Acta Protozool. (2006) 45: 313 - 316

ACTA
PROTOZOOLOGICA

Short Communication

Codon Usage in *Amoeba proteus* Significantly Differs from *Entamoeba histolytica* and *Acanthamoeba castellanii*

Elżbieta KOCIK*, Magdalena SOBCZAK* and Maria Jolanta RĘDOWICZ

Department of Muscle Biochemistry, Nencki Institute of Experimental Biology, Warsaw, Poland

Summary. Codon usage analysis performed on 5718 expressing codons (including stop codons) of nine *Amoeba proteus* proteins deposited in GenBank revealed that there was no bias in *A. proteus* codon usage and nucleotide frequencies in the three codon positions. This contrasts with codon usage in other amoebozoans, such as *Acanthamoeba castellanii* and *Entamoeba histolytica,* which are biased towards GC and AT, respectively. Interestingly, codon usage in *A. proteus* resembled that of human and *Escherichia coli.* The presence of tRNAs for all possible codons indicates that all heterologous genes may be expressed in these giant cells. However, based on this analysis it is impossible to deduce whether the observed differences and/or similarities are due to conservation or convergence.

Key words: *Acanthamoeba castellanii*, *Amoeba proteus*, codon usage, *Entamoeba histolytica.*

INTRODUCTION

Amoeba proteus has been widely used as a model to study cell motility (Jeon 1995). However, molecular mechanisms governing locomotion and intracellular trafficking remain poorly understood. Parallel to that, the studies on the genome of these giant cells have been also neglected. As the result of that, to date only several actin-binding proteins have been detected including caldesmon (Gągola *et al*. 2003), spectrin (Choi and Jeon 1992), α-actinin and vinculin (Brix *et al*. 1990), Rho and Rac (Kłopocka and Rędowicz 2003) as well as two myosin heavy chains (Oh and Jeon 1998, Dominik *et al*. 2005). Moreover, only nine *A. proteus* proteins have been cloned and sequenced, among them are: actin (Fahrni *et al*. 2003), myosin II heavy chain (Oh and Jeon 1998), and recently a novel actin-binding protein interacting with anti-filamin antibody (ApABP-FI; GenBank accession no: DQ374440).

Therefore, there is very little information on codon usage in *A. proteus*. Moreover, data available at the www.kazusa.orjp/codon web site are based only on the cDNA sequence of four *A. proteus* proteins. This

Address for correspondence: Maria Jolanta Rędowicz, Nencki Institute of Experimental Biology, Department of Muscle Biochemistry, ul. Pasteura, 02-093 Warszawa, Poland; Fax +48-22-8225342; E-mail: j.redowicz@nencki.gov.pl

* these authors have equally contributed to this manuscript

makes very difficult to perform studies on cloning *A. proteus* proteins based on the amino acid sequence obtained, for example, from mass spectrometry analysis.

In order to get more knowledge on the codon usage we performed codon usage analysis based on 5718 codons of nine *A. proteus* proteins (including the stop codons) deposited in the GenBank database. The analysis revealed that, unlike for other amoebozoans, there was no bias in codon usage in *Amoeba proteus*.

MATERIALS AND METHODS

The analysis of codon usage in *Amoeba proteus* was performed using Codon Usage tool, available at the web site http://bioinformatics.org/sms2/. 5718 codons (including the stop ones) of nine *A. proteus* protein sequences deposited in the GenBank database were subjected to the analysis: polyubiquitin (AF034789), S-adenosylmethionine synthetase (U91602), S-adenosylmethionine synthetase 2 (AY324626), pepstatin-insensitive carboxyl proteinase 2 (AF142415), pepstatin-insensitive carboxyl

Table 1. Codon frequency in *Amoeba proteus**.

Amino acid	Codon	Frequency per 10,000 codons	Amino acid	Codon	Frequency per 10,000 codons
Gly	GGG	49	Thr	ACG	61
Gly	GGA	229	Thr	ACA	100
Gly	GGT	267	Thr	ACT	185
Gly	GUC	74	Thr	ACC	153
Glu	GAG	375	Trp	TGG	89
Glu	GAA	539	End	TGA	2
ASP	GAT	438	Cys	TGT	66
Asp	GAC	194	Cys	TGC	56
Val	GTG	124	End	TAG	2
Val	GTA	88	End	TAA	12
Val	GTT	217	Tyr	TAT	128
Val	GTC	124	Tyr	TAC	105
Ala	GCG	82	Leu	TTG	354
Ala	GCA	294	Leu	TTA	74
Ala	GCT	324	Phe	TTT	177
Ala	GCC	257	Phe	TTC	151
Arg	AGG	135	Ser	TCG	68
Arg	AGA	154	Ser	TCA	95
Ser	AGT	67	Ser	TCT	221
Ser	AGC	33	Ser	TCC	140
Lys	AAG	436	Arg	CGG	25
Lys	AAA	190	Arg	CGA	91
Asn	AAT	196	Arg	CGT	154
Asn	AAC	180	Arg	CGC	39
Met	ATG	138	Gln	CAG	184
Ile	ATA	61	Gln	CAA	270
Ile	ATT	210	His	CAT	100
Ile	ATC	250	His	CAC	102

* Determined with the Codon Usage tool available at http://bioinformatics.org/sms2/

Table 2. Nucleotide frequencies (as percentage) in the three codon positions in *Amoeba proteus* (Ap), *Entamoeba histolytica* (Eh), *Acanthamoeba castellanii* (Ac), *Homo sapiens* (Hs) and *Escherichia coli* (Ec).

	First position					Third position				
	Ap	Eh	Ac	Ec	Hs	**Ap**	Eh	Ac	Ec	Hs
G	**37**	33	37	36	34	**23**	8	39	28	29
A	**27**	34	21	25	29	**25**	41	<0.01	18	19
T	**17**	21	12	15	24	**31**	43	4	26	22
C	**19**	12	26	24	13	**21**	8	57	28	30

proteinase 1 (AF142414), myosin II heavy chain (AF136711), peroxiredoxin (AY869722), actin (AY294160), and a novel 110-kDa actin binding proteins interacting with anti-filamin antibody (ApABP-FI; DQ374440).

RESULTS AND DISCUSSION

We have recently cloned and sequenced cDNA of a novel 110-kDa actin-binding protein interacting with anti-filamin antibody and we deposited its sequence in GenBank (accession number: DQ374440). This protein, termed as ApABP-FI, is encoded by 2634 nucleotides that give 878 codons, including a stop codon T(U)AG. Before that only eight mRNA sequences encoding eight *A. proteus* proteins, including two cytoskeletal proteins: myosin II heavy chain (Oh and Jeon 1998) and actin (Fahrni *et al*. 2003) were deposited in the GenBank database. These nine proteins are encoded by the total number of 5718 codons (including stop ones) that, based on a report on codon usage in pathogenic *Entamoeba histolytica* (Tannich and Horstmann 1992), is sufficient to perform the codon usage analysis.

The obtained results (see Table 1) showed that the A+T(U) content in *A. proteus* was 53.8%. This is in contrast with the codon usage in other amoebae, which was found to be biased towards G+C in *Acanthamoeba castellanii* (Hammer *et al*. 1987, Nakamura *et al*. 1996; www.kazusa.orjp/codon) and towards A+T(U) in *D. discoideum* and *E. histolytica* (Tannich and Horstmann 1992, Nakamura *et al*. 1996; www.kazusa.or.jp/codon). The codon usage similar to *A. proteus* was found in organisms evolutionary distant such as *Homo sapiens* and *Escherichia coli*, that had balanced (A+U):(G+C) ratio in the coding sequences of 52:48 and 48:52, respectively (www.kazusa.or.jp/codon, Wada *et al*. 1991).

Accordingly to the balanced nucleotide content in *A. proteus*, codon usage (Table 1) and the nucleotide frequencies in the three codon positions (Table 2) were also not found to be biased: C and G were found to be in the first position in 19% and 37% of codons, and in third position in 21% and 23% of codons, respectively. A and T were present in first position in 27% and 17% of codons, and in third position in 25% and 31% of codons, respectively. These nucleotide frequencies (see Table 2) are not similar to *E. histolytica* or *A. castellanii* but are similar to those found in *E. coli* and *H. sapiens* (www.kazusa.or.jp/codon).

As it can be seen in Table 1, all the possible codons are used in *A. proteus*, and no preference in codon usage for any amino acid has been found. For example, histidine is coded by CAT and CAC with nearly identical frequencies (49.6 and 50.4%, respectively) and all six codons for arginine and leucine are used with the frequency varying from ~ 4% to 25% or 35%, respectively. On the other hand, a stop codon T(U)AA was found in seven out of nine *A. proteus* protein sequences characterized so far. This finding may indicate the preference for this particular stop codon in these giant cells. The presence of tRNAs for all the possible codons suggests that all heterologous genes may be expressed in *A. proteus*.

The interpretation of these observations with regard to the evolutionary divergence between amoebozoan organisms, and evolution *per se*, remains open as it is impossible to deduce whether these differences and/or similarities are due to conservation or convergence.

Acknowledgments. This work was supported by a grant to the Nencki Institute from the State Committee for Scientific Research.

REFERENCES

Brix K., Reinecke A. Stockem W. (1990) Dynamics of the cytoskeleton in *Amoeba proteus*. Influence of microinjected antibodies on the organization and function of the microfilament system. *Eur. J. Cell Biol.* **51:** 279-284

Choi E. Y., Jeon K. W. (1992) Role of spectrin in *Amoeba proteus*, as studied by microinjection of anti-spectrin monoclonal antibodies. *Exp. Cell Res.* **199:** 174-178

Dominik M., Kłopocka W., Pomorski P., Kocik E., Rędowicz M. J. (2005) Characterization of *Amoeba proteus* myosin VI-immunoanalog. *Cell Motil. Cytoskeleton* **61:** 172-188

Fahrni J. F., Bolivar I., Berney C., Nassonova E., Smirnov A., Pawlowski J. (2003) Phylogeny of lobose amoebae based on actin and small-subunit ribosomal RNA genes. *Mol. Biol. Evol.* **20:** 1881-1886

Gągola M., Kłopocka W., Grębecki A., Makuch, R. (2003) Immunodetection and intracellular localization of caldesmon-like proteins in *Amoeba proteus*. *Protoplasma* **222:** 75-83

Hammer J. A. 3-rd, Bowers B., Paterson B. M., Korn E. D. (1987) Complete nucleotide sequence and deduced polypeptide sequence of a nonmuscle myosin heavy chain gene from *Acanthamoeba*: evidence of a hinge in the rodlike tail. *J. Cell Biol.* **105:** 913-25

Jeon K. W. (1995) The large, free-living amoebae: wonderful cells for biological studies. *J. Euk. Microbiol.* **42:** 1-7

Kłopocka W., Rędowicz M. J. (2003) Effect of Rho family GTP-binding proteins on *Amoeba proteus*. *Protoplasma* **220:** 163-172

Nakamura Y., Wada K., Wada Y., Doi H., Kanaya S., Gojobori T., Ikemura T. (1996) Codon usage tabulated from the international DNA sequence databases. *Nucleic Acids Res.* **24:** 214-215

Oh S. W., Jeon K. W. (1998) Characterization of myosin heavy chain and its gene in *Amoeba proteus. J. Euk. Microbiol*. **45:** 600-605

Tannich E., Hortsmann R. D. (1992) Codon usage in pathogenic *Entamoeba histolytica. J. Mol. Evol.* **34:** 272-273

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 (Suppl):** 1981-1986

Received on 24th March, 2006; revised version on 25th April, 2006; accepted on 19th May, 2006