

Marx Gomes Van der Linden · Sávio Torres de Farias

## Correlation between codon usage and thermostability

Received: 7 April 2006 / Accepted: 23 March 2006 / Published online: 8 July 2006  
© Springer-Verlag 2006

**Abstract** Variations of arginine codon usage between organisms may have important implications to thermostability. The preferential usage of AGR codons for arginine in thermophiles and hyperthermophiles implies positive error minimization, contributing to avoid mutations that could harm protein thermostability. This bias is not a mere consequence of increased  $G + C$  content, as it has been previously suggested, and may represent a new mechanism of adaptation to protein thermostability.

**Keywords** Codon usage · Thermostability · Adaptation · Error minimization

### Introduction

Hyperthermophiles have increasingly fascinated researchers all over the world in the last decades, as evidenced by the increasing number of hyperthermophilic organisms that have been described in the last years. A more complete comprehension of the molecular mechanisms of thermostability is currently an important goal of structural biology.

The structural determinants of thermostability are now beginning to be understood. Most frequently reported trends in hyperthermophilic proteins include an increase in van der Waals interactions (Berezovsky et al. 1997), higher core hydrophobicity (Schumann et al. 1993), additional networks of hydrogen bonds (Jaenicke 2000), enhanced secondary structure propensity (Querol et al.

1996), ionic interactions (Ventriani et al. 1998), increased packing density (Hurley et al. 1992) and decreased length of surface loops (Thompson and Eisenberg 1999).

Farias and Bonato (2003) analyzed 28 proteomes from organisms belonging to the three domains of life and related the ability of a protein to be stable in high temperatures to an increased number of Glu ( $E$ ) and Lys ( $K$ ) residues, as well as a decrease in the number of Gln ( $Q$ ) and His ( $H$ ) residues. Based on these results, they suggested that the  $(E + K)/(Q + H)$  ratio could be used as an indicator of adaptations to high temperatures: hyperthermophiles would have proteomes with an average ratio higher than 4.5; thermophiles proteomes ratio would range between 3.2 and 4.6; mesophiles proteomes would have a ratio below 2.5. A codon usage bias was also observed: thermophiles and hyperthermophiles tend to employ AGR to encode for arginine, whereas mesophiles tend to employ CGN. Codon usage can be an evidence of positive and negative error minimization in translation. The choice for AGR in thermophiles and hyperthermophiles implies positive error minimization: mutations from G to A in the second base of these codons would change them to codons for lysine, which is also positively charged and important for thermostability. The choice for CGN in mesophiles implies a negative error minimization: these codons are a mutation away from histidine and glutamine, which are avoided in hyperthermophiles.

To test the importance of codon usage in protein thermostability, we analyzed all 173 organisms whose complete genomes are available in the TIGR database comprehensive microbial resource (CMR): 10 hyperthermophiles, 10 thermophiles and 153 mesophiles.

Communicated by J. N. Reeve

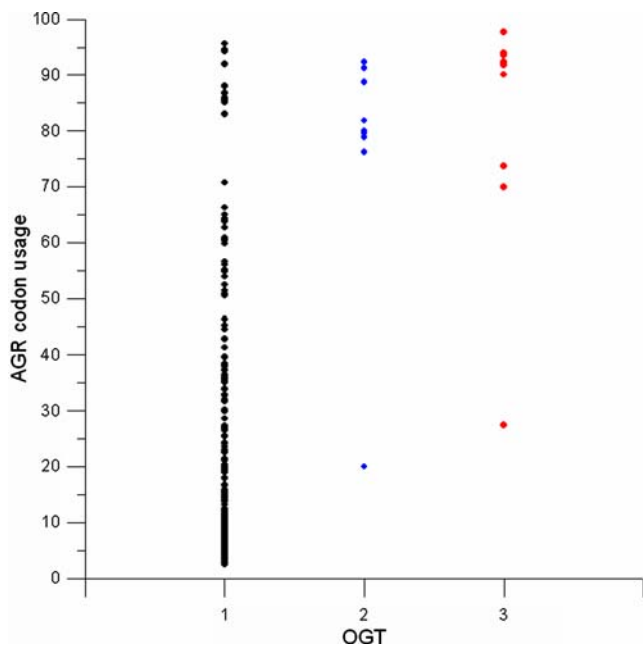
M. G. Van der Linden  
Laboratório Nacional de Computação Científica,  
Petrópolis, Rio de Janeiro, Brazil

S. T. de Farias (✉)  
Departamento de Biologia Geral,  
Universidade Federal de Minas Gerais,  
Belo Horizonte, MG, Brazil  
e-mail: stfarias@yahoo.com.br

### Results and discussion

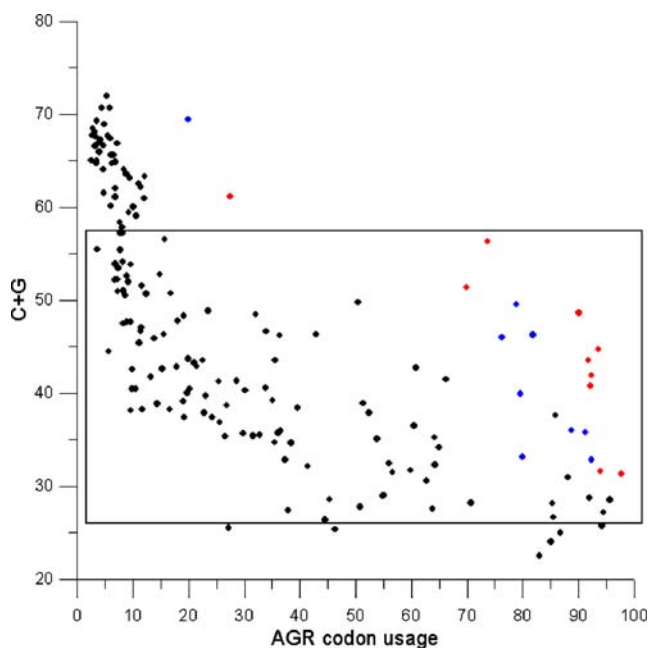
AGR codon usage and lifestyle

A comparative analysis of codon usage for arginine has been undertaken with three classes of organisms,

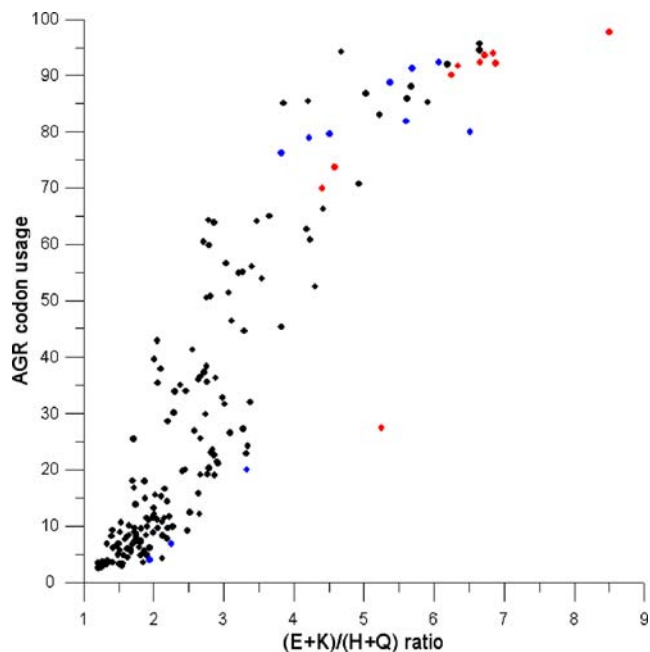


**Fig. 1** Correlation between ARG codon usage and lifestyle, in 1 and black mesophiles, 2 and blue thermophiles and 3 and red hyperthermophiles

according to their optimal growth temperature (OGT): mesophiles (OGT < 50°C), thermophiles (50°C < OGT < 80°C) and hyperthermophiles (80°C < OGT). Arginine codons were split into two groups: one containing CGN codons, the other containing AGR codons. Figure 1 shows the codon usage distribution between these two groups. A high bias can be observed for AGR



**Fig. 2** Correlation between AGR codon usage and C + G content in black mesophiles, blue thermophiles and red hyperthermophiles



**Fig. 3** Correlation between ARG codon usage and  $(E + K)/(H + Q)$  ratio in black mesophiles, blue thermophiles and red hyperthermophiles

codon usage in thermophiles and hyperthermophiles (T/HT), in contrast to lower AGR codon usage in mesophiles. The same observation has been previously made by Farias and Bonato (2003), albeit with a much smaller sample of organisms. These data indicate that some evolutionary force, such as founding effect or adaptive selection, probably has had a role in codon usage bias for arginine in T/HT. The founding effect hypothesis seems unlikely, since there are some mesophiles whose codon usage for arginine is similar to T/HT. To investigate the hypothesis of adaptive selection, we examined how C + G contents would affect codon usage for arginine among the organisms in our study. We have also studied the correlation between codon bias and the  $(E + K)/(H + Q)$  ratio, whose correlation with thermostability has been previously demonstrated (Farias and Bonato 2003; Farias et al. 2004).

#### AGR codon usage and C + G content

C + G content was one of the first aspects to be analyzed in organisms at the genome level. Some authors argue that the variations in codon usage among different organisms is a direct consequence of the proportion of C + G in their genomes (Porter 1995; Mita et al. 1991). Might the opposite be true, i.e., could C + G content be in fact a consequence of adaptive selection towards certain codons?

If the former hypothesis is true, we should expect an inverse correlation between C + G content and AGR codon bias, since these codons have a lower concen-

tration of  $C + G$  than the CGN codons, and the  $C + G$  concentration in organisms with a preference for AGR codons should not superpose those of organisms with a preference for CGN. Should the latter hypothesis be correct, we would expect to see organisms with similar  $C + G$  content but different AGR codon usage.

Figure 2 shows the results of this comparison. There appears to be a selective pressure for codon usage that is not directly dependent on  $C + G$  concentration. Many organisms with distinct codon biases for arginine share about the same  $C + G$  concentrations.

#### Arginine codon usage and $(E + K)/(H + Q)$ ratio

The correlation between AGR codon bias and the  $(E + K)/(Q + H)$  turned out to be strongly positive. Figure 3 shows that the increase in the ratio is followed by an increase in the bias for AGR codons. These observations confirm our suggestion that adaptations to thermostability include changes in codon usage. Whenever there is an increase in lysine usage (a relevant amino acid to thermostability), there is also a bias for AGR codons for arginine, consequently a positive error minimization mechanism between arginine and lysine codons. The decrease in CGN codon usage in organisms that live in high temperatures can be interpreted as a negative error minimization, since a single mutation could turn CGN arginine codons into codons for histidine and glutamine (which can be harmful for thermostability). Therefore, a single evolutionary event, namely the ARG codon bias for arginine, has a double effect in protecting thermostability of proteins, making amino acid usage and codon preference walk along the same way.

---

#### Conclusion

Our results suggest that codon bias have a role in protecting protein properties that ensure thermostability.

This bias is not a consequence of  $C + G$  variation, as evidenced by the fact that organisms with similar  $C + G$  proportions may have strongly different patterns of codon usage for arginine. Temperature adaptation of proteins therefore lies not only at the protein structure level, but also at the translation level, through mechanisms of error minimization.

---

#### References

- Anderson DE, Hurley JH, Nicholson H, Baase WA, Matthews BW (1993) Hydrophobic core repacking and aromatic-aromatic interaction in the thermostable mutant of T4 lysozyme Ser 117. *Protein Sci* 2(8):1285–1290
- Berezovsky IN, Tumanyan VG, Esipova NG (1997) Representation of amino acid sequences in terms of interaction energy in protein globules. *FEBS Lett* 418(1–2):43–46, 24 November 1997
- Farias ST, van der Linden MG, Rego TG, Araujo DA, Bonato MC (2004) Thermo-search: lifestyle and thermostability analysis. *Silico Biol* 4(3):377–380
- Farias ST, Bonato MC (2003) Preferred amino acids and thermostability. *Genet Mol Res* 2(4):383–393, 30 December 2003
- Jaenicke R (2000) Do ultrastable proteins from hyperthermophiles have high or low conformational rigidity? *Proc Natl Acad Sci USA* 97(7):2962–2964, 28 March 2000
- Mita K, Ichimura S, Nenoï M (1991) Essential factors determining codon usage in ubiquitin genes. *J Mol Evol* 33(3):216–225
- Porter TD (1995) Correlation between codon usage, regional genomic nucleotide composition, and amino acid composition in the cytochrome P-450 gene superfamily. *Biochim Biophys Acta* 1261(3):394–400, 26 April 1995
- Querol E, Perez-Pons JA, Mozo-Villarias A (1996) Analysis of protein conformational characteristics related to thermostability. *Protein Eng* Mar 9(3):265–271
- Schumann J, Böhm G, Schumacher G, Rudolph R, Jaenicke R (1993) Stabilization of creatinase from *Pseudomonas putida* by random mutagenesis. *Protein Sci* 2(10):1612–1620
- Thompson MJ, Eisenberg D (1999) Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability. *J Mol Biol* 290(2):595–604, 9 July 1999
- Ventriani C, Maeder DL, Tolliday N, Yip KSP, Stillman TJ, Britton KL, Rice DW, Klump HH, Robb FT (1998) Protein thermostability above 100°C: a key role ionic interactions. *Proc Natl Acad Sci USA* 95:12300–12305