J Mol Evol (1992) 34:272–273

Journal of Molecular Evolution

Letter to the Editor

Codon Usage in Pathogenic *Entamoeba histolytica*

Egbert Tannich and Rolf D. Horstmann

Bernhard Nocht Institute for Tropical Medicine, 2000 Hamburg 36, Germany

Summary. An analysis of 4680 codons expressed by pathogenic *Entamoeba histolytica* showed the A+U content of coding sequences to be 67%. The preference for A+U resulted in an unusual codon usage with an A+U content of 84% in the third codon position. The data show a remarkable similarity to those obtained for *Plasmodium falciparum.*

Key words: *Entamoeba histolytica* — Codon usage — A+U content

The frequencies with which genetic condons are used may differ considerably between various species (Wada et al. 1991). Technically, codon usage information is of importance (1) to predict the most likely open reading frame from a novel sequence, (2) to design oligonucleotides from an amino acid sequence, and (3) to compare the codon usage of two species in order to estimate the degree of heterologous gene expression (Makoff et al. 1989).

The protozoon *Entamoeba histolytica* is a human intestinal parasite. Recently, evidence was presented for the existence of two genetically distinct forms, pathogenic and nonpathogenic ones (Tannich et al. 1989). The coding sequences of pathogenic *E. histolytica* published so far (Edman et al. 1987, 1990; Huber et al. 1988; Stanley et al. 1990; Torian et al. 1990; De Meester et al. 1991; Tannich et al. 1991a,b,c) and additional ones obtained in our laboratory comprise a total of 4680 expressed codons, which form a database comparable to those used in analyses of codon usage in other organisms (Wada et al. 1991).

The results obtained for pathogenic *E. histolytica* showed the A+U content to be 67% in the coding sequences, and accordingly, codon usage (Table 1) and the nucleotide frequencies in the three codon positions (Table 2) were found to be biased toward A and U.

The impact of the preference for A+U may be analyzed by comparing pathogenic *E. histolytica* with organisms that have a balanced (A+U):(G+C) ratio in the coding sequences such as *Escherichia coli,* in which (A+U):(G+C) is 48:52 (Wada et al. 1991). For instance, a comparison of the nucleotide frequencies in the three codon positions reveals that C being in the first position is twice as frequent in *E. coli* (24%) as it is in pathogenic *E. histolytica* (12%), and in the third codon position, the ratio (A+U):(G+C) is 84:16 in pathogenic *E. histolytica* compared to 44:56 in *E. coli* (Table 2). Accordingly, a number of codons are rarely used by pathogenic *E. histolytica,* and some have not yet been found at all (CCG, GCG, CGC, CGG). Athough no data are available on the structure and abundance of tRNAs in *E. histolytica,* it is reasonable to assume that their levels are shifted according to the codon usage (Bulmer 1987). Underrepresentation of certain tRNAs, for instance of those accepting the CGN codons for arginine (see Table 1), may largely restrain the capacity of *E. histolytica* to express heterologous genes (Makoff et al. 1989). It should be noted that the database studied here may not be representative for rare codons (Gouy and Gautier 1982) as the average expression of the genes included may be higher than the average of all genes expressed by the amoebae.

Of the nonpathogenic form of *E. histolytica* not more than 878 codons have been sequenced so far (Edman et al. 1990; Tannich et al. 1991a,c). From the limited data available, no difference in the codon usage of pathogenic forms has become evident.

Table 1. Codon frequency in pathogenic *Entamoeba histolytica*[a]

Amino acid	Codon	Frequency[b]	Amino acid	Codon	Frequency	Amino acid	Codon	Frequency	Amino acid	Codon	Frequency
Phe	TTT	188	Ser	TCT	150	Tyr	TAT	380	Cys	TGT	318
Phe	TTC	169	Ser	TCC	11	Tyr	TAC	73	Cys	TGC	41
Leu	TTA	226	Ser	TCA	318	End	TAA	17	End	TGA	2
Leu	TTG	75	Ser	TCG	4	End	TAG	2	Trp	TGG	107
Leu	CTT	248	Pro	CCT	32	His	CAT	135	Arg	CGT	21
Leu	CTC	30	Pro	CCC	4	His	CAC	32	Arg	CGC	0
Leu	CTA	11	Pro	CCA	344	Gln	CAA	286	Arg	CGA	17
Leu	CTG	4	Pro	CCG	0	Gln	CAG	13	Arg	CGG	0
Ile	ATT	466	Thr	ACT	291	Asn	AAT	530	Ser	AGT	175
Ile	ATC	75	Thr	ACC	49	Asn	AAC	107	Ser	AGC	43
Ile	ATA	43	Thr	ACA	342	Lys	AAA	615	Arg	AGA	263
Met	ATG	207	Thr	ACG	9	Lys	AAG	231	Arg	AGG	9
Val	GTT	425	Ala	GCT	259	Asp	GAT	530	Gly	GGT	165
Val	GTC	62	Ala	GCC	28	Asp	GAC	64	Gly	GGC	4
Val	GTA	126	Ala	GCA	318	Glu	GAA	709	Gly	GGA	483
Val	GTG	17	Ala	GCG	0	Glu	GAG	83	Gly	GGG	15

[a] Determined using the DNASIS computer package (Pharmacia LKB)
[b] Frequencies are per 10,000 codons

Table 2. Nucleotide frequencies in the three codon positions in *Entamoeba histolytica* (E.h.), compare to *Plasmodium falciparum* (P.f.) and *Escherichia coli* (E.c.)

	First position			Second position			Third position		
	E.h.	P.f.	E.c.	E.h.	P.f.	E.c.	E.h.	P.f.	E.c.
T	21[a]	17	15	24	22	29	43	39	26
C	12	13	24	21	19	23	8	10	28
A	34	36	25	38	49	30	41	44	18
G	33	34	36	17	10	18	8	7	28

[a] Percentage

The data presented here show a striking similarity to those obtained for *Plasmodium falciparum* (Table 2) (Saul and Battistutta 1988). The interpretation of this finding with regard to the evolutionary relationship between the two protozoa remains open as it cannot be deduced from the data whether the similarity is due to conservation or convergence.

References

Bulmer M (1987) Coevolution of codon usage and transfer RNA abundance. Nature 325:728–730

De Meester F, Bracha R, Huber M, Keren Z, Rozenblatt S, Mirelman D (1991) Cloning and characterization of an unusual elongation factor-1α cDNA from *Entamoeba histolytica*. Mol Biochem Parasitol 44:23–32

Edman U, Meza I, Agabian N (1987) Genomic and cDNA actin sequences from a virulent strain of *Entamoeba histolytica*. Proc Natl Acad Sci USA 84:3024–3028

Edman U, Meraz MA, Rausser S, Agabian N, Meza I (1990) Characterization of an immuno-dominant variable surface antigen from pathogenic and nonpathogenic *Entamoeba histolytica*. J Exp Med 172:879–888

Gouy M, Gautier C (1982) Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Res 10:7055–7074

Huber M, Garfinkel L, Gitler C, Mirelman D, Revel M, Rozenblatt S (1988) Nucleotide sequence analysis of an *Entamoeba histolytica* ferredoxin gene. Mol Biochem Parasitol 31: 27–34

Makoff AJ, Oxer MD, Romanos MA, Fairweather NF, Ballantine S (1989) Expression of tetanus toxin fragment C in *E. coli*: high level expression by removing rare codons. Nucleic Acids Res 17:10191–10202

Saul A, Battistutta D (1988) Codon usage in *Plasmodium falciparum*. Mol Biochem Parasitol 27:35–42

Stanley SL, Becker A, Kunz-Jenkins C, Foster L, Li E (1990) Cloning and expression of a membrane antigen of *Entamoeba histolytica* possessing multiple tandem repeats. Proc Natl Acad Sci USA 87:4976–4980

Tannich E, Horstmann RD, Knobloch J, Arnold HH (1989) Genomic DNA differences between pathogenic and nonpathogenic *Entamoeba histolytica*. Proc Natl Acad Sci USA 86: 5228–5122

Tannich E, Bruchhaus I, Walter RD, Horstmann RD (1991a) Pathogenic and nonpathogenic *Entamoeba histolytica*: identification and molecular cloning of an iron-containing superoxide dismutase. Mol Biochem Parasitol 49:61–72

Tannich E, Ebert F, Horstmann RD (1991b) Primary structure of the 170-kDa surface lectin of pathogenic *Entamoeba histolytica*. Proc Natl Acad Sci USA 88:1849–1853

Tannich E, Scholze H, Nickel R, Horstmann RD (1991c) Homologous cysteine proteinases of pathogenic and nonpathogenic *Entamoeba histolytica*: differences in structure and expression. J Biol Chem 266:4798–4803

Torian BE, Flores BM, Stroeher VL, Hagen FS, Stamm WE (1990) cDNA sequence analysis of a 29-kDA cysteine-rich surface antigen of pathogenic *Entamoeba histolytica*. Proc Natl Acad Sci USA 87:6358–6362

Wada K, Wada Y, Doi H, Ishibashi F, Gojobori T, Ikemura, T (1991) Codon usage tabulated from the GenBank genetic sequence data. Nucleic Acids Res 19:s1981–s1986

Received July 10, 1991