUG-CGU-UAC-CAC-AUC-GAA-GAA-GAA-GGU-UCC-AGA-AAG-GUC-GAU-GGU-CAA-

  K   V   K   A   S   K   E   D   V   Q   K   F   R   H   E   L   S
AAG-GUC-AAG-GCU-UCC-AAG-GAA-GAU-GUU-CAA-AAG-UUC-AGA-CAC-GAA-UUG-AGC-

  S   L   A   D   V   Y   I   N   D   A   F   G   T   A   H   R   A
UCU-UUG-GCU-GAU-GUU-UAC-AUC-AAC-GAU-GCC-UUC-GGU-ACC-GCU-CAC-AGA-GCU-

  H   S   S   M   V   G   F   D   L   P   Q   R   A   A   G   F   L
CAC-UCU-UCU-AUG-GUC-GGU-UUC-GAC-UUG-CCA-CAA-CGU-GCU-GCC-GGU-UUC-UUG-

  L   E   K   E   L   K   Y   F   G   K   A   L   E   N   P   T   R
UUG-GAA-AAG-GAA-UUG-AAG-UAC-UUC-GGU-AAG-GCU-UUG-GAG-AAC-CCA-ACC-AGA-

  P   F   L   A   I   L   G   G   A   K   V   A   D   K   I   Q   L
CCA-UUC-UUG-GCC-AUC-UUA-GGU-GGU-GCC-AAG-GUU-GCU-GAC-AAG-AUU-CAA-UUG-

  I   D   N   L   L   D   K   V   D   S   I   I   I   G   G   G   M
AUU-GAC-AAC-UUG-UUG-GAC-AAG-GUC-GAC-UCU-AUC-AUC-AUU-GGU-GGU-GGU-AUG-

  A   F   T   F   K   K   V   L   E   N   T   E   I   G   D   S   I
GCU-UUC-ACC-UUC-AAG-AAG-GUU-UUG-GAA-AAC-ACU-GAA-AUC-GGU-GAC-UCC-AUC-

  F   D   K   A   G   A   E   I   V   P   K   L   M   E   K   A   K
UUC-GAC-AAG-GCU-GGU-GCU-GAA-AUC-GUU-CCA-AAG-UUG-AUG-GAA-AAG-GCC-AAG-

  A   K   G   V   E   V   V   L   P   V   D   F   I   I   A   D   A
GCC-AAG-GGU-GUC-GAA-GUC-GUC-UUG-CCA-GUC-GAC-UUC-AUC-AUU-GCU-GAU-GCU-

  F   S   A   D   A   N   T   K   T   V   T   D   K   E   G   I   P
UUC-UCU-GCU-GAU-GCC-AAC-ACC-AAG-ACU-GUC-ACU-GAC-AAG-GAA-GGU-AUU-CCA-
```

```
    A   G   W   Q   G   L   D   N   G   P   E   S   R   K   L   F   A
GCU-GGC-UGG-CAA-GGG-UUG-GAC-AAU-GGU-CCA-GAA-UCU-AGA-AAG-UUG-UUU-GCU-

    A   T   V   A   K   A   K   T   I   V   W   N   G   P   P   G   V
GCU-ACU-GUU-GCA-AAG-GCU-AAG-ACC-AUU-GUC-UGG-AAC-GGU-CCA-CCA-GGU-GUU-

    F   E   F   E   K   F   A   A   G   T   K   A   L   L   D   E   V
UUC-GAA-UUC-GAA-AAG-UUC-GCU-GCU-GGU-ACU-AAG-GCU-UUG-UUA-GAC-GAA-GUU-

    V   K   S   S   A   A   G   N   T   V   I   I   G   G   G   D   T
GUC-AAG-AGC-UCU-GCU-GCU-GGU-AAC-ACC-GUC-AUC-AUU-GGU-GGU-GGU-GAC-ACU-

    A   T   V   A   K   K   Y   G   V   T   D   K   I   S   H   V   S
GCC-ACU-GUC-GCU-AAG-AAG-UAC-GGU-GUC-ACU-GAC-AAG-AUC-UCC-CAU-GUC-UCU-

    T   G   G   G   A   S   L   E   L   L   E   G   K   E   L   P   G
ACU-GGU-GGU-GGU-GCU-UCU-UUG-GAA-UUA-UUG-GAA-GGU-AAG-GAA-UUG-CCA-GGU-

    V   A   F   L   S   E   K   K
GUU-GCU-UUC-UUA-UCC-GAA-AAG-AAA-UAA
```