ELSEVIER

FEMS Microbiology Letters 200 (2001) 79–84

FEMS MICROBIOLOGY Letters

www.fems-microbiology.org

On the origin of Ser/Thr kinases in a prokaryote

Guoniu Han [a], Cheng-Cai Zhang [b,*]

[a] *Institut de Recherche Mathématique Avancée, Université Louis Pasteur, C.N.R.S., 7 rue René Descartes, F-67084 Strasbourg, France*
[b] *Laboratoire de Chimie Bactérienne, UPR9043-C.N.R.S., 31 Chemin Joseph Aiguier, F-13402 Marseille Cedex 20, France*

Received 7 March 2001; received in revised form 19 April 2001; accepted 19 April 2001

First published online 11 May 2001

Abstract

The family of Ser/Thr and/or Tyr kinases and that of His kinases play essential roles in signal transduction. For a long time, the former has been found in eukaryotes, the latter in prokaryotes. Studies in the last decade have shown, however, that most bacteria possess from one to more than 10 genes encoding Ser/Thr kinases. This observation raises an important question concerning the evolutionary origin of Ser/Thr kinases found in bacteria. To answer this question, we have analyzed a family of 11 genes encoding Ser/Thr kinases in the cyanobacterium *Synechocystis* sp. PCC 6803. This bacterium contains the largest number of Ser/Thr kinases among all bacteria whose genomic sequences have been released so far. In this study, we have developed a user-friendly computer program for statistical analysis of codon usages and GC content. The results demonstrate that Ser/Thr kinases have similar codon usages and GC contents as the average of all possible open reading frames (ORFs) deduced from the genome. In contrast, ORFs encoding transposases, as a control in our analysis, display a disparity in both codon usage and GC content, confirming their multiple origin and genetic promiscuity. In light of our results, we propose that Ser/Thr kinases existed before the divergence between prokaryotes and eukaryotes during evolution, or were laterally transferred into prokaryotes at the early stages of bacterial evolution. If Ser/Thr kinases have persisted ever since in prokaryotes under evolutionary pressure, it is then expected that they play important, possibly even essential roles in regulating bacterial activities as do their counterparts in eukaryotes.

Keywords: Ser/Thr kinase; Genomics; Gene transfer; Evolution; Cyanobacterium

1. Introduction

Signal transduction is an essential mechanism required for cell adaptation to changing environments. Bacterial cells use two-component systems for signal transduction and these systems are found in almost all bacterial strains [1–3]. A typical two-component system contains a His kinase and a response regulator. Upon ligand binding, the His kinase autophosphorylates on a conserved His residue and then transfers the phospho-group to a conserved Asp residue of the response regulator. As a comparison, eukaryotic cells employ Ser/Thr and Tyr kinases for signal transduction [4,5]. Thus protein phosphorylation is the predominant form of signaling in both prokaryotes and eukaryotes, but the enzymes involved have been considered different for a long time. This view is now changing as Ser/Thr and Tyr kinases have been reported in many bacterial strains during the last few years [6,7].

Protein phosphorylation was first discovered in eukaryotes in the mid-1950s (for a review, see [5]). A large family of Ser/Thr and Tyr kinases is now found to operate in a eukaryotic cell, and many of these enzymes form signaling networks for the coordination of various cellular processes. Ser/Thr and Tyr phosphorylation in prokaryotes was found much later, in the late 1970s, through biochemical analysis (for a review, see [8]). In 1991, the first example of protein Ser/Thr and Tyr kinases similar to those of eukaryotic ones was identified in the bacterium *Myxococcus xanthus* by a molecular approach [9]. Since then several laboratories have isolated genes encoding eukaryotic-like protein kinases from different bacterial species [6,10]. Analyses of more than 20 bacterial genomes sequenced so far indicate that, contrary to earlier belief, the presence of eukaryotic-type protein kinases in bacteria is a general rule rather than an exception [7]. It is true though that His kinases of two-component systems outnumber Ser/Thr

* Corresponding author. Tel.: +33 (4) 9116 4096;
Fax: +33 (4) 9171 8914; E-mail: cczhang@ibsm.cnrs-mrs.fr

0378-1097/01/$20.00 © 2001 Federation of European Microbiological Societies. Published by Elsevier Science B.V. All rights reserved.
PII: S0378-1097(01)00206-3

kinases in any given bacterial strain, in contrast to the dominant presence of Ser/Thr and Tyr kinases in eukaryotes.

The discovery of Ser/Thr kinases in many bacterial species raises one fundamental question about the evolutionary origin of such enzymes in evolution. Genes encoding Ser/Thr kinases could either be genuine prokaryotic enzymes or they were acquired through horizontal gene transfer from eukaryotes. The answer to this question will provide important guidance to the functional analysis of eukaryotic-like Ser/Thr kinases in bacteria. If Ser/Thr kinases existed before the divergence of prokaryotes and eukaryotes and were maintained under selective pressure during evolution, one would argue that these enzymes play important roles in regulating cell activities in prokaryotes. On the contrary, if Ser/Thr kinases were acquired later from eukaryotes through one or multiple horizontal gene transfer events, they may then play mostly secondary roles for the molecular fitness or for cell activities added later in prokaryotes.

In order to get an insight into the question of the evolutionary origin of Ser/Thr kinases in prokaryotes, we have used statistical analysis to compare several parameters between a gene family of Ser/Thr kinases and other genes on the cyanobacterial genome *Synechocystis* sp. PCC 6803. The genome of *Synechocystis* PCC 6803 was chosen for several reasons. Firstly, this genome has been completely sequenced and is thus suitable for statistical analysis on a genomic scale [11]. Secondly, it contains a relatively large number of Ser/Thr kinase genes compared to other bacterial strains [7], which can generate relevant statistical values. Thirdly, cyanobacteria are of ancient origin, and signs of their presence were recorded in fossils as early as more than 3 billion years ago [12]. Our studies strongly suggest that Ser/Thr kinase genes in cyanobacteria are genuine prokaryotic ones as they share similar values of GC contents and codon usages as other cyanobacterial genes on average.

2. Materials and methods

All genomic information of *Synechocystis* sp. PCC 6803 can be accessed on the CyanoBase internet site (http://www.kazusa.or.jp/cyano/cyano.html). According to various reports [7,13], 11 genes encoding Ser/Thr kinases can be found on the genome of *Synechocystis* PCC 6803, and they are compiled here as sll1770, sll0005, slr0889, slr1919, slr1225, slr1443, sll1575, slr0152, slr1697, slr0599, [7,13]. Ninety-nine genes encoding transposases are recorded in the CyanoBase and were used for comparative studies [11].

For the comparison of GC content, an average GC content was first obtained by analyzing all open reading frames (ORFs) of the genome. The difference in GC content between a given gene and the average was then calculated according to the following mathematical model: $M_{\mathrm{GC}} = (\mathrm{GC}_a - \mathrm{GC}_g)^2$, where GC_a is the average GC content of all ORFs deduced from the whole genome, and GC_g is the GC content of a given ORF. If the M_{GC} value is zero or close to zero, it means that the GC content of the gene is the same as or close to the average GC content of the whole genome. The bigger the M_{GC} value, the larger the difference in GC content between this gene and the genomic average. The distribution of M_{GC} values was then generated and compared.

For the comparison of codon usages, the average usage in each codon was calculated for all the ORFs on the genome. For each ORF, the usage of each codon was compared to the average, and this difference of all codon usages is summed up by using the following mathematical formula: $M_{\mathrm{CU}} = (\mathrm{CU}_a1 - \mathrm{CU}_g1)^2 + (\mathrm{CU}_a2 - \mathrm{CU}_g2)^2 + (\mathrm{CU}_a3 - \mathrm{CU}_g3)^2, \ldots + (\mathrm{CU}_a64 - \mathrm{CU}_g64)^2$. In this formula, CU_a is the average usage of one codon for the whole genome, CU_g is the usage of the same codon for a given gene, and 1–64 depict numbers 1–64 of all possible codons. M_{CU} is thus the sum of all differences in codon usages between a specific gene and all genes on the genome. The bigger the M_{CU} value, the larger the difference in codon usage between this gene and the genomic average. The distribution of M_{CU} was then generated and analyzed.

3. Results

3.1. Analysis of GC content and codon usages of the whole genome of Synechocystis sp. PCC 6803

In order to analyze statistically the codon usages and GC contents of ORFs in *Synechocystis* sp. PCC 6803, a computer program, CodonWarrior, was written based on the two mathematical models described above. The coding sequences of all ORFs of *Synechocystis* sp. PCC 6803 were downloaded from the CyanoBase, and the codon usage and the GC content of any gene can be compared to those of the average of the whole genome by using this computer program. The codon usage and GC content of each ORF were compared to the average values generated from all ORFs. These differences, expressed as M_{CU} (for codon usages) and M_{GC} (for GC content), respectively, as outline above, are shown in Figs. 1 and 2.

A bell-shaped distribution curve for M_{CU} values was obtained (Fig. 1A). Most ORFs give a M_{CU} value less than 20, indicating that they are not very different from the average of codon usages of the genome. However, some ORFs gave a M_{CU} value as big as 90, meaning that they display an extremely different codon usage as compared to the average of the genome. These genes could be either acquired by *Synechocystis* sp. PCC 6803 through horizontal gene transfer, or strongly regulated in their expression level through the bias of codon usage [14,18].

Since the M_{CU} value represents the sum of differences in

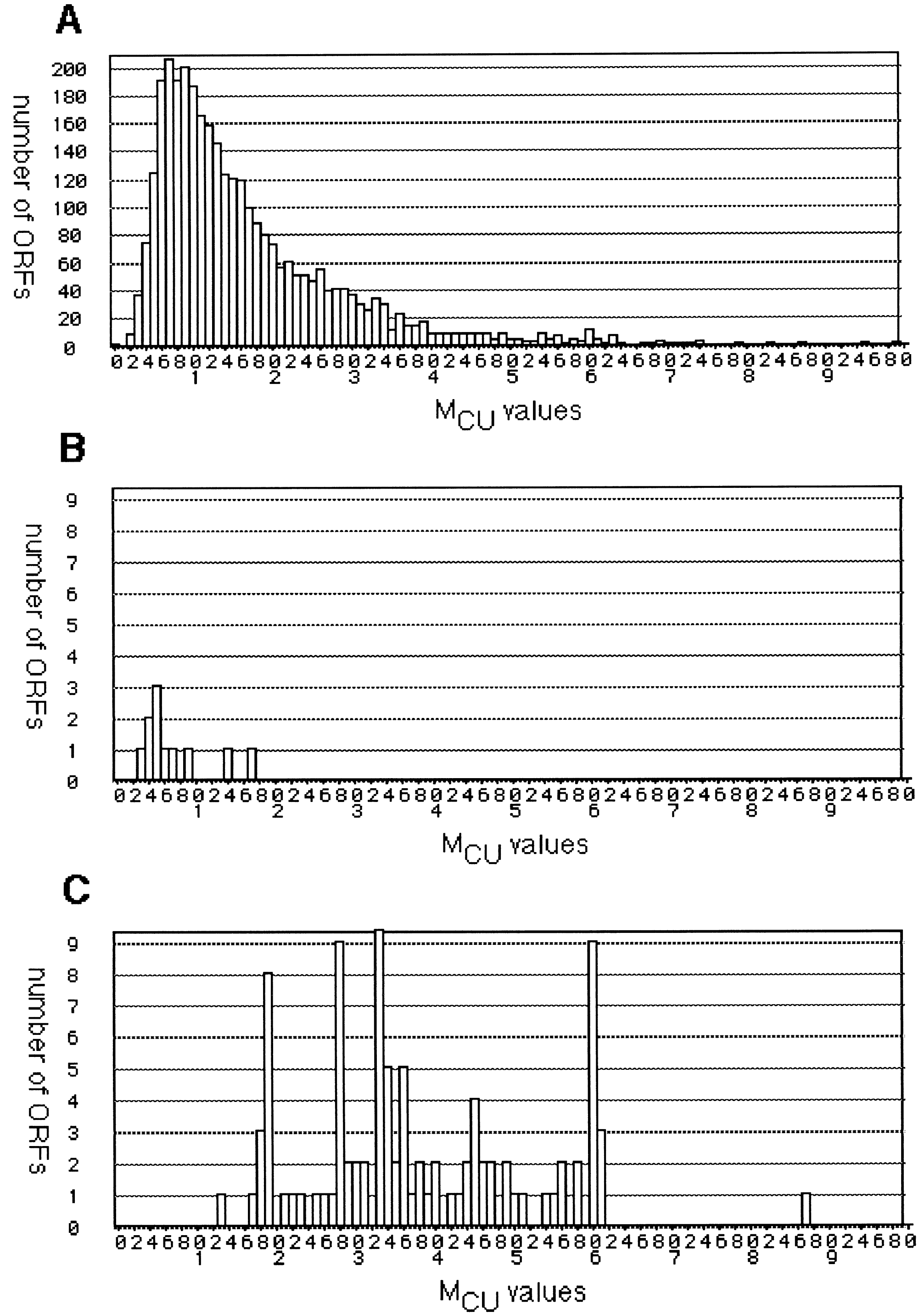

Fig. 1. Distribution of M_{CU} values reflecting codon usages of different ORFs on the genome of *Synechocystis* sp. PCC 6803. The M_{CU} value of each ORF, generated from the mathematical model described in the text, was plotted against the number of ORFs. For example, there are about 200 ORFs on the genome displaying a M_{CU} value of 6. A: Distribution of M_{CU} values of all ORFs deduced from the genome. B: M_{CU} values of the gene family encoding Ser/Thr kinases. C: M_{CU} values of genes encoding transposases.

codon usage of up to 64 codons, one could ask whether it is valid in reflecting the overall difference in usages of individual codons. For that reason, two chosen ORFs, representing two different ranges of M_{CU} values (7 for slr1697, and 60 for sll0699), were analyzed in detail. The usage of each codon was compared between a given ORF and the genomic average. As shown in Fig. 3, the higher the M_{CU} value, the more marked the difference between usages in individual codons as compared to the genomic average. Thus the M_{CU} values do reflect the overall difference in codon usages as well as difference in usages of individual codons.

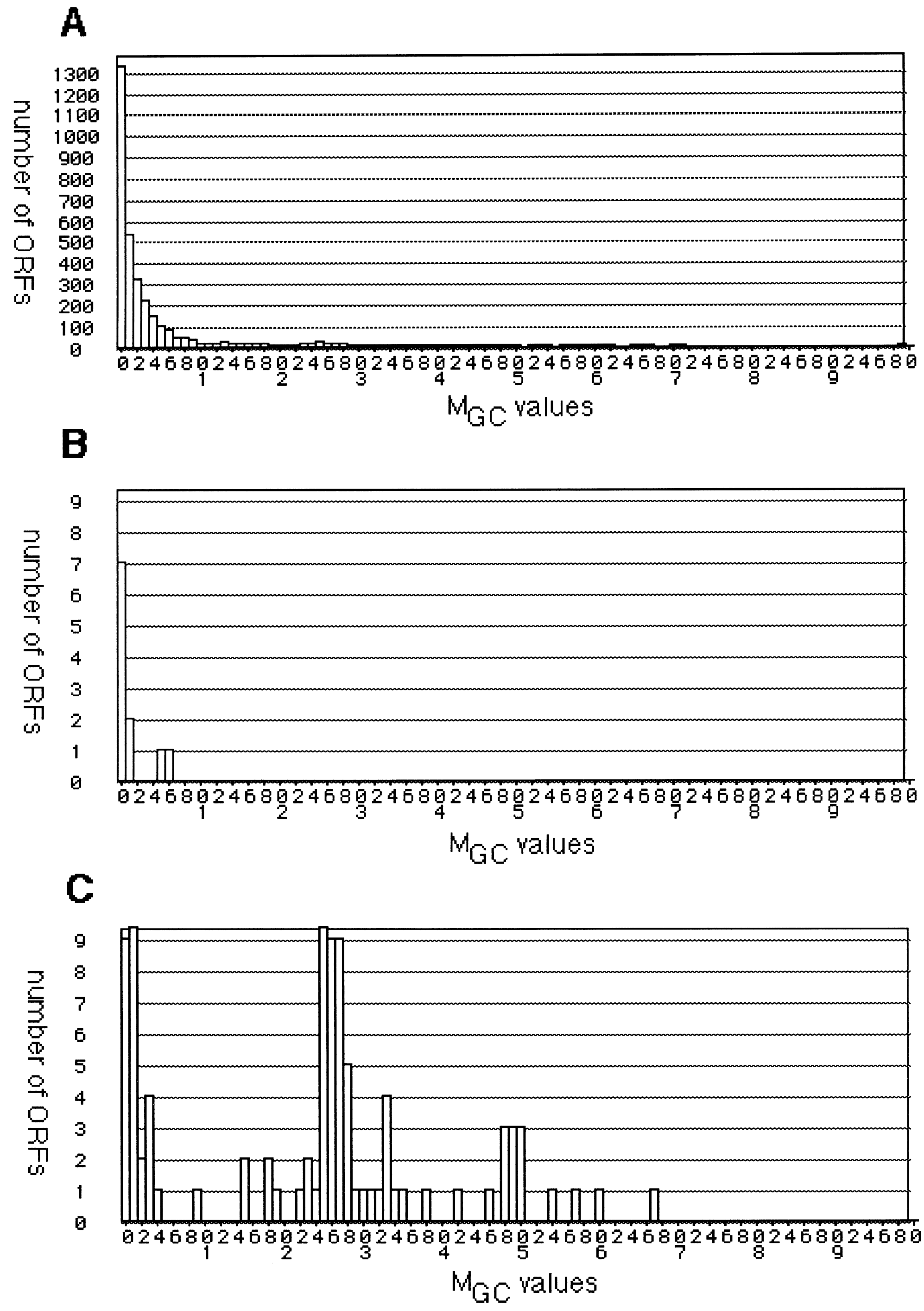

Fig. 2. Distribution of M_{GC} values based of the GC content of coding sequences of genes on the genome of *Synechocystis* sp. PCC 6803. See text for the calculation of GC content. A: Distribution of M_{GC} values of all ORFs deduced from the genome. B: M_{GC} values of the gene family encoding Ser/Thr kinases. C: M_{GC} values of genes encoding transposases.

The GC content of each ORF was also compared to the average GC content of the whole genome. As shown in Fig. 2A, the GC content of the majority of ORFs in the genome of *Synechocystis* sp. PCC 6803 is the same as, or very close to, the overall GC content of the whole genome. A few ORFs have a GC content very different from that of the whole genome, and they may represent the results of horizontal gene transfer events.

A good correlation between M_{GC} and M_{CU} could be found for each ORF analyzed in this study. For example, slr1697, sll0700 and sll0699 gave M_{CU} values of 7, 28 and 60, respectively, and M_{GC} values of 0, 25 and 27, respectively. These results further validate our approach in analyzing the relationship among different gene families on a bacterial genome.

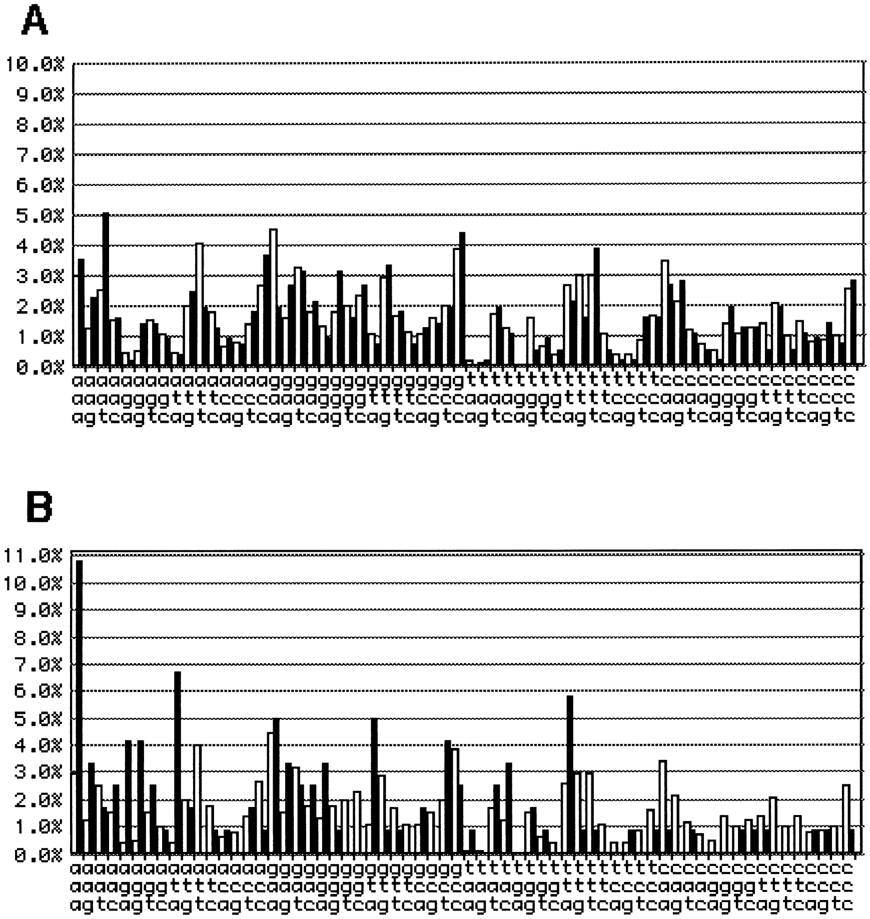

Fig. 3. Comparison of codon usages of two individual ORFs, slr1697 (A) and sll0699 (B), with the average codon usages of all ORFs on the genome of *Synechocystis* PCC 6803. Each codon is given below, and the usage of each codon (in %) is indicated on the left on each panel. The open bar depicts the codon usages of the whole genome, and the filled bar the codon usages of one specific ORF. slr1697 displays M_{CU} and M_{GC} values of 7 and 0, respectively, and the M_{CU} and M_{GC} values for sll0699 were 60 and 27, respectively.

3.2. *Analysis of the Ser/Thr kinase gene family*

Both the GC content and the codon usages were analyzed for all members of the gene family encoding eukaryotic-like protein kinases in *Synechocystis* sp. PCC 7120. Both parameters of all members of this gene family are close to the average of the whole genome. Their M_{CU} values range from 3 to 7, similar to the majority of ORFs in the genome (Fig. 1B). Similarly, their M_{GC} values obtained were from 0 to 6, which also fall within the range of those of most ORFs of the genome (Fig. 2B).

3.3. *Genetic mosaics of genes encoding transposable elements*

To further validate our results, we have also analyzed all deduced ORFs related to genes encoding transposases on the genome of *Synechocystis* sp. PCC 6803 [11]. Transposon elements are known for their great genetic promiscuity and could be transferred among different bacterial strains [15,16,18]. As shown in Figs. 1 and 2, the distribution of M_{CU} and M_{GC} values derived from genes encoding transposases is random and does not show a bell shape, in contrast to those of all ORFs on the genome. Their M_{CU} values range from 13 to 87 (Fig. 1C), and M_{GC} values from 0 to 67 (Fig. 2C). The disparity in M_{GC} and M_{CU} values for genes encoding transposases confirms the multiple origin of these genes, as expected from such mobile genetic elements [15,16].

4. Discussion

Horizontal gene transfer is one important force for generating genetic diversity in living organisms. Such events are believed to have occurred among eukaryotes and prokaryotes, as well as between eukaryotes and prokaryotes [16]. These lateral transfers could enable an organism to acquire new characters and new possibilities to colonize different environments. By certain estimation with sequenced bacterial genomes, the range of foreign DNA varies from 0.0 (*Mycoplasma genitalium*) to as high as 16.6% (*Synechocystis* sp. PCC 6803) of total genomic information [16]. It thus seems that horizontal gene transfer events have been particularly frequent for cyanobacteria.

Our statistical analyses presented here strongly suggest that Ser/Thr kinases in *Synechocystis* sp. strain PCC 6803 are genuine prokaryotic enzymes since their corresponding genes show a GC content and codon usages similar to the averages of all genes from the genome. Or at least, they were acquired by cyanobacteria through lateral gene transfer at the early stage of their evolution in order for Ser/Thr kinase genes to adapt to the genetic background of this bacterial strain. The discovery in almost all sequenced bacterial genomes of genes encoding putative Ser/Thr kinases argues rather against the idea that they were acquired through horizontal gene transfer events from eukaryotes [6]. All of these observations favor the possibility that Ser/Thr kinases existed before the divergence between prokaryotes and eukaryotes in evolution. The persistence of genes encoding Ser/Thr kinases in prokaryotes despite the evolutionary pressure suggests that they are required for cell growth. It can then be expected that some Ser/Thr kinases are important or even essential in regulating bacterial activities.

Although computer programs for analyzing GC content and codon usages have existed for many years, most of them are rather complex to use [14]. The computer program described here, CodonWarrior, is small in size and easy to handle. It can be used for other analytical purposes as well. For example, the comparison of codon usages, between two ORFs or between one ORF and the average of a whole genome, can be easily performed with this program (Fig. 3). Such information may be of help in cases where an optimized-expression level is sought in a heterologous expression system by recognizing biased codon usages that often limit the expression level of recombinant proteins [17].

While our article was being reviewed, Mrazek et al. published a paper predicting highly expressed or horizontally transferred genes in *Synechocystis* sp. PCC 6803 [18]. Their study was also based on the analysis, on a genomic scale, of the codon usage of all genes in this cyanobacterium. This analysis, together with ours, should provide guidance for functional genomic study in cyanobacteria.

Acknowledgements

This work was supported by CNRS and the ATIPE program. We would like to thank Dr. Y. Quentin for helpful discussion.

References

[1] Alex, L.A. and Simon, M.I. (1994) Protein kinase and signal transduction in prokaryotes and eukaryotes. Trends Genet. 10, 133–138.

[2] Dutta, R., Qin, L. and Inoue, M. (1999) Histidine kinases: diversity of domain organization. Mol. Microbiol. 34, 633–640.

[3] Stock, A.M., Robinson, V.L. and Goudreau, P.N. (2000) Two-component signal transduction. Annu. Rev. Biochem. 69, 183–215.

[4] Hanks, S.K. and Hunter, H. (1995) The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. FASEB J. 9, 576–596.

[5] Hunter, T. (2000) Signaling – 2000 and beyond. Cell 100, 113–127.

[6] Zhang, C.-C. (1996) Bacterial signalling involving eukaryotic-type protein kinases. Mol. Microbiol. 20, 9–15.

[7] Shi, L., Potts, M. and Kennelly, P.J. (1998) The serine, threonine, and/or tyrosine-specific protein kinases and protein phosphatases of prokaryotic organisms: a family portrait. FEMS Microbiol. Rev. 22, 229–252.

[8] Cozzone, A.J. (1988) Protein phosphorylation in prokaryotes. Annu. Rev. Microbiol. 42, 97–125.

[9] Munoz-Dorado, J., Inouye, M. and Inouye, S. (1991) A gene encoding protein serine/threonine kinase is required for normal development of *M. xanthus*, a gram-negative bacterium. Cell 67, 995–1006.

[10] Kennelly, P.J. and Potts, M. (1996) Fancy meeting you here! A fresh look at 'prokaryotic' protein phosphorylation. J. Bacteriol. 178, 4759–4764.

[11] Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M. and Tabata, S. (1996) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. DNA Res. 3, 109–136.

[12] Schopf, J.W. and Packer, B.M. (1987) Early archean (3.3-billion to 3.5-billion-years-old) microfossils from Warsanoona group, Australia. Science 237, 71–73.

[13] Zhang, C.-C., Gonzalez, L. and Phalip, V. (1998) Survey, analysis and genetic organization of genes encoding eukaryotic-like signalling proteins by a cyanobacterial genome. Nucleic Acids Res. 26, 3619–3625.

[14] Moszer, I., Rocha, E.P.C. and Danchin, A. (1999) Codon usage and lateral gene transfer in *Bacillus subtilis*. Curr. Opin. Microbiol. 2, 524–528.

[15] Liebert, C.A., Hall, R.M. and Summers, A.O. (1999) Transposon Tn21, flagship of the floating genome. Microbiol. Mol. Biol. Rev. 63, 507–522.

[16] Ochman, H., Lawrence, J.G. and Groisman, E.A. (2000) Lateral gene transfer and the nature of bacterial innovation. Nature 405, 299–304.

[17] Hannig, G. and Makrides, S.C. (1998) Strategies for optimized heterologous protein expression in *Escherichia coli*. Trends Biotechnol. 16, 54–60.

[18] Mrazek, J., Bhaya, D., Grossman, A.R. and Karlin, S. (2001) Highly expressed and alien genes of the *Synechocystis* genome. Nucleic Acids Res. 29, 1590–1601.