

---

# Synonymous codon usage in different protein secondary structural classes of human genes: Implication for increased non-randomness of GC<sub>3</sub> rich genes towards protein stability

PAMELA MUKHOPADHYAY, SURAJIT BASAK and TAPASH CHANDRA GHOSH\*

Bioinformatics Centre, Bose Institute, P 1/12, CIT Scheme VII M, Kolkata 700 054, India

\*Corresponding author (Fax, 91-33-2355 3886; Email, tapash@boseinst.ernet.in)

The relationship between the synonymous codon usage and different protein secondary structural classes were investigated using 401 *Homo sapiens* proteins extracted from Protein Data Bank (PDB). A simple Chi-square test was used to assess the significance of deviation of the observed and expected frequencies of 59 codons at the level of individual synonymous families in the four different protein secondary structural classes. It was observed that synonymous codon families show non-randomness in codon usage in four different secondary structural classes. However, when the genes were classified according to their GC<sub>3</sub> levels there was an increase in non-randomness in high GC<sub>3</sub> group of genes. The non-randomness in codon usage was further tested among the same protein secondary structures belonging to four different protein folding classes of high GC<sub>3</sub> group of genes. The results show that in each of the protein secondary structural unit there exist some synonymous family that shows class specific codon-usage pattern. Moreover, there is an increased non-random behaviour of synonymous codons in sheet structure of all secondary structural classes in high GC<sub>3</sub> group of genes. Biological implications of these results have been discussed.

[Mukhopadhyay P, Basak S and Ghosh T C 2007 Synonymous codon usage in different protein secondary structural classes of human genes: Implication for increased non-randomness of GC<sub>3</sub> rich genes towards protein stability; *J. Biosci.* 32 947–963]

---

## 1. Introduction

Genetic code is degenerate and not all synonymous codons are used with equal frequencies. The non-random use of synonymous codons creates codon usage bias that act in a species-specific way (Grantham *et al* 1980). The factors responsible for codon usage bias in the coding sequences includes (i) diversity in the (G+C)% at the third codon position (Alvarez *et al* 1994), (ii) abundance of t-RNA molecule (Ikemura 1985), (iii) overall base composition of genes (Ellis and Morrison 1995), (iv) differences in the expression level of the genes (Pouwels and Leunissen 1994), (v) in the cellular location of the genes in the genome (Chiappello *et al* 1999), (vi) optimal growth temperature (Lynn *et al* 2002; Basak *et al* 2004; Basak and Ghosh 2006) and (vii) protein secondary structures (Kahali *et al* 2007; Adzhubei *et al* 1996). Emphasizing on individual codons,

earlier worker demonstrated that first, second and third positions of the codon has been associated respectively with the biosynthetic pathway, hydrophobicity pattern, and the alpha helix or beta strand forming potentiality of the coded amino acid (Volkenstein 1966; Taylor and Coates 1989; Siemion and Siemion 1994). It is also demonstrated that first and second position of the codons are the structure determining positions whereas the third position of the codon is the species determining position (Majumdar *et al* 1999). For the respective secondary structural units, protein alpha helices are preferentially coded by translationally fast mRNA regions while beta strands and coils are preferentially coded by slow mRNA regions (Thanaraj and Argos 1996 a,b). However, by comparing three-dimensional structures of proteins of *Escherichia coli* and human with their corresponding mRNA sequences came to conclusion that species-specific correlation exists between the use of two

**Keywords.** Aggregation reaction; codon usage; non-randomness; protein folding; protein secondary structure

synonymous codons and protein secondary structural units (Oresic and Shalloway 1998). On contrary to this hypothesis Tao and Dafu (1998) found no significant correlation between the synonymous codon usage and the protein secondary structural units in *E. coli* proteins. Working on the same hypothesis with *E. coli* and *Homo sapiens* Gu *et al* (2004) found no significant correlation between the use of synonymous codons and protein secondary structural class in *H. sapiens*. However, compositional heterogeneity in *H. sapiens* genes has been completely ignored by Gu *et al* (2004) in their analysis.

In the present study a simple chi-square test was performed to assess the significance of codon usage among four different (all-alpha, all-beta, alpha+beta, alpha/beta) secondary structural classes in *H. sapiens* genes. Same analysis was performed after partitioning the genes in three different groups according to their GC<sub>3</sub> levels to remove any noise due to the compositional bias on codon usage. We showed that there exist significant correlation between synonymous codon usage and secondary structural class in *H. sapiens* and the non-randomness in synonymous codon usage further increased in genes having higher GC<sub>3</sub> levels. Biological implications regarding the interrelationships between GC richness and nonrandom use of synonymous codons have been discussed.

## 2. Materials and methods

A dataset of 401 *H. sapiens* protein sequences were collected from Protein Data Bank (PDB). The extracted protein sequences were classified into four secondary structural classes (all-alpha, all-beta, alpha+beta and alpha/beta) from the structural information provided by ASTRAL Structural Classification of Proteins (SCOP) database 1.61 (Berman *et al* 2000; Brenner *et al* 2000; Conte *et al* 2002). The respective four secondary structural classes include 86 all-alpha proteins, 103 all-beta proteins, 119 alpha+beta proteins and 93 alpha/beta proteins. Further the mutated proteins and the same protein that have been classified to more than one class by SCOP database are removed from our analysis. This accounts for lower number of *H. sapiens* proteins in our datasets as compared to 563 *H. sapiens* proteins being collected by Gu *et al* (2004). TBLASTN program against the 'nr' database was used to retrieve the corresponding coding sequences for the 401 *H. sapiens* proteins. The amino acid sequences, which have 100% similarity scores with the 'nr' sequence database are only chosen for avoiding any ambiguity of one to one correspondence between the amino acid and the codon.

The significance of deviation of the observed and expected frequencies of codons for the 59 synonymous codons was tested by chi-square test for the individual codon families. The expected frequencies were calculated with an

assumption that the bases were randomly associated to form codons (Zhang *et al* 1991; Gupta *et al* 2000). For example, there were 120 U, 150 C, 180 A and 210 G, for a sum of 660 total bases. The expected number for a codon was calculated from the probability that any base will occur at a specific position of the codon. In the calculation, the base frequencies were used as the probabilities. For example, the probability that a base occurs at a specific position of the codon is  $120 \div 660 = 0.181818$  for U,  $150 \div 660 = 0.227272$  for C,  $180 \div 660 = 0.272727$  for A and  $210 \div 660 = 0.318181$  for G. The probability at which a codon is expected to occur is the product of the probabilities of the bases in the codon. The probability of the AGG (Arg) codon, for example, is calculated as  $0.272727 \times 0.318181 \times 0.318181 = 0.02761$ . Since a total of 220 codons were counted, the expected number of AGG codon is therefore  $220 \times 0.02761 = 6.0742$ . The expected frequencies of individual codon have been calculated from the overall base frequencies of the genes.

Considering the compositional heterogeneity of *H. sapiens*, the coding sequences were partitioned into three groups according to GC<sub>3</sub> level: low (0-46.5), midrange (46.6-64.7) and high (64.8-100). These GC<sub>3</sub> boundaries correspond to those of the genes as distributed in the isochore of the human genome (Arhondakis *et al* 2004).

The secondary structural assignments of the individual residues for the coding sequences in four different secondary structural classes were done by Database of Secondary Structure Assignment (DSSP) program (Kabsch and Sander *et al* 1983). The alpha helices are annotated by H and G in the DSSP file, beta-sheets by E and B and coils by the rest.

## 3. Results and discussion

Observed and expected frequencies of codon for four different protein classes (all-alpha, all-beta, alpha+beta, alpha/beta) have been tabulated in table 1. The significance of deviation of the observed and expected frequencies of 59 codons in individual synonymous families was tested by chi-square test. At the level of individual synonymous family it was observed that 5 synonymous families comprising 19 codons have significant non-random codon distribution in all four protein classes whereas in 11 synonymous families comprising 32 codons have displayed random distribution of codon frequencies among the four protein classes.

Gu *et al* (2004) while performing variance analysis on 563 *H. sapiens* proteins, observed no significant difference in synonymous codon usage in different *H. sapiens* protein secondary structural classes. Therefore, they claimed that synonymous codon usage is not related to protein secondary structural classes in *H. sapiens*. But, the present analysis shows clear evidence that the synonymous codon usage for at least 2 synonymous families representing threonine and

proline is related to the protein secondary structural classes in *H. sapiens*.

The results seem to be in apparent discrepancy with the conclusions reached in previously published paper of Gu *et al* (2004). In this regard it is worthwhile to mention that the inclusion of mutated proteins and the proteins classified to more than one class by SCOP database in the dataset of Gu *et al* might result non-significant difference in synonymous codon usage in different *H. sapiens* protein secondary structural classes. Moreover, we have analysed our dataset at the level of individual synonymous codon families to show the relationship between synonymous codon usages with different protein secondary structural classes.

### 3.1 Non randomness in synonymous codon usage in three groups of genes classified according to GC<sub>3</sub> level

In the vertebrate genome, G+C content at third codon position is the causative factor of synonymous codon usage (Ikemura 1985; Karlin and Mrazek 1996). Thus considering the compositional heterogeneity of vertebrate genome, such as human; and to minimize the influence of GC<sub>3</sub> composition on synonymous codon usage, we have classified the genes of each protein class into three groups according to their GC<sub>3</sub> values (Arhondakis *et al* 2004).

Observed and expected frequencies of codons for four different gene classes having high GC<sub>3</sub> values have been tabulated in table 2. Chi-square test of observed and expected frequencies of all the codons on the basis of their synonymous group reveals that 8 synonymous families comprising 27 codons have non-random codon distribution whereas 1 synonymous family comprising 2 codons have displayed random distribution of codon frequencies among the four protein classes. In other 9 synonymous families the distribution pattern of codon is markedly different in four protein classes. However, chi-square test of observed and expected frequencies of all the codons for four different protein classes (table 3) having low GC<sub>3</sub> values show that 10 synonymous families comprising 28 codons have random codon distribution in all the four protein classes, whereas 3 synonymous families comprising 12 codons show non-random distribution of codon frequencies among the four protein classes. In other 5 synonymous families comprising 17 codons display distribution pattern of codons not same in all the four protein classes.

In the intermediate GC<sub>3</sub> range, there are 9 synonymous families comprising 24 codons have random codon distribution in all the protein classes. However, 5 families comprising 19 codons show non-random distribution of codon frequencies among the protein classes. The distribution pattern of 4 synonymous families is not same in all the four protein classes (data not shown).

Considering the deviation of observed frequency from expected frequency of different synonymous groups of codons in each protein class, it is clear from tables 2 and 3 that significant deviation between observed and expected frequencies have been increased in all protein classes having high value of GC<sub>3</sub>, compared to the protein classes having lower value of GC<sub>3</sub>. Hence it is reasonable to conclude that for human genome, relationship between synonymous codon usage and different protein classes are stronger in those genes having higher level of GC<sub>3</sub>. The folding type structural information contained in mRNA sequences is found to increase in genes containing high level of GC<sub>3</sub>.

To investigate the contribution of respective secondary structural units in determining the folding type structural information, genes with higher level of GC<sub>3</sub> are further analysed to account for increased non-randomness in the respective secondary structural units of four different protein folding classes.

### 3.2 Relationship between synonymous codon usage and secondary structural unit in different folding classes

The respective secondary structural units of four different protein classes are separated by using DSSP program. Observed and expected frequencies of codons for respective sheet portion for four different protein classes at higher level of GC<sub>3</sub> have been tabulated in table 4. Chi-square test of observed and expected frequencies of all the codons on the basis of their synonymous group reveals that 11 synonymous families comprising 39 codons show non-random codon distribution in all four protein classes whereas 7 synonymous families comprising 20 codons display distribution pattern that is different in all four protein classes. The codons contributing to the non-random distribution in 7 synonymous families display class-specific codon usage. However chi-square test of observed and expected frequencies of all the 59 codons on the basis on their synonymous group in helix portion for four different protein classes at higher level of GC<sub>3</sub> (table 5) reveals that 10 synonymous families comprising 33 codons show non-random codon distribution in all four protein classes, whereas 1 synonymous family comprising 2 codons have random codon distribution. In the rest 7 synonymous families comprising 24 codons have codon distribution different in all four protein classes.

The observed and expected frequencies of codon for coil portion of four different protein classes in higher level of GC<sub>3</sub> has been tabulated in table 6. The Chi-square test between observed and expected frequencies of 59 codons on the basis of their synonymous group reveals that 6 synonymous families comprising 19 codons show non-random codon distribution in all four protein classes, whereas 6 synonymous families comprising 20 codons have random codon distribution. In other 6 synonymous

**Table 1.** The significance of deviation between observed and expected frequencies of codons at the level of individual synonymous codon families in four different protein secondary structural classes

Amino acid	Codons	All-alpha				All-beta				Alpha+beta				Alpha/Beta		
		Exp- frequency	Obs- frequency	Significance level	Exp- frequency	Obs- frequency	Significance level	Exp- frequency	Obs- frequency	Significance level	Exp- frequency	Obs- frequency	Significance level	Exp- frequency	Obs- frequency	Significance level
Ala	gca	27.65	22.4(253)	$P < 0.01$	25.27	18.2(187)	$P < 0.01$	27.11	22.9(327)	$P < 0.01$	25.46	21.3(473)	$P < 0.01$	24.88	43.6(968)	
	gcc	24.36	39.8(441)		26.38	45.5(480)		24.14	39.4(538)		24.88	43.6(968)		27.17	8.9(195)	
	geg	25.62	9.3(94)		27.08	12(126)		26.08	9.3(119)		22.47	26.3(583)		52.54	53.8(296)	NS
Cys	get	22.35	28.5(320)		21.25	24.4(262)		22.64	28.3(402)		47.45	46.1(253)		52.54	53.8(296)	NS
	tgc	52.15	50.7(146)	NS	55.38	61(258)	NS	51.60	50.9(242)	NS	47.45	46.1(253)		52.54	53.8(296)	NS
Asp	tgt	47.84	49.3(143)		44.61	39(162)		48.39	49.1(245)		52.54	53.8(296)		47.45	46.1(253)	
	gac	52.15	53.9(472)	NS	55.38	59.7(588)	NS	51.60	51.9(617)	NS	52.54	53.8(296)		47.45	46.1(253)	
Glu	gat	47.84	46.1(409)		44.61	40.3(385)		48.39	48.1(594)		52.54	53.8(296)		47.45	46.1(253)	
	gaa	51.90	43(636)	NS	48.27	40.6(453)	NS	50.96	43.9(648)	NS	48.37	37.9(758)	NS	52.54	53.8(296)	NS
Phe	gag	48.09	57(705)		51.72	59.4(677)		49.03	56.1(795)		51.62	62.1(1250)		52.54	53.8(296)	
	ttc	52.15	52.4(328)	NS	55.38	62.3(435)	NS	51.60	48.6(443)	NS	52.54	53.8(296)		47.45	46.1(253)	
Gly	ttt	47.84	47.6(298)		44.61	37.7(263)		48.39	51.4(480)		47.45	46.9(614)		47.45	46.9(614)	
	gga	27.65	28(194)	NS	25.27	26.5(375)	NS	27.11	29(440)	NS	25.48	23.8(514)	NS	24.88	35(751)	
His	ggc	24.36	34.6(250)		26.38	34.9(527)		24.14	31.7(444)		24.88	35(751)		27.17	25.5(554)	
	ggg	25.62	22.3(160)		27.08	26.5(382)		26.06	20.3(297)		27.17	25.5(554)		22.47	15.8(341)	
Ile	ggt	22.35	17.2(125)		21.25	12.1(176)		22.64	19(285)		22.47	15.8(341)		22.47	15.8(341)	
	cac	52.15	57.7(234)	NS	55.38	64(336)	NS	51.60	53.6(293)	NS	52.54	53.8(296)	NS	47.45	46.9(614)	NS
Leu	cat	47.84	42.3(171)		44.61	36(176)		48.39	46.4(265)		47.45	46.9(614)		47.45	46.9(614)	
	ata	37.17	14.4(116)	$P < 0.01$	34.88	16(136)	$P < 0.001$	36.68	15.4(175)	$P < 0.01$	34.97	14(211)	$P < 0.01$	34.16	50.4(768)	
Lys	atc	32.76	48.8(386)		36.18	52.7(455)		32.67	48.7(541)		34.16	50.4(768)		30.85	35.7(544)	
	att	30.05	36.8(296)		29.14	31.3(265)		30.63	35.8(406)		30.85	35.7(544)		48.37	37.2(706)	NS
Leu	aaa	51.90	43(470)	NS	48.27	39.2(398)	NS	50.96	43.6(641)	NS	48.37	37.2(706)	NS	51.62	82.8(1208)	
	aag	48.09	57(626)		51.72	60.8(635)		49.03	58.4(817)		51.62	82.8(1208)		17.25	5.3(160)	$P < 0.001$
Leu	cta	18.57	6.5(120)	$P < 0.01$	17.78	5.8(86)	$P < 0.001$	18.09	7.6(145)	$P < 0.01$	17.25	5.3(160)	$P < 0.001$	18.86	18.5(544)	
	ctc	16.36	18(334)		18.55	20.5(322)		16.11	18.3(340)		18.86	18.5(544)		18.41	44.2(1297)	
Leu	ctg	17.20	39.5(714)		19.04	46.2(730)		17.40	39.7(727)		18.41	44.2(1297)		15.23	12.4(362)	
	ctt	15.01	14(259)		14.94	10.4(168)		15.10	14.7(290)		15.23	12.4(362)		15.58	6.4(191)	
Leu	fta	17.04	8.1(148)		14.32	5.7(82)		16.96	7.4(146)		15.58	6.4(191)		16.63	13.2(388)	
	ttg	15.78	14(257)		15.34	11.4(195)		16.32	12.3(237)		16.63	13.2(388)				

Asn	aac	52.15	57.2(377)	NS	55.38	56.9(449)	NS	51.60	55(477)	NS	52.54	56(704)	NS
	aat	47.84	42.8(280)		44.61	43.1(328)		48.39	45(401)		47.45	42(508)	
Pro	cca	27.65	26.3(165)	NS	25.27	23.4(244)	NS	27.11	30.7(336)	$P < 0.05$	25.46	28.6(436)	$P < 0.05$
	ccc	24.36	35.7(218)		26.38	35.9(381)		24.14	29.5(305)		24.88	32.6(490)	
	ccg	25.62	11.9(72)		27.08	12.7(132)		26.08	9.7(100)		27.17	10.5(155)	
	cct	22.35	26(162)		21.25	28(297)		22.64	30.1(330)		22.47	28.1(418)	
Gln	caa	51.90	27.2(213)	$P < 0.001$	48.27	23.6(182)	$P < 0.001$	50.96	27.5(255)	$P < 0.001$	48.37	25.1(292)	$P < 0.001$
	cag	48.09	72.8(578)		51.72	76.4(615)		49.03	72.5(643)		51.82	74.9(873)	
Arg	aga	19.55	19.2(157)	NS	16.12	22.2(179)	NS	19.06	22.1(259)	NS	16.93	19.5(278)	NS
	agg	18.12	21.9(179)		17.27	21.9(199)		18.33	17.8(207)		18.07	18.8(266)	
	cga	17.23	11.3(89)		16.83	8.9(81)		16.97	13.1(156)		16.54	11.4(166)	
	cgc	15.18	17.9(142)		17.57	21(209)		15.11	19(220)		16.16	20.1(293)	
	cgg	15.96	22.4(177)		18.03	17.2(167)		16.33	19.8(218)		17.66	21.4(311)	
	cgt	13.93	7.3(60)		14.15	8.8(87)		14.17	8.2(96)		14.80	8.8(127)	
Ser	agc	19.71	24.6(245)	NS	20.36	26.9(405)	NS	19.45	23.1(313)	NS	19.41	22.6(408)	NS
	agt	18.08	15.1(150)		18.40	13.4(193)		18.24	16.4(224)		17.53	15.4(278)	
	tea	17.20	15.2(151)		15.98	12.4(187)		16.89	15.6(211)		16.05	12.7(231)	
	tec	15.15	22.3(223)		18.88	24.6(383)		15.04	21.6(297)		15.88	22.6(407)	
	tgg	15.93	6(59)		17.12	5.4(83)		16.25	4.8(64)		17.13	5.7(104)	
	tct	13.90	16.9(169)		13.43	17.4(250)		14.10	18.5(256)		14.16	21.2(384)	
Thr	aca	27.65	29.9(248)	NS	25.27	25.6(275)	$P < 0.025$	27.11	27.8(319)	$P < 0.05$	25.46	27.3(414)	$P < 0.05$
	acc	24.36	36(300)		26.36	40.2(454)		24.14	37.4(412)		24.88	36.2(552)	
	acg	25.62	11.4(94)		27.08	11.3(134)		26.08	12(134)		27.17	12.1(185)	
	act	22.35	22.6(190)		21.25	22.8(253)		22.64	22.7(262)		22.47	24.4(366)	
Val	gta	27.65	12.2(116)	$P < 0.01$	25.27	8.1(111)	$P < 0.001$	27.11	12.9(178)	$P < 0.025$	25.46	10.4(217)	$P < 0.01$
	gtc	24.38	22.3(212)		28.38	26.2(371)		24.14	21.2(294)		24.88	21.2(440)	
	gtg	25.62	47(445)		27.08	50.8(727)		26.08	46.1(596)		27.17	50.4(1034)	
	gtt	22.35	18.8(177)		21.25	14.8(215)		22.64	19.9(277)		22.47	17.9(374)	
Tyr	tac	52.15	55.8(251)	NS	55.38	59.8(360)	NS	51.60	56.7(461)	NS	52.54	54.3(528)	NS
	tat	47.84	44.4(205)		44.61	40.2(251)		48.39	43.3(361)		47.45	45.7(445)	

The values in parenthesis indicate the number of codon in the dataset.

**Table 2.** The significance of deviation between observed and expected frequencies of 59 codons at the level of individual synonymous families for the genes having high GC<sub>3</sub> values among four different protein secondary structural classes

Amino-acids	Codons	All-alpha(high)				All-beta(high)				Alpha+beta(high)				Significance level
		Expected	Observed	Significance level	Expected	Observed	Significance level	Expected	Observed	Significance level	Expected	Observed	Significance level	
Ala	gca	22.81	12.5(60)	$P < 0.01$	21.09	10.7(63)	$P < 0.01$	22.95	14(63)	$P < 0.001$	21.47	13.9(143)	$P < 0.001$	
	gcc	29.46	55.2(266)		30.55	54.8(324)		29.06	57(229)		29.18	55.6(580)		
Cys	geg	28.77	16.6(72)		30.26	18.6(109)		29.18	14(51)		29.93	11.6(121)		
	gct	18.94	15.6(75)		18.09	15.9(91)		18.78	14.9(59)		19.39	18.9(201)		
Asp	tgc	60.85	64(70)	NS	62.80	70.9(161)	NS	60.73	73.2(101)	NS	60.06	65.4(161)	NS	
	tgt	39.14	36(40)		37.19	29.1(66)		39.26	26.8(41)		39.93	34.6(83)		
Glu	gac	60.86	74(255)	$P < 0.05$	62.80	75(353)	NS	60.73	72.3(240)	NS	60.06	70.3(504)	NS	
	gat	39.13	26(88)		37.19	25(120)		39.26	27.7(101)		39.93	29.7(211)		
Phe	gaa	44.23	17.4(80)	$P < 0.001$	41.07	19.9(101)	$P < 0.01$	44.02	19.4(83)	$P < 0.001$	41.77	19(158)	$P < 0.001$	
	gag	55.76	82.6(378)		58.92	80.1(402)		55.97	80.6(321)		58.22	81(670)		
Gly	ttc	60.86	68.2(175)	NS	62.81	80.7(256)	$P < 0.01$	60.73	69.3(173)	NS	60.06	68.5(367)	NS	
	ttt	39.13	31.8(82)		37.18	19.3(64)		39.26	30.7(78)		39.93	31.5(167)		
His	gga	22.81	12(43)	$P < 0.005$	21.09	15.6(126)	NS	22.95	12(55)	NS	21.47	12.5(124)	NS	
	ggc	29.46	47.3(161)		30.55	46.1(378)		29.06	43.2(168)		29.18	44.7(455)		
Ile	ggg	28.76	28.6(97)		30.26	29.3(227)		29.18	29.5(123)		29.93	30.8(307)		
	ggt	18.94	12(40)		18.09	9(70)		18.78	15.3(62)		19.40	12.1(121)		
Leu	cac	60.86	77.1(144)	$P < 0.025$	62.80	76.7(192)	$P \leq 0.05$	60.73	70.9(121)	NS	60.06	73.3(238)	$P < 0.05$	
	cat	39.13	22.9(41)		37.19	23.3(57)		39.26	29.1(50)		39.93	26.7(87)		
Lys	ata	32.03	6.5(18)	$P < 0.001$	30.24	7.7(26)	$P < 0.001$	32.41	6(19)	$P < 0.001$	30.65	6.1(34)	$P < 0.001$	
	atc	41.36	72.9(207)		43.80	73.7(269)		41.04	78.4(208)		41.65	71.6(408)		
Met	att	26.60	20.5(59)		25.94	18.7(71)		26.53	15.6(44)		27.68	22.4(126)		
	aaa	44.23	19.2(75)	$P < 0.001$	41.07	22.2(94)	$P < 0.01$	44.02	25.2(86)	$P < 0.01$	41.77	21(143)	$P < 0.01$	
Val	aag	55.76	80.8(326)		58.92	77.8(314)		55.97	74.8(267)		58.22	79(538)		
	cta	17.13	3.7(31)	$P < 0.001$	16.17	3.1(260)	$P < 0.001$	17.16	3.9(32)	$P < 0.001$	16.00	2.8(41)	$P < 0.001$	
Phe	ctc	22.12	23.4(192)		23.42	24.3(200)		21.74	22.7(117)		21.74	21.7(301)		
	ctg	21.60	55.7(442)		23.20	59.5(487)		21.82	58.2(323)		22.31	59.6(808)		
Tyr	ctt	14.22	6.3(51)		13.87	5(41)		14.05	6.2(35)		14.45	6.1(82)		
	fta	11.01	2.2(18)		9.57	1.7(14)		11.09	1.3(7)		10.64	1.9(25)		
Trp	tgg	13.89	8.8(71)		13.74	6.5(55)		14.11	7.7(46)		14.83	7.9(108)		

Asn	aac	60.86	76.5(200)	$P < 0.025$	62.80	70.5(253)	NS	60.73	71.8(161)	NS	60.06	75.6(360)	$P < 0.025$
	aat	39.13	23.5(59)		37.19	29.5(96)		39.26	28.2(68)		39.93	24.4(116)	
Pro	cca	22.81	17.7(55)	NS	21.09	16.1(88)	NS	22.95	14.6(49)	$P < 0.025$	21.47	18.6(122)	$P < 0.025$
	ccc	29.46	45.8(135)		30.55	45.5(252)		29.06	44.6(130)		29.18	46.8(309)	
Gln	ccg	28.76	18.1(53)		30.26	20(109)		29.18	16.4(50)		29.93	13.7(91)	
	cct	18.94	18.4(56)		18.09	18.4(109)		18.78	24.4(83)		19.40	21(137)	
Arg	caa	44.23	12.4(39)	$P < 0.001$	41.07	11.1(43)	$P < 0.001$	44.02	12.6(43)	$P < 0.001$	41.77	11.7(62)	$P < 0.001$
	cag	55.76	87.6(279)		58.92	88.9(372)		55.97	87.4(264)		58.22	88.3(471)	
Ser	aga	12.62	5.6(20)	NS	10.75	11.8(49)	NS	12.84	9.1(35)	$P < 0.05$	11.46	7.4(50)	NS
	agg	15.91	21.1(77)		15.42	20(103)		16.32	15.6(62)		15.98	16.1(110)	
Thr	cga	16.30	8.8(31)		15.57	6.8(34)		16.26	7.7(30)		15.58	8.1(54)	
	cgc	21.05	27.8(98)		22.55	32.1(170)		20.58	30(112)		21.17	30.4(203)	
Val	cgg	20.55	30.2(105)		22.33	21.8(116)		20.67	33.1(107)		21.71	30.2(205)	
	cgt	13.53	6.4(23)		13.35	7.5(39)		13.30	4.6(17)		14.07	7.8(53)	
Tyr	agc	22.07	31.9(143)	NS	22.59	34.3(248)	$P < 0.05$	22.46	32.2(126)	NS	21.35	31.7(242)	NS
	agt	14.19	10.3(45)		13.37	7.2(50)		14.52	12.2(46)		14.19	8.5(63)	
Ile	tea	14.54	8.1(36)		13.50	9.2(63)		14.46	8.6(35)		13.84	9.1(68)	
	tcc	18.77	28.9(130)		19.56	30.9(225)		18.31	29.8(116)		18.80	27.8(210)	
Leu	teg	18.33	9.8(44)		19.37	8.8(62)		18.39	7.4(25)		19.29	8.1(64)	
	tct	12.07	10.9(49)		11.58	9.6(68)		11.83	9.8(39)		12.50	14.9(113)	
Met	aca	22.81	17.1(60)	$P < 0.025$	21.09	19.1(94)	$P < 0.025$	22.95	17.8(64)	$P < 0.05$	21.47	17.7(113)	$P < 0.05$
	acc	29.46	51.3(181)		30.55	51.9(268)		29.06	49.3(162)		29.18	49(311)	
Phe	acg	28.76	16.2(56)		30.26	16.6(91)		29.18	19.9(64)		29.93	17.5(111)	
	act	18.94	15.4(54)		18.09	12.4(65)		18.78	12.9(44)		19.40	15.8(99)	
His	gta	22.81	4.4(17)	$P < 0.001$	21.09	3.5(26)	$P < 0.001$	22.95	3.2(11)	$P < 0.001$	21.47	5.3(46)	$P < 0.001$
	gtc	29.46	25.7(101)		30.55	31.1(233)		29.06	28(99)		29.18	25(221)	
Lys	gtg	28.77	60.6(240)		30.26	59.2(47)		29.18	59.8(219)		29.93	62.1(544)	
	gtt	18.94	9.3(37)		18.09	6.2(48)		18.78	9(37)		19.39	7.6(67)	
Trp	tac	60.86	78.1(139)	$P < 0.01$	62.80	73.8(193)	NS	60.73	69.1(157)	NS	60.06	70.2(300)	NS
	tat	39.13	21.9(40)		37.19	26.2(74)		39.26	30.9(73)		39.93	29.8(126)	

The values in parenthesis indicate the number of codon in the dataset.

**Table 3.** The significance of deviation between observed and expected frequencies of 59 codons at the level of individual synonymous families for the genes having low GC<sub>3</sub> values among four different protein secondary structural classes

Amino-acids	Codons	All-alpha(low)				All-beta(low)				Alpha+beta(low)			
		Expected	Observed	Significance level	Expected	Observed	Significance level	Expected	Observed	Significance level	Expected	Observed	Significance level
Ala	gca	32.76	33.2(120)	P < 0.001	32.72	34.8(48)	P < 0.001	30.86	33.79(52)	P < 0.01	30.57	36.3(168)	P < 0.01
	gcc	18.98	21.2(79)		18.63	26.5(31)		19.88	23(113)		19.14	22.3(1103)	
	gcg	22.10	2.6(9)		22.69	2.6(6)		23.48	6.1(25)		23.12	5.1(22)	
Cys	gct	26.14	43(149)		25.95	36.1(57)		25.76	37.2(186)		27.15	36.3(172)	
	tgc	42.05	42.5(40)	NS	41.79	38.8(32)	NS	43.57	31.2(53)	NS	41.35	42.9(54)	NS
	tgt	57.94	57.5(55)		58.20	61.2(42)		56.42	68.8(120)		58.64	57.1(71)	
Asp	gac	42.05	37.1(116)	NS	41.79	37(61)	NS	43.56	34.9(147)	NS	41.35	33.2(140)	NS
	gat	57.94	62.9(205)		58.20	63(105)		56.43	65.1(267)		58.64	66.8(257)	
	gaa	59.72	68.8(302)	NS	59.05	71.3(161)	NS	56.78	67.4(323)	NS	56.93	63.4(290)	NS
Glu	gag	40.27	31.2(148)		40.94	28.7(72)		43.21	32.6(173)		43.06	36.6(181)	
	ttc	42.05	32.1(79)	NS	41.79	34.8(34)	NS	43.57	31.2(101)	NS	41.35	33.6(100)	NS
	ttt	57.94	67.9(143)		58.20	65.2(86)		56.42	68.8(217)		58.64	66.4(210)	
Gly	gga	32.76	40.7(85)	NS	32.72	46.5(90)	NS	30.86	42.5(218)	NS	30.57	39.9(193)	NS
	ggc	18.98	23.1(50)		18.63	18.3(38)		19.88	18.6(105)		19.14	20.9(104)	
	ggg	22.10	12.6(24)		22.69	17.8(42)		23.48	15.8(77)		23.12	15.7(78)	
His	ggt	26.14	23.6(50)		25.95	17.4(38)		25.76	23.2(130)		27.15	23.4(112)	
	cac	42.05	25.5(37)	P < 0.05	41.79	32.6(27)	NS	43.56	31.4(64)	NS	41.35	36.8(60)	NS
	cat	57.94	74.5(89)		58.20	67.4(53)		56.43	68.6(125)		58.64	63.2(96)	
Ile	ata	42.06	21.9(68)	P < 0.01	42.32	27.6(56)	NS	40.33	22.2(87)	P < 0.025	39.77	24.1(100)	P < 0.05
	atc	24.36	30.5(82)		24.10	24.5(46)		25.99	32.7(148)		24.90	23.8(105)	
	att	33.56	47.7(144)		33.57	48(85)		33.66	45.2(195)		35.31	52.1(225)	
Lys	aaa	59.72	62.7(247)	NS	59.05	62.2(134)	NS	56.78	59.3(298)	NS	56.93	54.8(269)	NS
	aag	40.27	37.3(150)		40.94	37.8(87)		43.21	40.7(217)		43.06	45.2(227)	
	cta	18.66	9.7(53)	NS	18.46	11.6(20)	NS	18.11	11.4(61)	NS	17.35	9.2(56)	NS
Leu	cic	10.80	11.5(66)		10.51	11.6(23)		11.67	14(82)		10.86	10.4(74)	
	cig	12.58	20.1(111)		12.80	18.7(55)		13.78	23.5(131)		13.12	20.9(138)	
	cit	14.89	22.7(141)		14.64	21.2(50)		15.11	19(106)		15.41	26.1(151)	
Tyr	tta	25.71	16.1(92)		25.72	18.3(35)		23.45	15.8(81)		24.61	14.5(80)	
	tty	17.34	19.9(123)		17.83	18.7(56)		17.85	16.2(93)		18.61	18.9(116)	



Asn	aac	42.05	41.6(92)	NS	41.79	37.1(54)	NS	43.57	37.2(134)	NS	41.35	36.8(127)	NS
Pro	aat	57.94	58.4(128)		58.20	62.9(96)		56.42	62.8(189)		58.64	63.2(200)	
	cca	32.76	37.1(60)	$P < 0.0001$	32.72	35.3(55)	$P < 0.0001$	30.86	42.6(150)	$P < 0.0001$	30.57	42.9(146)	$P < 0.01$
	ccc	18.98	20.3(37)		18.63	16.7(24)		19.88	16.2(61)		19.14	13.9(45)	
	ccg	22.10	3.5(6)		22.69	2.7(3)		23.48	4.4(17)		23.12	6.4(22)	
	cct	26.14	39.2(66)		25.95	45.3(67)		25.76	36.8(127)		27.15	36.8(120)	
Gln	caa	59.72	45(98)	NS	59.05	45.2(46)	$P < 0.05$	56.78	45.8(134)	NS	56.93	48.3(116)	NS
	cag	40.27	55(133)		40.94	54.8(74)		43.21	54.2(158)		43.06	51.7(126)	
Arg	aga	29.05	38.5(89))	NS	29.12	42.8(49)	NS	25.97	37.9(133)	NS	26.28	40.9(110)	NS
	agg	19.59	20(49)		20.19	20.4(34)		19.77	15.3(62)		19.88	15.7(440)	
	cga	16.82	16.4(33)		16.58	15.8(23)		16.74	15.3(57)		16.46	16.9(46)	
	cgc	9.74	8.2(19)		9.44	3.9(8)		10.78	9(33)		10.30	7.4(23)	
	cgg	11.34	11.8(29)		11.50	5.9(10)		12.74	10.3(38)		12.44	6.2(17)	
Ser	cgt	13.42	5.1(16)		13.15	11.2(16)		13.97	12.3(41)		14.61	12.8(35)	
	agc	16.69	16.4(52)	NS	16.98	13.1(32)	$P < 0.025$	17.09	16.9(78)	NS	15.98	9.6(50)	$P < 0.01$
	agt	23.00	19.3(64)		23.65	22.4(63)		22.14	22.4(98)		22.66	23.6(105)	
	tca	19.75	26.8(78)		19.42	22.4(55)		18.75	21.7(109)		18.76	17.7(85)	
	tcc	11.44	12.6(43)		11.06	13.1(38)		12.08	13.9(71)		11.74	13.2(65)	
Thr	tcg	13.32	3(9)		13.47	1.6(4)		14.27	2.8(13)		14.18	2.6(12)	
	tct	15.76	21.9(74)		15.40	27.3(70)		15.65	22.4(114)		16.65	33.2(146)	
	aca	32.77	42.1(115)	$P < 0.01$	32.72	40.3(74)	$P < 0.01$	30.86	35.9(136)	$P < 0.025$	30.57	40.3(155)	$P < 0.01$
	acc	18.97	21.3(55)		18.63	17.5(41)		19.88	24.6(105)		19.14	20.5(82)	
	acg	22.10	5.1(16)		22.69	5.3(12)		23.48	6.7(27)		23.12	4.9(21)	
Val	act	26.14	31.5(94)		25.95	36.9(70)		25.76	32.8(132)		27.15	34.4(135)	
	gta	32.76	17.5(53)	$P < 0.025$	32.72	19.7(42)	NS	30.86	23.7(102)	NS	30.57	20.3(99)	NS
	gtc	18.98	13.9(43)		18.63	14.2(34)		19.88	17.6(81)		19.14	15.2(78)	
	gtg	22.10	38.5(110)		22.69	33.5(69)		23.48	29.3(138)		23.12	31.2(155)	
	gtt	26.14	30.2(90)		25.95	32.6(74)		25.75	29.3(141)		27.15	33.3(163)	
Tyr	tac	42.05	37.2(60)	NS	41.79	43.7(48)	NS	43.57	38.5(113)	NS	41.35	29.3(67)	NS
	tat	57.94	62.8(103)		58.20	56.3(66)		56.42	61.5(165)		58.64	70.7(157)	

The values in parenthesis indicate the number of codon in the dataset.

**Table 4** The significance of deviation between observed and expected frequencies of 59 codons at the level of individual synonymous families for respective sheet portion among four different protein secondary structural classes

Amino acid	Codons	All-alpha(sheet)				All-beta(sheet)				Alpha+beta(sheet)				Alpha/beta(sheet)			
		Exp-frequency	Obs-frequency	Significance level	Exp-frequency	Obs-frequency	Significance level	Exp-frequency	Obs-frequency	Significance level	Exp-frequency	Obs-frequency	Significance level	Exp-frequency	Obs-frequency	Significance level	
Ala	gca	19.09	0(0)	$P < 0.001$	18.60	9(18)	$P < 0.001$	20.12	8.7(7)	$P < 0.001$	19.26	13.2(15)	$P < 0.001$	26.43	56.1(64)	$P < 0.001$	
	gcc	30.91	75(9)		28.72	56.7(118)		27.70	62.5(50)		26.43	8.8(10)		27.79	21.9(25)		
	gcg	25.30	0(0)		29.11	16.7(36)		26.59	13.8(11)		26.52	49.91	61.2(30)	NS	50.09	38.8(19)	NS
Cys	gct	24.70	25(3)	$P < 0.001$	23.57	17.6(36)	$P < 0.01$	51.99	60.5(23)	NS	49.91	38.8(19)	NS	50.09	61.2(30)	NS	
	tgc	55.59	0(0)	$P < 0.001$	54.92	72.7(80)	$P < 0.01$	48.01	39.5(15)	$P < 0.001$	50.09	38.8(19)	NS	49.91	61.2(30)	NS	
Asp	tgt	44.41	100(1)		45.08	27.3(30)		48.01	39.5(15)		50.09	38.8(19)		49.92	61.2(41)	NS	
	gac	55.59	75(6)	$P < 0.01$	54.92	73.6(64)	$P < 0.01$	51.99	86.7(39)	$P < 0.001$	49.92	61.2(41)	NS	50.08	38.8(26)	$P < 0.001$	
	gat	44.41	25(2)		45.08	26.4(23)		48.01	13.3(6)		50.08	38.8(26)		40.94	15.5(17)	$P < 0.001$	
Glu	gaa	43.00	9.1(1)	$P < 0.001$	38.99	13(22)	$P < 0.001$	43.07	32.4(22)	NS	40.94	15.5(17)	$P < 0.001$	59.06	84.5(93)	$P < 0.001$	
	gag	57.00	90.9(10)		61.01	87(1410)		56.93	67.6(46)		59.06	84.5(93)		49.92	71.2(109)	$P < 0.01$	
Phe	ttc	55.59	87.5(14)	$P < 0.001$	54.92	80(162)	$P < 0.001$	51.99	69.9(72)	$P < 0.01$	49.92	71.2(109)	$P < 0.01$	50.08	28.8(44)	$P < 0.01$	
	ttt	44.41	12.5(2)		45.08	20(42)		48.01	30.1(31)		50.08	28.8(44)		19.26	12(13)	$P < 0.01$	
Gly	gga	19.09	0(0)	$P < 0.001$	18.60	13(21)	$P < 0.001$	20.12	14.3(6)	$P < 0.001$	19.26	12(13)	$P < 0.01$	26.43	43.5(47)	$P < 0.01$	
	ggc	30.91	25(1)		28.72	57.8(95)		27.71	57.1(24)		26.43	43.5(47)		27.79	32.4(35)	$P < 0.01$	
His	ggg	25.30	75(3)		29.11	20.5(33)		26.59	23.8(10)		27.79	32.4(35)		26.52	12(13)	$P < 0.025$	
	ggd	24.70	0(0)		23.57	8.7(14)		25.58	4.8(2)		26.52	12(13)		49.91	66.7(34)	$P < 0.025$	
Ile	cac	55.59	85.7(6)	$P < 0.001$	54.92	68.8(56)	$P < 0.05$	51.99	67.3(33)	$P < 0.05$	49.91	66.7(34)	$P < 0.025$	50.09	33.3(17)	$P < 0.001$	
	cat	44.41	14.3(1)		45.08	31.2(23)		48.01	32.7(16)		50.09	33.3(17)		26.68	4(8)	$P < 0.001$	
Lys	ata	25.56	0(0)	$P < 0.001$	26.24	7.5(15)	$P < 0.001$	27.41	6.7(7)	$P < 0.001$	26.68	4(8)	$P < 0.001$	36.60	74.5(149)	$P < 0.025$	
	atc	41.38	76(19)		40.51	75.6(156)		37.74	83.8(88)		36.60	74.5(149)		40.94	25.3(21)	$P < 0.025$	
Leu	att	33.06	24(6)		33.25	16.9(34)		34.85	9.5(10)		36.72	21.5(43)		59.06	74.7(62)	$P < 0.001$	
	aaa	43.00	33.3(4)	NS	38.99	23.7(33)	$P < 0.025$	43.07	31.6(18)	NS	40.94	25.3(21)	$P < 0.025$	13.09	3.2(9)	$P < 0.001$	
Phe	aag	57.00	66.7(8)		61.01	76.3(103)		56.93	68.4(39)		59.06	74.7(62)		17.95	25.3(72)	$P < 0.001$	
	cta	14.09	5.9(2)	$P < 0.001$	13.37	3.3(14)	$P < 0.001$	14.06	6(10)	$P < 0.001$	13.09	3.2(9)	$P < 0.001$	18.88	52.3(149)	$P < 0.001$	
Tyr	ctc	22.82	14.7(5)		20.64	26(111)		19.36	21(35)		17.95	25.3(72)		18.01	6(17)	$P < 0.001$	
	ctg	18.68	58.8(20)		20.92	60.2(259)		18.58	53.9(90)		18.88	52.3(149)		13.13	2.5(7)	$P < 0.001$	
Trp	ctt	18.23	11.8(4)		16.94	3(14)		17.87	7.8(13)		18.01	6(17)		18.94	10.9(31)	$P < 0.001$	
	tta	11.26	0(0)		10.97	1.6(8)		12.98	1.2(2)		13.13	2.5(7)		18.94	10.9(31)	$P < 0.001$	
Tyr	ttg	14.92	8.8(3)		17.17	5.8(26)		17.15	10.2(17)		18.94	10.9(31)					

Asn	aac	55.59	100(5)	$P < 0.001$	54.92	73.1(48)	$P < 0.01$	51.99	72.4(21)	$P < 0.01$	49.91	77(47)	$P < 0.001$
	aat	44.41	0(0)		45.08	26.9(22)		48.01	27.6(8)		50.09	23(14)	
Pro	cca	19.09	20(2)	NS	18.60	12.3(8)	$P < 0.01$	20.12	9.5(2)	$P < 0.001$	19.26	14(7)	$P < 0.001$
	ccc	30.91	20(2)		28.72	53.8(35)		27.71	33.3(7)		26.43	52(26)	
Gln	ccg	25.30	30(3)		29.11	18.5(12)		26.59	4.8(1)		27.79	8(4)	
	cct	24.70	30(3)		23.57	15.4(10)		25.58	52.4(11)		26.52	26(13)	
	caa	43.00	0(0)	$P < 0.001$	38.99	8.9(15)	$P < 0.001$	43.07	13.2(7)	$P < 0.001$	40.94	0(0)	$P < 0.001$
	cag	57.00	100(12)		61.01	91.1(144)		56.93	86.8(46)		59.06	100(53)	
Arg	aga	9.25	0(0)	$P < 0.001$	9.20	6.2(13)	$P < 0.0025$	10.91	6.8(6)	$P < 0.025$	10.46	7.1(7)	NS
	agg	12.27	7.7(1)		14.40	21.6(35)		14.42	13.6(12)		15.09	16.3(16)	
Ser	cga	14.98	7.7(1)		14.21	7.4(14)		15.02	5.7(5)		14.34	11.2(11)	
	cgc	24.26	46.2(6)		21.94	37.7(63)		20.69	35.2(31)		19.68	30.6(30)	
	egg	19.86	30.8(4)		22.24	19.8(33)		19.86	29.5(26)		20.69	25.5(25)	
	cgt	19.38	7.7(1)		18.01	7.4(12)		19.10	9.1(8)		19.74	9.2(9)	
	agc	14.47	20(2)	$P < 0.001$	16.20	36.8(83)	$P < 0.01$	14.92	37.5(21)	$P < 0.001$	14.38	30.6(26)	$P < 0.001$
	agt	11.56	30(3)		13.30	5(11)		13.77	12.5(7)		14.43	5.9(5)	
Thr	tca	14.12	10(1)		13.12	8.4(18)		14.35	3.6(2)		13.72	11.8(10)	
	tcc	22.86	40(4)		20.25	29.7(71)		19.76	37.5(21)		18.82	29.4(25)	
	teg	18.72	0(0)		20.52	9.6(22)		18.96	7.1(4)		19.79	4.7(4)	
	tct	18.27	0(0)		16.62	10.5(23)		18.24	1.8(1)		18.88	17.6(15)	
	aca	19.09	14.3(2)	$P < 0.001$	18.60	18.6(45)	$P < 0.01$	20.12	18.3(15)	$P < 0.01$	19.26	18.9(28)	$P < 0.025$
	acc	30.91	71.4(10)		28.72	53.8(128)		27.71	50(41)		26.43	47.3(70)	
Val	acg	25.30	7.1(1)		29.11	17(40)		26.59	15.9(13)		27.79	18.9(28)	
	act	24.70	7.1(1)		23.57	10.5(26)		25.58	15.9(13)		26.52	14.9(22)	
	gta	19.09	17.2(5)	$P < 0.001$	18.60	2.9(13)	$P < 0.001$	20.12	3.1(5)	$P < 0.001$	19.26	5.7(19)	$P < 0.001$
	gtc	30.91	31(9)		28.72	30.9(147)		27.71	28.4(46)		26.43	24.8(83)	
Tyr	gtg	25.30	48.3(14)		29.11	60.1(287)		26.59	61.1(99)		27.79	62.1(208)	
	gtt	24.70	3.4(1)		23.57	6.1(29)		25.58	7.4(12)		26.52	7.5(25)	
	tac	55.59	71.4(5)	$P < 0.025$	54.92	77.59(102)	$P < 0.001$	51.99	61.3(49)	NS	49.91	72(67)	$P < 0.01$
	tat	44.41	28.6(2)		45.08	22.5(30)		48.01	38.8(31)		50.09	28(26)	

The values in parenthesis indicate the number of codon in the dataset.

**Table 5.** The significance of deviation between observed and expected frequencies of 59 codons at the level of individual synonymous families for respective helix portion among four different protein secondary structural classes

Amino acid	Codons	All-alpha(helix)				All-beta(helix)				Alpha+beta(helix)				Alpha/beta(helix)			
		Exp-frequency	Obs-frequency	Significance level	Exp-frequency	Obs-frequency	Significance level	Exp-frequency	Obs-frequency	Significance level	Exp-frequency	Obs-frequency	Significance level	Exp-frequency	Obs-frequency	Significance level	
Ala	gca	22.59	12.1(40)	$P < 0.001$	22.56	13.4(11)	$P < 0.05$	24.09	16.1(24)	$P < 0.001$	21.52	11.7(65)	$P < 0.001$	21.52	11.7(65)	$P < 0.001$	
	gcc	27.97	57.3(189)		29.96	48.8(39)		27.40	57.7(86)		28.55	56.9(316)		28.55	56.9(316)		
	ggg	28.93	14.8(49)		29.36	23.2(19)		29.38	10.7(16)		30.65	13.2(73)		30.65	13.2(73)		
Cys	gct	20.51	15.8(52)		18.12	14.6(13)		19.13	15.4(23)		19.28	18.2(101)		19.28	18.2(101)		
	tgc	57.69	66.7(42)	NS	62.31	79.3(25)	$P < 0.01$	58.89	80(28)	$P < 0.01$	59.69	72(54)	NS	59.69	72(54)	NS	
	tgt	42.31	33.3(21)		37.69	20.7(6)		41.11	20(7)		40.31	28(21)		40.31	28(21)		
Asp	gac	57.69	76(117)	$P < 0.01$	62.31	73.9(34)	NS	58.89	68.6(59)	NS	59.69	73.2(156)	$P < 0.05$	59.69	73.2(156)	$P < 0.05$	
	gat	42.31	24(37)		37.69	26.1(12)		41.11	31.4(27)		40.31	26.8(57)		40.31	26.8(57)		
	gaa	43.85	17.5(52)	$P < 0.001$	43.46	21.7(13)	$P < 0.01$	45.05	19.5(33)	$P < 0.001$	41.24	16.5(68)	$P < 0.001$	41.24	16.5(68)	$P < 0.001$	
Phe	gag	56.15	82.5(245)		56.54	78.3(44)		54.95	80.5(136)		58.76	83.5(345)		58.76	83.5(345)		
	tic	57.70	65.4(106)	NS	62.31	75919	NS	58.89	67.1(47)	NS	59.69	68.5(139)	NS	59.69	68.5(139)	NS	
	ttt	42.30	34.6(56)		37.69	25(6)		41.11	32.9(23)		40.31	31.5(64)		40.31	31.5(64)		
Gly	gga	22.59	12.6(15)	$P < 0.025$	22.56	12.5(2)	$P < 0.025$	24.09	14(6)	$P < 0.05$	21.52	18.3(40)	$P < 0.05$	21.52	18.3(40)	$P < 0.05$	
	ggc	27.97	50.4(60)		29.96	37.5(9)		27.40	46.5(20)		28.55	44.7(98)		28.55	44.7(98)		
	ggg	28.93	22.7(27)		29.36	41.7(9)		29.38	25.6(11)		30.65	29.7(65)		30.65	29.7(65)		
His	ggt	20.51	14.3(17)		18.12	8.3(2)		19.13	14(6)		19.28	7.3(16)		19.28	7.3(16)		
	cac	57.69	76.7(79)	$P < 0.01$	62.31	87.9(29)	$P < 0.001$	58.89	77.1(37)	$P < 0.01$	59.69	77.8(98)	$P < 0.01$	59.69	77.8(98)	$P < 0.01$	
	cat	42.31	23.3(24)		37.69	12.1(5)		41.11	22.9(11)		40.31	22.2(28)		40.31	22.2(28)		
Ile	ata	31.79	7.7(15)	$P < 0.001$	31.94	2.6(2)	$P < 0.001$	34.11	8.5(7)	$P < 0.001$	31.03	7.8(16)	$P < 0.001$	31.03	7.8(16)	$P < 0.001$	
	atc	39.35	74.7(145)		42.41	65.8(25)		38.80	75.6(62)		41.17	70.6(144)		41.17	70.6(144)		
	att	28.86	17.5(34)		25.65	31.6(110)		27.09	15.9(13)		27.80	21.6(44)		27.80	21.6(44)		
Lys	aaa	43.85	20.8(51)	$P < 0.001$	43.46	23.4(11)	$P < 0.01$	45.05	23.8(25)	$P < 0.01$	41.24	19.7(57)	$P < 0.001$	41.24	19.7(57)	$P < 0.001$	
	aag	56.15	79.2(194)		56.54	76.6(35)		54.95	76.2(80)		58.76	80.3(232)		58.76	80.3(232)		
	cta	16.40	4.1(22)	$P < 0.001$	17.17	4.1(3)	$P < 0.001$	17.54	6.3(13)	$P < 0.001$	15.91	3.2(20)	$P < 0.001$	15.91	3.2(20)	$P < 0.001$	
Leu	cic	20.30	25.3(136)		22.80	26(20)		19.96	19.5(40)		21.11	20.9(132)		21.11	20.9(132)		
	cig	21.00	55(296)		22.34	47.9(36)		21.39	62.9(129)		22.67	60.7(383)		22.67	60.7(383)		
	ctt	14.88	5.6(30)		13.79	8.2(5)		13.93	4.4(9)		14.26	6.7(42)		14.26	6.7(42)		
Tyr	tta	12.02	2(11)		10.39	4.1(3)		12.25	1(2)		10.74	2.2(14)		10.74	2.2(14)		
	tig	15.40	8(43)		13.51	9.6(7)		14.93	5.9(12)		15.31	6.3(40)		15.31	6.3(40)		

Asn	aac	57.69	74.1(83)	$P < 0.025$	62.31	82.6(19)	$P < 0.01$	58.89	76.3(45)	$P < 0.01$	59.69	70.9(95)	NS
	aat	42.31	25.9(29)		37.69	17.4(4)		41.11	23.7(14)		40.31	29.1(39)	
Pro	cca	22.59	20.5(15)	$P < 0.025$	22.56	3.6(1)	$P < 0.001$	24.09	15.9(7)	NS	21.52	23(34)	$P < 0.01$
	ccc	27.97	39.7(29)		29.96	57.1(17)		27.40	36.4(16)		28.55	45.9(68)	
Gln	ccg	28.93	12.3(9)		29.36	17.9(5)		29.38	18.2(8)		30.65	12.2(18)	
	cct	20.51	27.4(20)		18.12	21.4(5)		19.13	29.5(13)		19.28	18.9(28)	
Arg	caa	43.85	12.9(26)	$P < 0.001$	43.46	16.3(6)	$P < 0.001$	45.05	9(10)	$P < 0.001$	41.24	12(30)	$P < 0.001$
	cag	56.15	87.1(175)		56.54	83.7(36)		54.95	91(101)		58.76	88(220)	
Ser	aga	12.89	3(7)	$P < 0.01$	12.22	0(1)	$P < 0.01$	14.41	10.3(12)	$P < 0.025$	11.64	5.8(18)	NS
	agg	16.50	24.1(56)		15.90	26.2(13)		17.57	19(22)		16.58	16.7(52)	
Thr	cga	15.95	6.9(16)		16.22	9.5(4)		16.39	10.3(12)		15.44	7.1(22)	
	cgc	19.75	27.6(64)		21.54	33.3(14)		18.64	26.7(31)		20.50	28.9(90)	
Val	cgg	20.43	31.5(73)		21.11	21.4(9)		19.99	31(36)		22.00	33.8(105)	
	cgt	14.48	6.9(16)		13.03	9.5(4)		13.01	2.6(3)		13.84	7.7(24)	
Tyr	agc	20.53	26.6(54)	NS	23.04	43.1(28)	$P < 0.025$	22.72	26.6(25)	$P < 0.01$	21.75	31(71)	$P < 0.01$
	agt	15.05	13.8(28)		13.94	9.2(6)		15.86	17(16)		14.68	9.2(21)	
Ile	tca	14.55	6.4(13)		14.22	4.6(4)		14.79	13.8(13)		13.68	6.1(14)	
	tcc	18.01	30.5(62)		18.88	23.1(15)		16.83	29.8(28)		18.15	31(71)	
Leu	tcg	18.64	9.9(20)		18.50	10.8(7)		18.04	1.1(1)		19.49	6.1(14)	
	tct	13.21	12.8(26)		11.42	9.2(5)		11.75	11.7(11)		12.26	16.6(38)	
Phe	aca	22.59	20(35)	$P < 0.01$	22.56	8.6(5)	$P < 0.001$	24.09	15.4(10)	$P < 0.05$	21.52	18.8(34)	$P < 0.025$
	acc	27.97	50.3(88)		29.96	65.7(23)		27.40	47.7(31)		28.55	49.7(90)	
Met	acg	28.93	16(28)		29.36	14.3(5)		29.38	24.6(16)		30.65	18.8(34)	
	act	20.51	13.7(24)		18.12	11.4(4)		19.13	12.3(8)		19.28	12.7(23)	
Lys	gta	22.59	1.5(4)	$P < 0.001$	22.56	13.3(6)	$P < 0.001$	24.09	2(2)	$P < 0.001$	21.52	4.2(13)	$P < 0.001$
	gtc	27.97	27.3(71)		29.96	24.4(12)		27.40	29.6(29)		28.55	23.9(74)	
His	gfg	28.93	60(156)		29.36	55.6(25)		29.38	58.2(57)		30.65	62.1(192)	
	gft	20.51	11.2(29)		18.12	6.793		19.13	10.2(10)		19.28	9.7(30)	
Trp	tac	57.70	79.6(90)	$P < 0.001$	62.31	61.9(13)	NS	58.89	70.8(46)	NS	59.69	69.2(108)	NS
	tat	42.30	20.4(23)		37.69	38.1(8)		41.11	29.2(19)		40.31	30.8(48)	

The values in parenthesis indicate the number of codon in the dataset.

**Table 6.** The significance of deviation between observed and expected frequencies of 59 codons at the level of individual synonymous families for the respective coil portion among four different protein secondary structural classes

Amino acid	Codons	All-alpha(coil)				All-beta(coil)				Alpha+beta(coil)				Alpha/beta(coil)			
		Exp-frequency	Obs-frequency	Significance level	Exp-frequency	Obs-frequency	Significance level	Exp-frequency	Obs-frequency	Significance level	Exp-frequency	Obs-frequency	Significance level	Exp-frequency	Obs-frequency	Significance level	
Ala	gea	23.48	15.8(18)	NS	22.01	10.2(29)	$P < 0.01$	23.79	19.3(28)	$P < 0.01$	22.51	17.4(53)	$P < 0.001$	22.51	17.4(53)	$P < 0.001$	
	gcc	31.53	50(57)		32.31	57.4(148)		30.30	52.4(76)		31.11	53.3(162)		31.11	53.3(162)		
	geg	28.20	18.4(21)		31.50	17.7(46)		29.96	13.8(20)		29.61	9.9(30)		29.61	9.9(30)		
Cys	gct	16.79	15.8(18)		14.18	14.7(40)		15.95	14.5(21)		16.78	19.4(59)		16.78	19.4(59)		
	tgc	65.25	61.4(27)	NS	69.50	65.8(52)	NS	65.52	73.3(44)	NS	64.96	66.1(72)	NS	64.96	66.1(72)	NS	
	tgt	34.75	38.6(17)		30.50	34.2(26)		34.48	26.7(16)		35.04	33.9(37)		35.04	33.9(37)		
Asp	gac	65.26	74.1(126)	NS	69.50	74.7(237)	NS	65.52	68.4(121)	NS	64.96	71.1(290)	NS	64.96	71.1(290)	NS	
	gat	34.74	25.9(44)		30.50	25.3(78)		34.48	31.6(56)		35.04	28.9(118)		35.04	28.9(118)		
Glu	gaa	45.44	17.8(23)	$P < 0.001$	41.13	22(56)	$P < 0.01$	44.26	13.7(19)	$P < 0.001$	43.19	25(64)	$P < 0.01$	43.19	25(64)	$P < 0.01$	
	gag	54.56	82.2(106)		58.87	78(197)		55.74	86.3(120)		56.81	75(192)		56.81	75(192)		
Phe	tic	65.26	66.2(49)	NS	69.50	81.2(67)	NS	65.52	68.2(45)	NS	64.96	65.5(112)	NS	64.96	65.5(112)	NS	
	ttt	34.74	33.8(25)		30.50	18.8(15)		34.48	31.8(21)		35.04	34.5(59)		35.04	34.5(59)		
Gly	gga	23.48	14.1(26)	$P < 0.025$	22.01	15.9(94)	NS	23.79	13.7(39)	NS	22.51	10.6(67)	NS	22.51	10.6(67)	NS	
	gge	31.53	49.2(91)		32.31	46.1(267)		30.30	38(108)		31.11	45.6(287)		31.11	45.6(287)		
His	ggg	28.20	29.2(54)		31.50	29.5(170)		29.96	31.7(90)		29.61	29.7(187)		29.61	29.7(187)		
	ggt	16.79	7.6(14)		14.18	8.5(50)		15.95	16.5(47)		16.78	14.1(89)		16.78	14.1(89)		
Ile	cac	65.26	80(52)	$P < 0.025$	69.50	79.1(104)	NS	65.52	69.2(45)	NS	64.96	72.3(102)	NS	64.96	72.3(102)	NS	
	cat	34.74	20(13)		30.50	20.9(27)		34.48	30.8(20)		35.04	27.7(39)		35.04	27.7(39)		
Lys	ata	32.71	5.2(3)	$P < 0.001$	32.13	6.5(7)	$P < 0.001$	33.97	6.5(5)	$P < 0.001$	31.97	6.3(10)	$P < 0.001$	31.97	6.3(10)	$P < 0.001$	
	atc	43.91	62.1(36)		47.16	71.3(81)		43.27	67.5(52)		44.19	70.3(111)		44.19	70.3(111)		
Leu	att	23.38	32.8(19)		20.70	22.2(24)		22.77	26(20)		23.84	23.4(37)		23.84	23.4(37)		
	aaa	45.44	14.2(18)	$P < 0.001$	41.14	22.1(44)	$P < 0.01$	44.26	20.9(34)	$P < 0.001$	43.19	20.1(56)	$P < 0.001$	43.19	20.1(56)	$P < 0.001$	
Leu	aag	54.56	85.8(109)		58.86	77.9(160)		55.74	79.1(129)		56.81	79.9(222)		56.81	79.9(222)		
	cta	18.42	2.7(6)	$P < 0.001$	17.82	2.7(7)	$P < 0.001$	18.55	3(5)	$P < 0.001$	17.57	2.7(11)	$P < 0.001$	17.57	2.7(11)	$P < 0.001$	
Leu	ctc	24.73	21.4(47)		26.16	20.9(62)		23.62	21.8(36)		24.28	21.4(88)		24.28	21.4(88)		
	ctg	22.11	56.4(124)		25.51	61(180)		23.35	56.4(93)		23.11	61.6(253)		23.11	61.6(253)		
Leu	ctt	13.16	6.4(14)		11.48	7.2(21)		12.43	7.9(13)		13.10	5.1(21)		13.10	5.1(21)		
	tta	9.81	2.7(6)		7.82	1(3)		9.76	1.2(2)		9.48	1(4)		9.48	1(4)		
Leu	tgg	11.77	10.5(23)		11.20	7.2(21)		12.29	9.7(16)		12.47	8.3(34)		12.47	8.3(34)		

Asn	aac	65.26	77.2(95)	NS	69.49	73.4(169)	NS	65.52	66.9(85)	NS	64.96	77.9(212)	$P < 0.05$
	aat	34.74	22.8(28)		30.51	26.6(65)		34.48	33.1(42)		35.04	22.1(60)	
Pro	cca	23.49	15.3(27)	NS	22.01	16(67)	NS	23.79	15.9(34)	$P < 0.025$	22.51	18.1(78)	$P < 0.025$
	ccc	31.53	48.9(86)		32.31	43.3(179)		30.30	44.9(96)		31.11	48.4(208)	
Gln	ccg	28.20	21(37)		31.50	20.6(84)		29.96	15.4(33)		29.61	14(60)	
	cct	16.79	14.8(26)		14.18	20.1(84)		15.95	23.8(51)		16.78	19.5(84)	
Arg	caa	45.44	10.9(10)	$P < 0.001$	41.13	9.5(18)	$P < 0.001$	44.26	18.5(22)	$P < 0.001$	43.19	13.5(28)	$P < 0.001$
	cag	54.56	89.1(82)		58.87	90.5(176)		55.74	81.5(97)		56.81	86.5(180)	
Ser	aga	12.63	12.5(12)	NS	10.99	10.4(33)	NS	13.14	10.8(16)	NS	11.83	8.6(21)	NS
	agg	15.16	18.8(18)		15.73	17.7(51)		16.54	18.9(28)		15.56	15.6(38)	
Thr	cga	16.96	11.5(11)		16.13	5.4(15)		16.73	8.8(13)		16.34	8.2(20)	
	cgc	22.77	26(25)		23.68	32.7(87)		21.31	31.1(46)		22.59	32.1(78)	
Val	cgg	20.36	25(24)		23.08	25.4(71)		21.07	26.4(39)		21.50	27.6(67)	
	cgt	12.12	6.2(6)		10.39	8.5(22)		11.21	4.1(6)		12.19	7.8(19)	
Tyr	agc	24.58	35.3(72)	NS	28.70	30.9(117)	NS	26.57	31.5(64)	NS	24.64	33.5(136)	$P < 0.01$
	agf	13.09	5.9(12)		12.60	7.7(26)		13.98	10.8(22)		13.29	7.9(32)	
Leu	tea	14.64	8.3(170)		12.92	9.6(37)		14.14	8.9(18)		13.97	9.4(38)	
	tec	19.65	30.9(63)		18.97	34.4(129)		18.01	28.1(57)		19.31	26.6(108)	
Phe	teg	17.57	10.3(21)		18.49	8.3(31)		17.81	7.9(16)		18.37	9.9(40)	
	tct	10.46	9.3(19)		8.33	9.1(34)		9.48	12.8(26)		10.41	12.8(52)	
Met	aca	23.48	13(18)	$P < 0.05$	22.01	18.4(37)	NS	23.79	17.9(29)	NS	22.51	16.2(47)	$P < 0.025$
	acc	31.53	50(69)		32.31	48.3(101)		30.30	49.4(80)		31.11	50.5(147)	
Ile	acg	28.20	19.6(27)		31.50	18.8(38)		29.96	19.8(32)		29.61	15.5(45)	
	act	16.79	17.4(24)		14.18	14.5(30)		15.95	13(21)		16.78	17.9(52)	
Lys	gta	23.48	8.9(8)	$P < 0.001$	22.01	2.9(6)	$P < 0.001$	23.79	4.2(4)	$P < 0.001$	22.51	6.1(13)	$P < 0.001$
	gtc	31.53	21.1(19)		32.31	32.2(66)		30.30	24.2*(23)		31.11	28.2(60)	
His	gfg	28.20	64.4(58)		31.50	58.5(118)		29.96	55.8(53)		29.61	61(130)	
	gft	16.79	5.6(5)		14.18	6.3(14)		15.94	15.8(15)		16.78	4.7(10)	
Trp	tac	65.25	75.4(43)	NS	69.49	69.7(71)	NS	65.52	75.4(52)	NS	64.96	71.4(120)	NS
	tat	34.75	24.6(14)		30.51	30.3(30)		34.48	24.6(70)		35.04	28.6(48)	

The values in parenthesis indicate the number of codon in the dataset.

families comprising 20 codons display codon distribution markedly different in four different protein classes. There is an increased non-random behaviour of synonymous codons in sheet portion of protein in higher level of GC<sub>3</sub>, as compared to helix portion whereas non-random behaviour of synonymous codons is markedly less in coil portion of all four protein classes.

It is known that translational speed and co-translational folding are the main factors that effect the correlation between synonymous codon usage and protein structure (Gu *et al* 2004; D'Onofrio 2002). Our results also indicate occurrence of relationship between synonymous codon usage and protein secondary structural classes in human proteins as it was observed that synonymous codon families show non-randomness in codon usage in four different secondary structural classes. However, when the genes were partitioned according to their GC<sub>3</sub> levels there was an increase in non-randomness in synonymous codon usage in high GC<sub>3</sub> group of genes irrespective of the protein secondary structural classes. It is demonstrated that the genes with higher level of GC<sub>3</sub> have increased hydrophobicity for the encoded protein as compared to genes with lower level of GC<sub>3</sub> and are therefore accompanied with increase in protein stability (Bernardi 2004). Increase in non-randomness of synonymous codon usage in GC<sub>3</sub> rich genes as well as in sheet structure indicates that increased levels of GC<sub>3</sub> might have some selective advantages towards protein stability for the increased non-random synonymous codon usage, as it has already been reported that sheet structures of human genes have higher levels of GC<sub>3</sub> than helix and coil structures (D'Onofrio *et al* 2002).

### Acknowledgements

Authors are thankful to the Department of Biotechnology, New Delhi for financial help.

### References

- Adzhubei A A, Adzhubei I A, Krashennnikov I A and Neidle S 1996 Non-random usage of 'degenerate' codons is related to protein three-dimensional structure; *FEBS Lett.* **399** 78–82
- Alvarez F, Robello C and Vignali M 1994 Evolution of codon usage and base contents in kinetoplastid protozoans; *Mol. Biol. Evol.* **11** 790–802
- Arhondakis S, Anletta F, Torelli G and D'Onofrio G 2004 Base composition and expression level of human genes; *Gene* **325** 165–169
- Basak S, Banerjee T, Gupta S K and Ghosh T C 2004 Investigation on the causes of codon and amino acid usages variation between thermophilic *Aquifex aeolicus* and mesophilic *Bacillus subtilis*; *J. Biomol. Struct. Dyn.* **22** 205–214
- Basak S and Ghosh T C 2006 Temperature adaptation of synonymous codon usage in different functional categories of genes: a comparative study between homologous genes of *Methanococcus jannaschii* and *Methanococcus maripaludis*; *FEBS Lett.* **580** 3895–3899
- Berman H M, Westbrook J, Feng Z, Gilliland G, Bhat T N, Weissig H, Shindyalov I N and Bourne P E 2000 The Protein Data Bank; *Nucleic Acids Res.* **28** 235–242
- Bernardi G 2004 *Structural and evolutionary genomics. Natural selection in genome evolution* (Amsterdam: Elsevier)
- Brenner S E, Koehl P and Levitt M 2000 The ASTRAL compendium for protein structure and sequence analysis; *Nucleic Acids Res.* **28** 254–256
- Chiappello H, Ollivier E, Landes-Devauchelle C, Nitschke P and Risler J L 1999 Codon usage as a tool to predict the cellular location of eukaryotic ribosomal proteins and aminoacyl-tRNA synthetases; *Nucleic Acids Res.* **27** 2848–2851
- Conte L L, Brenner S E, Hubbard T J P, Chothia C and Murzin A G 2002 SCOP database in 2002: refinements accommodate structural genomics; *Nucleic Acids Res.* **30** 264–267
- D'Onofrio G 2002 Expression patterns and gene distribution in the human genome; *Gene* **300** 155–160
- D'Onofrio G, Ghosh T C and Bernardi G 2002 The base composition of the human genes is correlated with the secondary structures of the encoded proteins. *Gene* **300** 179–187
- Ellis J T and Morrison D A 1995 *Schistosoma mansoni*: patterns of codon usage and bias. *Parasitology* **110** 53–60
- Grantham R, Gautier C, Gouy M, Mercier R and Pavé A 1980 Codon catalog usage and the genome hypothesis; *Nucleic Acids Res.* **8** r49–r62
- Gupta S K, Majumdar S, Bhattacharya T K and Ghosh T C 2000 Studies on the relationships between the synonymous codon usage and protein secondary structural units; *Biochem. Biophys. Res. Commun.* **269** 692–696
- Gu W, Zhou T, Ma J, Sun X and Lu Z 2004 The relationship between synonymous codon usage and protein structure in *Escherichia coli* and *Homo sapiens*; *Biosystems* **73** 89–97
- Ikemura T 1985 Codon usage and tRNA content in unicellular and multicellular organisms; *Mol. Biol. Evol.* **2** 13–34
- Kabsch W and Sander C 1983 Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features; *Biopolymers* **12** 2577–2637
- Karlin S and Mrazek J 1996 What drives codon choices in human genes?; *J. Mol. Biol.* **262** 459–472
- Kahali B, Basak S and Ghosh T C 2007 Reinvestigating the codon and amino acid usage of *S. cerevisiae* genome: A new insight from protein secondary structure analysis; *Biochem. Biophys. Res. Commun.* **354** 693–699
- Lynn D J, Singer G A and Hickey D A 2002 Synonymous codon usage is subject to selection in thermophilic bacteria; *Nucleic Acids Res.* **30** 4272–4277
- Majumdar S, Gupta S K, Sundararajan V S and Ghosh T C 1999 Compositional correlation studies among the three different codon positions in 12 bacterial genomes; *Biochem. Biophys. Res. Commun.* **266** 66–71
- Oresic M and Shalloway D 1998 Specific correlations between relative synonymous codon usage and protein secondary structure; *J. Mol. Evol.* **281** 31–48



- Pouwels P H and Leunissen J A 1994 Divergence in codon usage of *Lactobacillus* species; *Nucleic Acids Res.* **22** 929–936
- Siemion I Z and Siemion P J 1994 The informational context of the third base in amino acid codons; *Biosystems* **33** 139–148
- Tao X and Dafu D 1998 The relationship between synonymous codon usage and protein Structure; *FEBS Lett.* **434** 93–96
- Taylor F and Coates D 1989 The code within the codons; *Biosystems* **22** 177–187
- Thanaraj T A and Argos P 1996a Protein secondary structural types are differentially coded on messenger RNA; *Protein Sci.* **5** 1973–1983
- Thanaraj T A and Argos P 1996b Ribosome-mediated translational pause and protein domain organization; *Protein Sci.* **5** 1594–1612
- Volkenstein M V 1966 The genetic coding of protein structure; *Biochim. Biophys. Acta* **119** 421–424
- Zhang S, Zubay G and Goldman E 1991 Low-usage codons in *Escherichia coli*, yeast, fruit fly and primates; *Gene* **105** 61–72

ePublication: 21 June 2007