ISSN 1330-9862
(FTB-2193)

original scientific paper

Highly Expressed Genes in Marine Sponge *Suberites domuncula* Prefer C- and G-Ending Codons

Drago Perina[1], *Matija Harcet*[1], *Andreja Mikoč*[1], *Kristian Vlahoviček*[2], *Werner E. G. Müller*[3] *and Helena Ćetković*[1]*

[1]Department of Molecular Biology, Ruđer Bošković Institute,
Bijenička 54, HR-10000 Zagreb, Croatia

[2]Faculty of Science, Division of Biology, Molecular Biology Department,
Horvatovac 102a, HR-10000 Zagreb, Croatia

[3]Institute for Physiological Chemistry, Department for Applied Molecular Biology,
Johannes Gutenberg University, Duesbergweg 6, D-55099 Mainz, Germany

Received: November 11, 2008
Accepted: May 27, 2009

Summary

Sponges are the simplest extant phylum of Metazoa; they are closest to the common ancestor of all multicellular animals. A total of 223 coding sequences from *Suberites domuncula* (Demospongiae) represent the dataset for the codon usage analysis. A total of 46038 codons had an average guanine and cytosine (G+C) content of 45.8 % and an average content of guanine and cytosine at the synonymous third position of codons (GC3S) of 43.4 %. In this sample of genes considerable variations in synonymous codon usage were found. The G+C content of the coding sequences varied from 34 to 56.1 % and GC3S from 19 to 58.7 %. Correspondence analysis revealed that highly expressed genes preferentially use a limited subset of codons (preferred codons). A total of 15 preferred codons were found and they all, with one exception, end with C or G. The preferential use of C- or G-ending codons in highly expressed genes was possibly developed in a common ancestor of sponges and other Metazoa and it has remained conserved throughout the sponge evolution.

Key words: porifera, *Suberites domuncula*, codon usage

Introduction

Genetic code is degenerative; more than one synonymous codon can encode the same amino acid. Since 1979 it has frequently been observed that synonymous codons are not used randomly and that some are used more often than the others. This phenomenon is termed codon usage bias (*1–5*). Well documented studies correlate codon usage bias and various biological characteristics such as gene expression level (*6,7*), mutational bias (*8,9*), gene length (*10*), gene function (*11*), translational selection (*12*), relative abundance of isoaccepting tRNAs (*13*), mRNA secondary structure (*14*), protein structure and folding (*15*), and the composition of guanine and cytosine (*16*). The first recognized correlation was with the level of gene expression. Coding sequences of highly expressed genes were constituted of 'optimal' codons recognized by the most abundant isoaccepting tRNAs (*17*). This phenomenon has been found in many organisms, including *Escherichia coli* (*18*), *Saccharomyces cerevisiae* (*19*), *Caenorhabditis elegans* (*20*) and *Drosophila melanogaster* (*21*). Codon usage bias occurs in species from all three domains of life. It varies through genes within a genome as well as through taxa. Codon usage bias observed in one organism can be different in another either in terms of preferentially used codons or in terms of correlation with biological func-

*Corresponding author; Phone: ++385 1 4571 239; Fax: ++385 1 4561 177; E-mail: cetkovic@irb.hr

This paper was presented at the Congress of Croatian Society of Biochemistry and Molecular Biology in Osijek, Croatia, September 17–20, 2008

D. Perina and M. Harcet contributed equally to this work tions, or both. In multicellular eukaryotes this applies even to different tissues which may have different isoaccepting tRNA pools adjusted specifically to maximize the production of particular protein (*22*). Codon usage bias may also vary in different developmental stages (*23*). In organisms with extremely high adenine and thymine (A+T) or guanine and cytosine (G+C) content (*24*), mutation bias seems to be the most important factor involved in codon usage bias. Similar was found in human genes. Synonymous codon usage depends on genomic G+C content of the region where gene is located and appears not to be related to the level of gene expression (*16*). In some plants, it was observed that translational selection at silent sites and gene function affect codon usage the most (*25*).

Sponges (Porifera) are the simplest Metazoa, and probably the earliest branching metazoan phylum (*26*). They do not have true tissues and organs, and lead a sessile lifestyle. In many aspects of their molecular biology, such as genome size, gene content and intron positions, sponges probably reflect the situation in a metazoan ancestor (Urmetazoa). Until now, there has been very little information about codon usage in sponges. The only published research was done on a relatively small number of *Geodia cydonium* genes (39 sequences) and without the expression level data (*27*). G+C content in sponge genomes is relatively low: 43.9 % in *G. cydonium* and 39.6 % in *S. domuncula* (*28*). In this paper synonymous codon usage is analysed on a set of highly and lowly expressed *S. domuncula* genes.

Materials and Methods

A sample of 223 coding sequences of *S. domuncula* was collected for analysis of synonymous codon usage. Construction of *S. domuncula* EST (expressed sequence tag) database had been described earlier (*29,30*). In order to estimate transcription/expression levels of sponge genes, approx. 13 000 randomly sequenced ESTs were assembled into 4646 unique cDNAs using CAP3 sequence assembly program (*31*). From the pool of assembled *S. domuncula* ESTs, sequences of 38 most abundant transcripts, *i.e.* 'contigs' consisting of no less than 25 sequences were chosen. Those contigs comprised 9312 codons in total and included sequences coding for sponge homologues of 18 ribosomal proteins, natterin, gelsolin, cyclophilin, ferritin, Ube1c, profilin, porin 31HM, α1-tubulin, serum response factor, actin, Y-box-binding protein 1, tubulin, thioredoxin, coactosin, annexin, HNRPDL protein, Col protein, PBEF-1, cofilin and 1-cysPrx. These proteins have known functions (mostly ribosomal and structural) and are widespread throughout the eukaryote domain. From the same pool, 51 sequences ('singlets') found only in one copy were collected with 8798 codons. For the reasons of brevity, these sequences are not listed and the list is available from the authors on request. In order to identify selected sponge sequence homologies, BlastX was used to search a database consisting of all proteins from six model organisms with complete genome sequences: cnidaria *Nematostella vectensis* – starlet sea anemone, nematode *Caenorhabditis elegans* – roundworm, arthropode *Drosophila melanogaster* – fruit fly, echinodermate *Strongylocentrotus purpuratus* – purple sea urchin, urochordate *Ciona intestinalis* – sea squirt and vertebrate *Homo sapiens* – human. *S. domuncula* EST database is available at Sponge Base (*32*). Also, 44 cDNAs coding for the Ras family proteins (*29*) and 58 sequences coding for ribosomal proteins were used (*30*). Additionally, 32 well defined sequences from NCBI were added to our pool: coding sequences for LAGL protein (accession number AJ250580), vacuolar proton pump protein (AJ297976), apoptosis MA3 (Y15421), ethylene responsive receptor ERR (Y19159), BHP1 protein (Y19158), L27 protein (AY857441), sorcin (AM040448), ISG12 protein (BN000244), AdaPTin-1 protein (AJ699167), lysozyme (AJ699166), macrophage expressed protein (AJ890501), LPS-binding protein (AJ890500), Toll-like receptor adapter protein (AJ890499), nonmuscle myosin II regulatory light chain (AJ784435), leucine zipper and ICAT homologous protein (AJ784432), glycogen synthase kinase 3 (AJ784431), TCF/LEF transcription factor (AJ784430), arginine kinase (AJ744770), cortactin (Y18860), allograft inflammatory factor-1 (Y18439), δ-aminolevulinic acid dehydratase (AJ575745), retinoid X receptor (AJ517420), noggin-1 (AJ535747), SNO protein (AJ277954), c-jun N-terminal kinase (AJ291511), MyoI protein (AJ252240), nucleoside diphosphate kinase Nm23-SD1 (AY764256), defender against cell death 1-like molecule (AJ632073), Fas apoptotic inhibitory-related molecule (AJ632072), apoptosis-linked protein 2 (AJ632071), (1,3)-β-D-glucan binding protein (AJ606470), and epidermal growth factor precursor (AJ606469). Sequences shorter than 100 amino acids were not included into analysis unless they were complete. The longest coding sequence had 714 codons. The total number of codons was 46038 (not including STOP codons).

CODONW program was used for synonymous codon usage analysis (*33*). This program counts the number of codons used in genes and calculates other codon usage indices. RSCU (relative synonymous codon usage) is a value that indicates codon usage bias or random usage; it is the frequency of a particular codon divided by the frequency of synonymous codons, presuming that all synonymous codons are used equally. GC3S is a fraction of codons in a gene which has either a guanine or cytosine at the third codon position. N_c is effective number of codons, a simple measure of overall codon bias, which varies from 20 (extreme bias) to 61 (lack of bias). Fgc is G+C content of a gene. Fop is the frequency of optimal codons used in a gene, the ratio of optimal codons to synonymous codons. This value is always between 0, which indicates no usage of optimal codon, and 1, which indicates that only optimal codons are used. COA (correspondence analysis) is a multivariate statistical approach in codon usage analysis. This method plots genes according to their synonymous codon usage in a multi-dimensional space and identifies major trends as axes which account for the largest fractions of codon usage variation among genes.

Results and Discussion

Overall codon usage of 223 *S. domuncula* genes with total of 46038 codons is presented in Table 1. The average G+C content of genes used in the analyses was 45.8 % and the average GC3S was 43.4 %. Higher frequencies of C- and/or G-ending codons are found only in six

Table 1. Overall codon usage in 223 *S. domuncula* genes

		N	RSCU			*N*	RSCU			*N*	RSCU			*N*	RSCU
Phe	UUU	999	1.12	Ser	UCU	790	1.53	Tyr	UAU	689	0.92	Cys	UGU	558	1.37
	UUC	784	0.88		UCC	396	0.77		UAC	817	1.08		UGC	256	0.63
Leu	UUA	321	0.53		UCA	654	1.27	TER	UAA	–	–	TER	UGA	–	–
	UUG	673	1.10		UCG	194	0.38		UAG	–	–	Trp	UGG	488	1.00
	CUU	723	1.19	Pro	CCU	699	1.47	His	CAU	499	0.99	Arg	CGU	643	1.32
	CUC	771	1.27		CCC	350	0.74		CAC	505	1.01		CGC	214	0.44
	CUA	483	0.79		CCA	736	1.55	Gln	CAA	975	1.03		CGA	362	0.74
	CUG	685	1.12		CCG	116	0.24		CAG	920	0.97		CGG	96	0.20
Ile	AUU	1039	1.29	Thr	ACU	904	1.33	Asn	AAU	829	0.92	Ser	AGU	680	1.32
	AUC	928	1.15		ACC	613	0.90		AAC	967	1.08		AGC	382	0.74
	AUA	454	0.56		ACA	980	1.44	Lys	AAA	1509	0.82	Arg	AGA	1031	2.12
Met	AUG	1142	1.00		ACG	232	0.34		AAG	2189	1.18		AGG	574	1.18
Val	GUU	922	1.12	Ala	GCU	1375	1.77	Asp	GAU	1322	1.00	Gly	GGU	1130	1.34
	GUC	792	0.96		GCC	796	1.02		GAC	1321	1.00		GGC	590	0.70
	GUA	589	0.72		GCA	823	1.06	Glu	GAA	1238	0.90		GGA	1197	1.42
	GUG	992	1.20		GCG	119	0.15		GAG	1517	1.10		GGG	466	0.55

N – number of codons, RSCU – relative synonymous codon usage

amino acids (Leu, Val, Tyr, Lys, Asn and Glu), while in two amino acids (His and Asp), C- and U-ending codons are used almost equally. In this dataset, NCG type of codons is used significantly less. Due to the randomness of EST sequencing process, the number of sequences in contigs reflects the abundance of different mRNAs in sponge 'tissue'. Although mRNA abundance is not always entirely correlated with translation level and protein abundance, in our dataset this correlation is most likely present. The most abundant sponge mRNAs (those found as contigs with high number of sequences) encode proteins that are universally known to be highly expressed, such as ribosomal proteins, collagen, actin and other structural proteins. On the other hand, most mRNAs found as singlets encode different regulatory proteins.

The G+C values of coding sequences varied from 34 to 56.1 %. Actin, found as contig consisting of 86 sequences, had the highest GC3S value, 58.7 %. Ferritin (56.7 %), found as contig consisting of 234 sequences and ribosomal proteins L27 (57.4 %) and S20 (56.6 %) had values close to this one. CD63 antigen (melanoma antigen), found as singlet, had the lowest GC3S value (only 19 %). The next lowest GC3S values were observed in SdRab-like 1 protein (29.9 %), SdRab21-like protein (32.5 %), glycosylasparaginase (32.2 %), and amyloid beta precursor member 1 protein (32.9 %), all found as singlets, but, surprisingly, also in collagen as contig of 47 sequences (26.1 %). Cysteine-rich heart protein showed the most extreme codon usage bias; its N_c value was 35. Values lower than 40 were observed in CD63 antigen (melanoma 1 antigen) and skd/vacuolar sorting protein found as singlets, but also in ribosomal proteins L37 and S21. N_c plot, where N_c values are plotted against GC3S values, are shown in Fig. 1. Continuous curve represents the expected relation between GC3S and N_c values if codons are used randomly. It is interesting to note that the majority of the points are positioned well bellow the expected curve. This is an indication of codon usage bias. Genes found as contigs and ribosomal proteins are located more to the right of the plot. These are highly expressed genes and their position indicates their preferential use of C- and G-ending codons. Correspondence analysis (COA) on RSCU values was performed to examine the major trends in codon usage. Fig. 2 shows the position of genes along the first and second axes produced by COA. COA on codon count accounted for 8.9 and 6.1 % on the first and second axes, respectively. The first axis is highly correlated with GC3S content. Although it does not explain a large proportion of the total variation in the data, it is interesting to note how genes are clustered on the first axis. Out of 20 coding sequences clustered on one side, nine were found as singlets, seven were collected from NCBI which mostly code for regulatory pro-

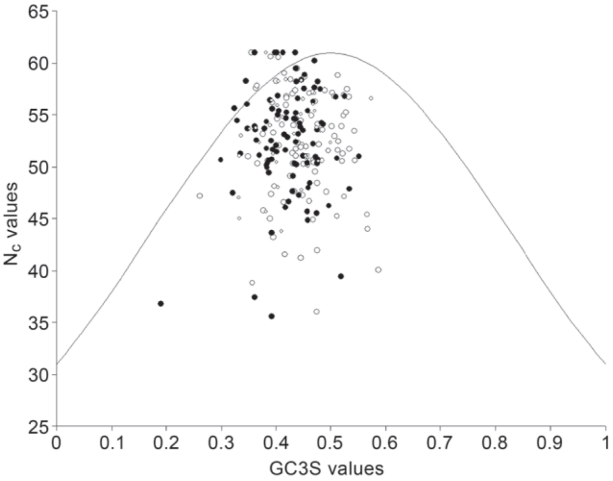

Fig. 1. N_c plot of 223 *S. domuncula* genes. N_c values are plotted against the GC3S values. Contig and EST coding for ribosomal proteins is marked with O, singlet and cDNA coding for the Ras family proteins is marked with ●. Remaining sequences are marked with ○

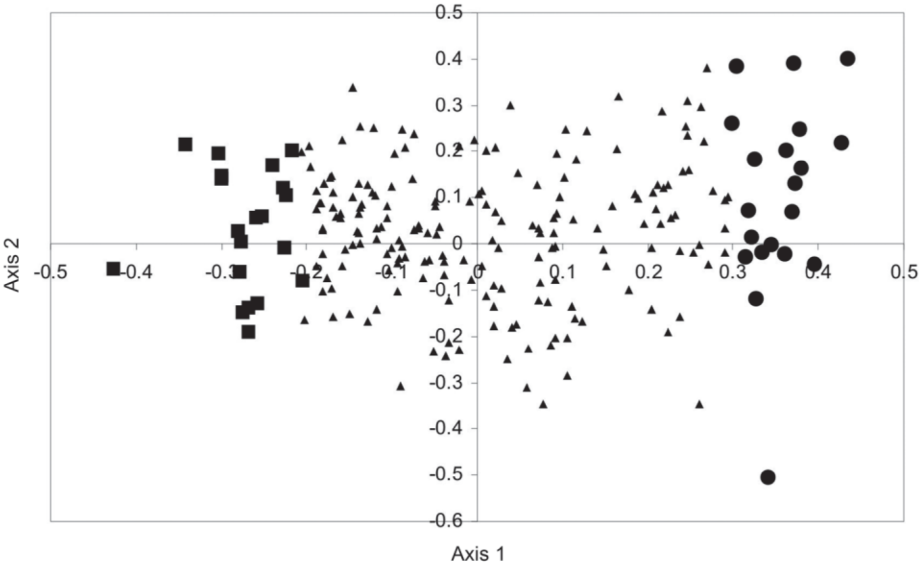

Fig. 2. Correspondence analysis of the relative synonymous codon usage (RSCU) variations among 223 *S. domuncula* genes. Positions of genes on the first two axes are shown. Genes from both extremities are marked with dots and squares

tein, three were Ras family proteins and only one was found as cluster (collagen). The average G+C value of these 20 coding sequences was 44.3 % and GC3S was 36.3 %. On the other side of the first major axis, 19 ribosomal proteins and actin (found as contig) were clustered. The average G+C value of these 20 coding sequences was 46.5 % and GC3S was 49.8 %. According to these results, expression levels can distinguish genes along the first major explanatory axis based on their codon usage. To explore differences in codon usage variation between these two clusters and to determine the preferred codons, twelve sequences from two extremities on the first axis were used. Chi square test was performed with probability $p<0.01$ as significance criterion and twelve preferred codons for twelve amino acids were found (Table 2). They all, except one, end with C or G. For three amino acids (Val, Tyr and Asp), preferred codons have probability $0.01<p<0.05$. These three codons end with C. Overall, 15 preferred codons for 15 amino acids were found and they all, except one, end with C or G. Finally, we wanted to explore the frequency of optimal codons (Fop) used in genes and whether the potentially highly expressed genes show highly optimal codon usage. Fop varies from 13.7 to 60.6 %. Potentially highly and lowly expressed genes are distinguished in Fig. 3 according to their Fop and G+C content (Fgc). Conspicuously, ribosomal proteins and contigs are clustered more to the right, which indicates higher optimal codon usage. Moreover, from 20 genes with the highest Fop, only one (SDRab8) was not ribosomal protein or sequence found as contig. GC3S preference in highly expressed genes is interesting because *S. domuncula* genome overall has low G+C content of 39.6 % (*28*). Our results are in accordance with the observed preference for C- and G-ending codons in highly expressed genes in many organisms (*18–21*). This codon bias is probably a result of selection of optimal codons recognized more efficiently and/or more accurately by more abundant isoaccepting tRNA. Genes using more of these optimal codons may be translated more efficiently with fewer mistakes than genes consisting of less frequent codons, so the selection may favour the usage of more frequent codons (*17*). In accordance with this selection model is the observation that G+C content of introns in *Drosophila* and *C. elegans* is not positively correlated with the level of gene expression. However, no difference in codon usage was found between ribosomal protein genes and potentially low expressed genes in humans and other vertebrates (*5*). Significant correlation between codon usage and gene expression in humans can only be observed if G+C content of the isochore where gene is located is taken into account. Sueoka and Kawanishi (*16*) proposed that this pattern is a result of directional mutation pressure rather than of directional selection pressure. Data of bias in synonymous codon usage in basal metazoan taxa, Porifera, is very limited. Relationship be-

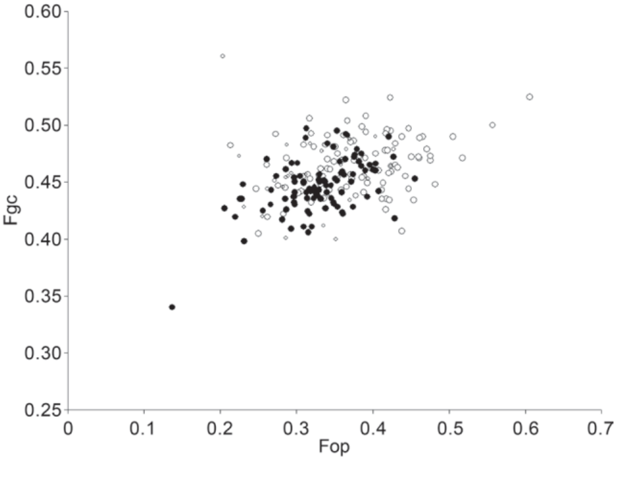

Fig. 3. Fop plot (frequency of optimal codons is ploted against G+C content – Fgc). Contig and EST coding for ribosomal proteins is marked with ○, singlet and cDNA coding for the Ras family proteins is marked with ●. Remaining sequences are marked with ○

Table 2. Codon usage in *S. domuncula* highly and lowly expressed genes with extreme positions on axis 1 of correspondence analysis

		High		Low				High		Low	
		RSCU	*N*	RSCU	*N*			RSCU	*N*	RSCU	*N*
Phe	UUU	0.79	25	1.28	97	Ser	UCU	1.77	36	1.39	73
	UUC*	1.21	38	0.72	54		UCC*	1.48	30	0.42	22
Leu	UUA	0.19	5	0.77	38		UCA	0.89	18	1.56	82
	UUG	1.01	27	1.08	53		UCG	0.44	9	0.38	20
	CUU	1.42	38	1.26	62	Pro	CCU	0.74	15	1.64	64
	CUC*	1.65	44	0.85	42		CCC*	1.63	33	0.33	13
	CUA	0.68	18	1.12	55		CCA	1.48	30	1.54	60
	CUG	1.05	28	0.92	45		CCG	0.15	3	0.49	19
Ile	AUU	1.00	35	1.42	92	Thr	ACU	1.45	38	1.32	93
	AUC*	1.80	63	0.65	42		ACC*	1.45	38	0.64	45
	AUA	0.20	7	0.93	60		ACA	0.72	19	1.74	122
Met	AUG	1.00	65	1.00	65		ACG	0.38	10	0.30	21
Val	GUU	1.13	42	1.18	78	Ala	GCU	1.74	64	1.80	90
	GUC**	1.02	38	0.80	53		GCC*	1.36	50	0.64	32
	GUA	0.81	30	0.95	63		GCA	0.63	23	1.44	72
	GUG	1.05	39	1.06	70		GCG	0.27	10	0.12	6
Tyr	UAU	0.67	22	0.98	64	Cys	UGU	1.00	15	1.38	73
	UAC**	1.33	44	1.02	66		UGC	1.00	15	0.62	33
TER	UAA	1.91	7	1.50	4	TER	UGA	0.82	3	1.13	3
	UAG	0.27	1	0.38	1	Trp	UGG	1.00	18	1.00	59
His	CAU	0.94	30	1.05	42	Arg	CGU*	1.67	58	0.99	32
	CAC	1.06	34	0.95	38		CGC	0.72	25	0.49	16
Gln	CAA	0.79	35	1.32	95		CGA	0.46	16	0.77	25
	CAG*	1.21	54	0.68	49		CGG	0.14	5	0.37	12
Asn	AAU	0.64	22	1.25	120	Ser	AGU	0.74	15	1.58	83
	AAC*	1.36	47	0.75	72		AGC	0.69	14	0.67	35
Lys	AAA	0.63	95	1.20	59	Arg	AGA	1.82	63	2.44	79
	AAG*	1.37	207	0.80	39		AGG	1.18	41	0.93	30
Asp	GAU	0.93	42	1.24	122	Gly	GGU	1.42	48	1.21	88
	GAC**	1.07	48	0.76	75		GGC	0.80	27	0.51	37
Glu	GAA	0.80	44	1.23	117		GGA	1.48	50	1.53	111
	GAG*	1.20	66	0.77	74		GGG	0.30	10	0.74	54

N – number of codons, RSCU – relative synonymous codon usage
Preferred codons are marked with *($p<0.01$) and **($0.01<p<0.05$)

tween GC3S and the level of gene expression was meaured in marine sponge *Geodia cydonium.* The major deficiency of the previous study on *G. cydonium* was a limited set of analyzed genes, lack of sponge's ribosomal protein genes (which are known to be expressed at high level) and estimation of abundance according to their relative abundance in vertebrates. *S. domuncula* had lower average GC3S (45.8 % in comparison with 55.7 % in *G. cydonium*) and G+C content of the analysed sequences (43.4 % in comparison with 51.2 % in *G. cydonium*). The differences are in accordance with the G+C composition of the genomes (43.9 % in *G. cydonium* and 39.6 % in *S. domuncula*). Many of the preferred codons in *S. domuncula* are known to be preferred by different eukaryotic organisms (*34*). Highly expressed genes in *G. cydonium* and *S. domuncula* preffer almost all the same codons, including CGU for Arg. The discrepancy in G+C composition between the two sponge species precludes us from making definite conclusions about the G+C composition of the common ancestor. In this work, a larger amount of random sequence data was analyzed, which could indicate that our data are more relevant and that the common ancestor did not have a G+C-rich genome. However, our results support the hypothesis that the preference for C- or G-ending codons appeared early in the metazoan evolution and remained conserved among sponges and, to a lesser extent, among other animals.

Conclusions

Investigation of the preferred codons in marine sponge *S. domuncula* revealed that highly expressed genes preferentially use C- and G-ending codons. Sponges as the simplest metazoans probably reflect the situation in the genome of the metazoan ancestor. The preference for C- and G-ending codons had previously been documented in other metazoans. Therefore, it is likely that the codon usage bias was present in the metazoan ancestor.

Acknowledgements

We wish to dedicate this paper to the late Dr Vera Gamulin.

This work is funded by Croatian MSES grants 098-0982913-2478 (H. Ćetković), 119-0982913-1211 (K. Vlahoviček), EMBO Young Investigator Program (Installation grant 1431/2006 to K. Vlahoviček) and Deutsche Forschungsgemeinschaft Mü/14-3 (W.E.G. Müller).

References

1. J.P. Holland, M.J. Holland, The primary structure of a glyceraldehyde-3-phosphate dehydrogenase gene from *Saccharomyces cerevisiae*, *J. Biol. Chem. 254* (1979) 9839–9845.
2. R. Grantham, C. Gautier, M. Gouy, R. Mercier, A. Pave, Codon catalog usage and the genome hypothesis, *Nucl. Acids Res. 8* (1980) r49–r62.
3. P.M. Sharp, T.M.F. Tuohy, K.R. Mosurski, Codon usage in yeast: Cluster analysis clearly differentiates highly and lowly expressed genes, *Nucl. Acids Res. 14* (1986) 5125–5143.
4. P.M. Sharp, M. Averof, A.T. Lloyd, G. Matassi, J.F. Peden, DNA-sequence evolution – The sounds of silence, *Phil. Trans. Roy. Soc. Lond. B. – Biol. Sci. 349* (1995) 241–247.
5. L. Duret, Evolution of synonymous codon usage in metazoans, *Curr. Opin. Genet. Dev. 12* (2002) 640–649.
6. P.M. Sharp, W.H. Li, An evolutionary perspective on synonymous codon usage in unicellular organisms, *J. Mol. Evol. 24* (1986) 28–38.
7. T. Nakamura, A. Suyama, A. Wada, Two types of linkage between codon usage and gene-expression levels, *FEBS Lett. 289* (1991) 123–125.
8. N. Sueoka, Directional mutation pressure and neutral molecular evolution, *Proc. Natl. Acad. Sci. USA, 85* (1988) 2653–2657.
9. N. Sueoka, Directional mutation pressure, selective constrains, and genetic equilibria, *J. Mol. Evol. 34* (1992) 95–114.
10. A. Eyre-Walker, Synonymous codon bias is related to gene length in *Escherichia coli*: Selection for translational accuracy?, *Mol. Biol. Evol. 13* (1996) 864–872.
11. Q. Liu, S. Dou, Z. Ji, Q. Xue, Synonymous codon usage and gene function are strongly related in *Oryza sativa*, *Biosystems, 80* (2005) 123–131.
12. J. Ma, A. Campbell, S. Karlin, Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures, *J. Bacteriol. 184* (2002) 5733–5745.
13. E.N. Moriyama, J.R. Powell, Codon usage bias and tRNA abundance in *Drosophila*, *J. Mol. Evol. 45* (1997) 514–523.
14. J.V. Chamary, L.D. Hurst, Evidence for selection on synonymous mutations affecting stability of mRNA secondary structures in mammals, *Genome Biol. 6* (2005) r75.
15. B. Kahali, S. Basak, T.C. Gosh, Reinvestigating the codon and amino acid usage of *S. cerevisiae* genome: A new insight from protein secondary structure analysis, *Biochem. Biophys. Res. Commun. 354* (2007) 693–699.
16. N. Sueoka, Y. Kawanishi, DNA G+C content of the third codon position and codon usage biases of human genes, *Gene, 261* (2000) 53–62.
17. T. Ikemura, Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optimal for the *E. coli* translational system, *J. Mol. Biol. 151* (1981) 389–409.
18. M. Gouy, C. Gautier, Codon usage in bacteria: Correlation with gene expressivity, *Nucl. Acids Res. 10* (1982) 7055–7074.
19. J.L. Bennetzen, B.D. Hall, Codon selection in yeast, *J. Biol. Chem. 257* (1982) 3026–3031.
20. M. Mitreva, M.C. Wendl, J. Martin, T. Wylie, Y. Yin, A. Larson, J. Parkinson, R.H. Waterston, J.P. McCarter, Codon usage patterns in Nematoda: Analysis based on over 25 million codons in thirty–two species, *Genome Biol. 7* (2006) r75.
21. H. Akashi, Synonymous codon usage in *Drosophila melanogaster*: Natural selection and translational accuracy, *Genetics, 136* (1994) 927–935.
22. D.C. Underwood, H. Knickerbocker, G. Gardner, D.P. Condliffe, K.U. Sprague, Silk gland-specific tRNA(Ala) genes are tightly clustered in the silkworm genome, *Mol. Cell. Biol. 8* (1988) 5504–5512.
23. S. Vicario, C.E. Mason, K.P. White, J.R. Powell, Developmental stage and level of codon usage bias in *Drosophila*, *Mol. Biol. Evol. 25* (2008) 2269–2277.
24. A. Saul, D. Battistutta, Codon usage in *Plasmodium falciparum*, *Mol. Biochem. Parasitol.* 27 (1988) 35–42.
25. H. Chiapello, F. Lisacek, M. Caboche, A. Hénaut, Codon usage and gene function are related in sequences of *Arabidopsis thaliana*, *Gene, 209* (1998) GC1-GC38.
26. W.E.G. Müller, Origin of Metazoa: Sponges as living fossils, *Naturwissenschaften, 85* (1998) 11–25.
27. V. Gamulin, J.F. Peden, I.M. Müller, W.E.G. Müller, Codon usage in the siliceous sponge *Geodia cydonium*: Highly expressed genes in the simplest multicellular animals prefer C- and G-ending codons, *J. Zool. Sist. Evol. Res. 39* (2001) 97–102.
28. M. Constantini, An analysis of sponge genomes, *Gene, 342* (2004) 321–325.
29. H. Cetkovic, A. Mikoc, W.E.G. Müller, V. Gamulin, Ras-like small GTPases form a large family of proteins in the marine sponge *Suberites domuncula*, *J. Mol. Evol. 64* (2007) 332–341.
30. D. Perina, H. Cetkovic, M. Harcet, M. Premzl, L. Lukic-Bilela, W.E.G. Müller, V. Gamulin, The complete set of ribosomal proteins from the marine sponge *Suberites domuncula, Gene, 366* (2006) 275–284.
31. X. Huang, A. Madan, CAP$_3$: A DNA Sequence Assembly Program, *Genome Res. 9* (1999) 868–877.
32. SpongeBase, The Center of Competence 'Molecular Biotechnology and Bioactive Compounds from Marine Sponges and Their Associated Organisms', University of Mainz, Mainz, Germany (*http://spongebase.uni-mainz.de*).
33. J.F. Peden, Analysis of codon usage, *PhD Thesis*, University of Nottingham, Nottingham, UK (1999) (*http://codonw.sourceforge.net/*).
34. B. Lafay, P.M. Sharp, Synonymous codon usage variation among *Giardia lamblia* genes and isolates, *Mol. Biol. Evol. 16* (1999) 1484–1495.