CHAPTER 5

Influence of Codon Bias on the Expression of Foreign Genes in Microalgae

Markus Heitzer,* Almut Eckert, Markus Fuhrmann and Christoph Griesbeck

Abstract

The expression of functional proteins in heterologous hosts is a core technique of modern biotechnology. The transfer to a suitable expression system is not always achieved easily because of several reasons: genes from different origins might contain codons that are rarely used in the desired host or even bear noncanonical codons, or the genes might hide expression-limiting regulatory elements within their coding sequence. These problems can also be observed when introducing foreign genes into genomes of microalgae as described in a growing number of detailed studies on transgene expression in these organisms. Particularly important for the use of algae as photosynthetic cell factories is a fundamental understanding of the influence of a foreign gene's codon composition on its expression efficiency. Therefore, the effect of codon usage of a chimeric protein on expression frequency and product accumulation in the green alga *Chlamydomonas reinhardtii* was analyzed. This fusion protein combines a constant region encoding the zeocin binding protein Ble with two different gene variants for the green fluorescent protein (GFP). It is shown that codon bias significantly affects the expression, but barely influences the final protein accumulation in this case.

General Aspects of Codon Bias in Pro- and Eukaryotic Expression Hosts

The genetic code is degenerated, i.e., there are multiple codons for the same amino acid. This allows organisms to select a subset of codons to efficiently encode all their proteins and to leave other base triplets underrepresented. Moreover, within one single organism the codon bias can vary between abundantly and moderately expressed genes and between long and short genes. For *E. coli*, the far most used prokaryotic expression host, the codon usage observed especially for Gly, Arg and Pro in the mRNA population of highly abundant genes is closely reflected by the corresponding tRNA populations.[1] Such a severe codon bias has been discussed for a long time as one possible hindrance in the high level expression of heterologous genes in general. It was reasoned that insufficient tRNA pools can lead to translational stalling, premature translation termination, translation frame shifting and amino acid misincorporation.[2] For example, multiple consecutive rare codons near the N-terminus of a coding sequence led to significant reduction in the expression efficiency of this gene.[3] In another case, supplementing the endogenous tRNA population of *E. coli* by additional plasmid encoded copies caused a strong increase in the expression yield of human tissue plasminogen activator.[4]

*Corresponding Author: Markus Heitzer—Universität Regensburg, Kompetenzzentrum für Fluoreszente Bioanalytik, Josef - Engert - Str. 9, 93053 Regensburg, Germany. Email: markus.heitzer@exfor.uni-regensburg.de

Transgenic Microalgae as Green Cell Factories, edited by Rosa León, Aurora Galván and Emilio Fernández. ©2007 Landes Bioscience and Springer Science+Business Media.

Effects of codon bias on the efficiency of endogenous and foreign gene expression are not limited to prokaryotic hosts, but have also been observed in different yeast expression systems,[5-7] higher plants like maize[8] or *Arabidopsis*,[9] animals like *Drosophila*[10] and mammalian cell lines.[11]

Although there have been several reports about the introduction of foreign genes into the genomes of various microalgae, stable genetic transformation of microalgal genomes is achieved routinely only for very few species. The use of microalgae for the efficient production of valuable compounds like recombinant proteins or their metabolic engineering using synthetic enzyme variants will depend largely on the efficiency of foreign gene expression. This efficiency is likely to be influenced by the codon composition of the transgene, as it is the case for many other organisms. Clearly, the detailed analysis of codon usage on expression has always to rely on wide information about the genome and even more the frequency of individual codons in abundant and average mRNAs. Yet, this combination of knowledge is available only for the diatom *Phaeodactylum tricornutum* and the green alga *Chlamydomonas reinhardtii*. For some other species genetic transformation has been reported, but the necessary information about codon frequency in mRNAs is restricted. Moreover, for some algae the number of experiments on transgene expression is still limited, or only transient expression has been observed (see also Chapter 1). Therefore judging the influence on expression efficiency of transgenes for *Cyclotella, Navicula, Cylindrotheca* or *Thalassiosira* species is still highly speculative.

Phaeodactylum tricornutum

The first reports about transformation of the *P. tricornutum* nuclear genome with the bacterial phleomycin resistance gene *ble* and the reporter chloramphenicol acetyltransferase demonstrated a stable integration into the nuclear genome.[12] These findings suggested a continuous expression of the foreign genes without the need for constant selection regardless of any differences in codon usage between donor and host organism. In contrast, similar experiments indicated unstable expression of the bacterial *S. hindustanus ble* gene in about 80% of all lines.[13] These effects are likely to be explained by one or more different epigenetic silencing mechanisms, a common phenomenon in transgene expression in plants and algae. A general inactivation of foreign genes could be excluded by the analysis of a second reporter gene, firefly luciferase. This gene was introduced in parallel and did not show a corresponding silencing effect. Both genes again originate from heterologous sources and show a different codon usage. The bacterial *ble* presents a very high GC-content (~70%), whereas the luciferase gene from *Photinus* has a much lower GC-content (~37%). Surprisingly, the average codon usage for *Phaeodactylum* as calculated from the Kazusa codon usage data base (http://www.kazusa.or.jp/codon/) is intermediate (54%), but with a preference for cytosine in the third position particularly in abundant transcripts.[14] In further reports, the expression of several GFP genes and glucose uptake proteins from different organisms were analyzed in detail. Wildtype *gfp* from *Aequorea* and a codon modified variant for maximal expression in *Arabidopsis* (*mgfp4*) did not result in detectable protein, but visible amounts of GFP were obtained by using *egfp*, a codon optimized form for human cells. Since contrary to *gfp* and *mgfp4* the codon composition of *egfp* is very close to that of *P. tricornutum*, the successful expression of *egfp* could be directly attributed to its codon bias.[15] In a second approach, the authors found that yeast glucose transporters (Hxt1, Hxt2 and Hxt4, ~39% GC) were not expressed, but the human erythrocyte Glut1 (~59% GC) and the *Chlorella* Hup1 (~61% GC) hexose transporters were produced efficiently, allowing heterotrophic growth of *P. tricornutum* on glucose as the only carbon source. The authors speculated, that these results "may reflect differences in codon usage between yeast and *P. tricornutum*".[16]

Chlamydomonas reinhardtii —Expression from Chloroplast and Nucleus

One microalga that has attracted attention as a promising cell factory system is *Chlamydomonas reinhardtii*, especially in view of the broad genetic and molecular toolset available for this organism. However, it has often been speculated that in this species a strong bias for specific codons in highly expressed nuclear and chloroplastic genes may significantly influence the expression efficiency of foreign genes after integration into any of the genomes. For example, only genes with a high GC-content similar to that found in *C. reinhardtii* (GC~61%) and with a low percentage of rarely used codons could be established as nuclear selection markers.[17]

Some systematic studies have been carried out for genes expressed from the chloroplast genome. It was demonstrated that a variant of GFP adapted to the chloroplast codon usage accumulated about 80-fold more in chloroplasts of transformed strains as compared to the non adapted gene.[18] The integration into the chloroplast genome occurs via homologous recombination, and therefore positional effects of genome integration can be excluded. Further experiments using different genes (for bacterial luciferase[19] and a full length single chain antibody)[20,21] optimized for the chloroplast codon usage were comparably successful. These findings suggest a significant influence of codon composition on the expression efficiency of a transgene when incorporated into the chloroplast genome of *Chlamydomonas* (see also Chapters 4 and 8).

Comparable experiments have been performed for nuclear encoded genes in *Chlamydomonas*. A first report using a synthetic gene encoding the green fluorescent protein from *Aequorea victoria* indicated a positive effect of the adaptation of codons to the nuclear bias of the alga.[22] In this study, no comparison with the AT-rich original gene was possible, because expression of the unmodified gene failed under the same promoter. Equal results were obtained for GFP expression in the closely related multicellular green alga *Volvox carteri* (GC~60%), where only the codon optimized gene variant produced a detectable amount of protein.[23] The successful use of synthetic genes with improved codon usage was continued in further experiments using a synthetic gene for the luciferase of *Renilla reniformis* adapted to the nuclear codon frequency of *C. reinhardtii*.[24]

In order to study the influence of codon usage on foreign gene expression from the nuclear genome in more detail, we compared the production of two variants of the GFP, the original AT-rich variant (*mgfp*, 39% GC) of the *Aequorea* gene and the synthetic *cgfp* (62% GC) with *Chlamydomonas* adapted codon usage (Fig. 1).[22] The *mgfp* was created from the plasmid pGFP (BD Clontech, USA) by introducing three individual mutations (F64L, S65T and T203I), which were also present in *cgfp*. As expression of *mgfp* failed before, both genes were cloned in frame with the bacterial *ble* gene from *S. hindustanus* in the previously described vector pSP124S-M,[22] allowing the direct selection of zeocin resistant clones expressing the Ble-GFP fusion proteins from the same constitutive promoter. Four independent transformations of *Chlamydomonas cw15 arg*⁻ were carried out as described[24] with the two gene variants transformed in parallel using the same culture divided into equal portions for comparison. Selection of zeocin resistant colonies was performed after an 18 h recovery phase in liquid nonselective medium on standard TAP (Tris-Acetate-Phosphate) medium supplemented with arginine (90 μg/ml) and zeocin (5-20 μg/ml) for about 10-14 days under constant illumination at 25°C.

The number of zeocin resistant clones was determined after two and four weeks (Table 1). This number was dependent on the variant of *gfp* that had been used for transformation, as the codon adapted gene resulted in a nearly 5-fold elevated number of resistant colonies. Obviously, there was a difference between the two lines of transformants, as determined by their growth after transfer from nonselective to selective conditions: whereas most of the cGFP-transformants showed very good growth on 5 μg/ml of zeocin (23 of 27 lines) none of the mGFP-strains grew as well. In this experiment 5 μl of liquid culture pregrown in nonselective medium for one week was spotted onto a selective plate and growth was monitored after five days. All three mGFP lines showed only sparse spotty growth, where most cGFP lines grew to dense green colonies (Table 2). This growth difference indicated a lower expression of

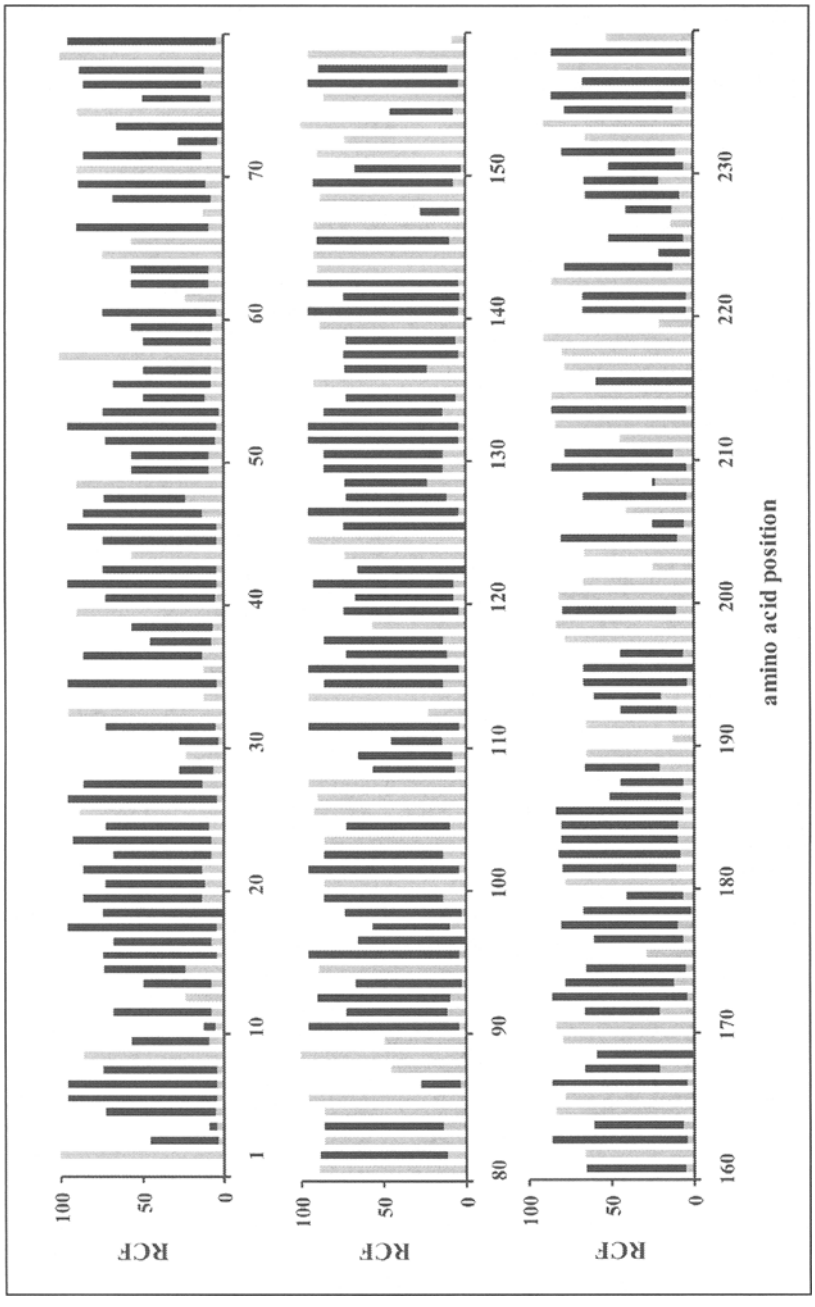

Figure 1. Codon usage of *mgfp* and *cgfp*. For each amino acid of the green fluorescent protein the relative frequency of individual codons (RCF) in the nuclear genome of *Chlamydomonas reinhardtii* is displayed. The *mgfp* sequence from *Aequorea* (in gray) contains several stretches with rare codons, which were replaced in the synthetic *cgfp* (in black) by adaptation of the gene to the nuclear codon usage of *Chlamydomonas*. The graphical codon usage analyser (gcua) tool was used for calculation.[24]

Table 1. Total numbers of zeocin resistant* Chlamydomonas reinhardtii *transformants obtained with two variants of* ble-gfp *from four independent transformations

To compare the transformation efficiency from different experiments, a comparison value (CV) was defined as

$$CV = \frac{number\ of\ clones\ (N)}{number\ of\ cells\ (n) \cdot amount\ on\ plasmid\ (p)} \left[\frac{1}{10^{-7} \cdot \mu g^{-1}}\right]$$

For each experiment 1-4·10^7 cells (n) were transformed with 0.6-2.5 µg plasmid (p).

Plasmid	N (Two Weeks)	CV	Expressing Clones (Further Cultivated)	N (Four Weeks)
pSP124S-mGFP	27	30	3	2
pMF124-cGFP	105	160	27	26

Table 2. Growth assay of primary transformants after a period of nonselective growth

Plasmid / Growth	+++	++	+	-
pSP124S-mGFP	0	0	2	1
pMF124-cGFP	20	3	3	1

Cells pregrown in liquid culture under nonselective conditions were transferred onto a selective plate. Growth was monitored after a period of five days in continuous light at 25°C and was grouped into four categories: +++ very dense growth, absolutely round shaped colony; ++ dense growth, not completely filled colony; + sparse growth, individual spots visible; - no growth detectable.

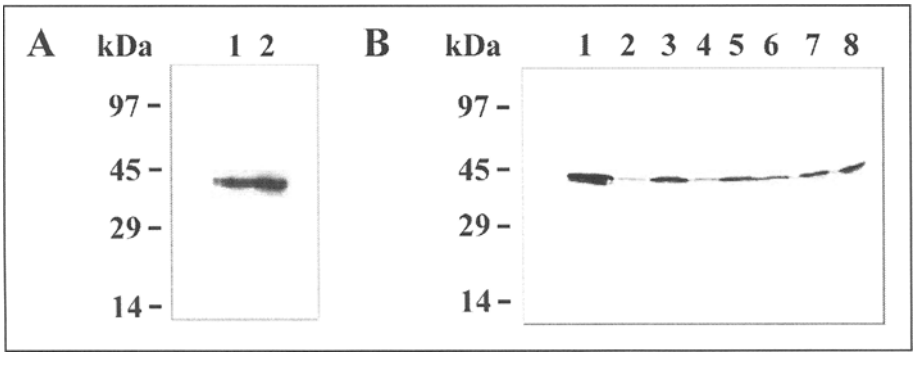

Figure 2. Western blot analysis of Ble-mGFP and Ble-cGFP lines with an anti-Ble-antibody. Samples with the same amount of protein according to a culture volume of approx. 30 µl were analyzed. A. Ble-mGFP lines. Lanes 1 and 2 correspond to clones AA1 and AA2. B. Ble-cGFP lines. Lanes 1-8 correspond to clones AA11 and AA15-21.

Ble-mGFP after a period of nonselective growth. Possible reasons for this observation could be active transcriptional or post-transcriptional gene silencing, but also stalled or prematurely terminated translation of the Ble-mGFP.

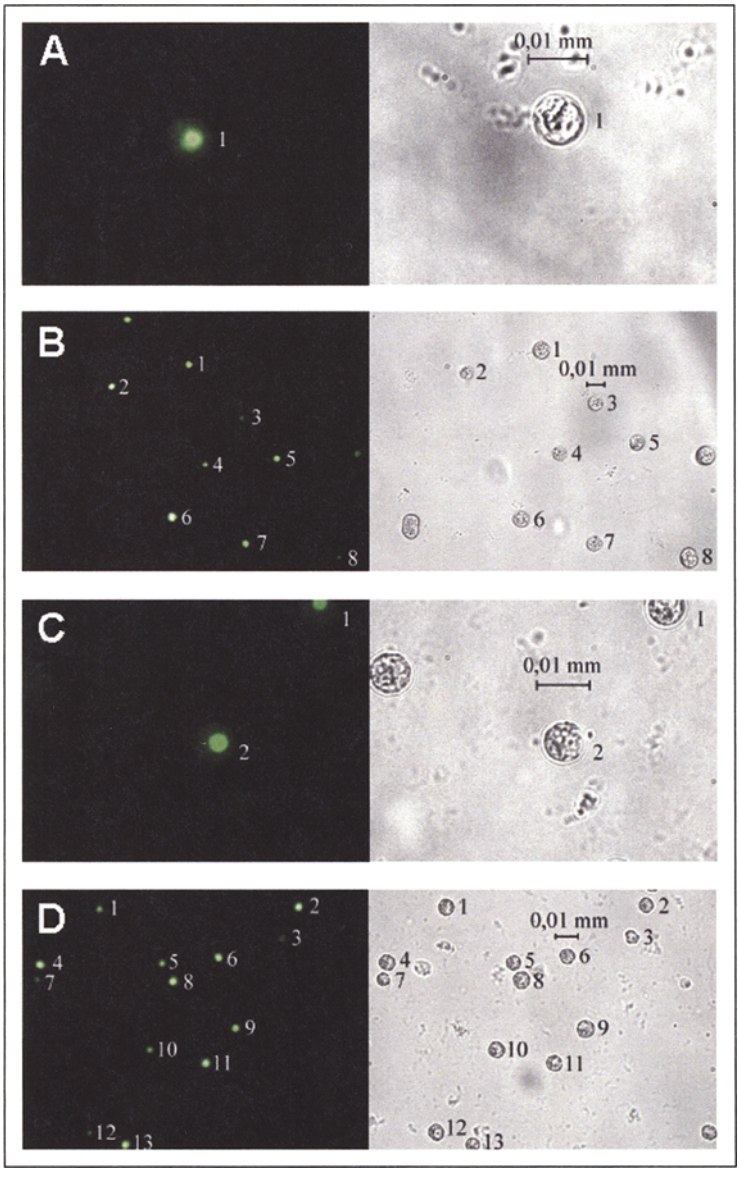

Figure 3. Fluorescence microscopy of Ble-mGFP and Ble-cGFP lines. A) and (B) Ble-mGFP clone AA2. The photographs show the localization of Ble-mGFP in the nucleus of a single cell (A) and with lower magnification expression levels in different cells (B) by fluorescence (left side) and light microscopy (right side). Identical cells are marked with the same numbers. C and D. Equal settings for Ble-cGFP clone AA11. There is no obvious difference detectable between these two strains. For detailed experimental procedures see reference 22.

To further analyze the expression of the Ble-GFP fusion protein both immunoblot assay and fluorescence microscopy were performed. Protein samples of all 28 lines grown in selective medium were separated by SDS-PAGE, transferred to nitrocellulose membranes and probed with a primary anti-Ble antibody (Cayla, France; dilution 1:1000). Detection of bound antibody was performed after three washing steps using an alkaline phosphatase conjugated secondary anti-rabbit-IgG antibody (Sigma, USA, dilution 1:1000) and a color precipitation assay. The results are shown in Figure 2. The Ble-GFP fusion protein was clearly detected at the correct size (43 kDa) for both resistant strains. No indication for premature termination of translation or protein degradation resulting in different products of various sizes could be observed. Also no significant difference between *mgfp* and *cgfp* lines in the amount of expressed protein as estimated from signal intensity was determined, although the expression levels of the *cgfp* clones covered a broad range.

The observed growth differences on selective medium could also result from a wide variation of expression efficiency within a single line, i.e., some cells showing a high expression level, some a very low. The Ble-GFP fusion protein can be directly visualized using fluorescence microscopy of living cells. We analyzed multiple sets of cells from different transformants of both gene variants and compared the expression levels (Fig. 3). No significant difference between the two types of transformants was obvious; all observed cells showed a comparably high level of GFP as determined by intensity measurement of fluorescence in the cell nucleus. However, an exact quantification of the three-dimensional fluorescence intensity and the conclusion for the expression level in the cell could not be achieved with this method.

Concluding Remarks

As observed in many other organisms, there are indications for a strong effect of codon usage on the expression efficiency of transgenes in *Chlamydomonas reinhardtii* for both nuclear and chloroplast encoded transgenes. Since the number of selectable transformants is significantly higher for codon optimized nuclear genes, the codon usage is likely to have an effect on the decision if a transgene is expressed or not and maybe also on the decision if silencing of this transgene takes place or not. So far, not adapted genes seem to be more susceptible to silencing effects. However, in clones that express transgenes, no differences regarding the expression quantity and quality can be observed between codon optimized and unmodified genes. But studies of the influence of codon bias on nuclear expression are always complicated by positional effects, since expression of transgenes strongly depends on the position within the genome and integration of foreign genes into the chromosomes takes place randomly in *C. reinhardtii*. Hence, a definite statement about codon effects cannot be made. Targeted integration of transgenes into the genome by homologous or Cre/*lox*-mediated site-specific recombination could be a helpful tool to circumvent this problem.[25]

References

1. Dong H, Nilsson L, Kurland CG. Covariation of tRNA abundance and codon usage in Escherichia coli at different growth rates. J Mol Biol 1996; 260(5):649-63.
2. Kurland C, Gallant J. Errors of heterologous protein expression. Curr Opin Biotechnol 1996; 7(5):489-93.
3. Goldman E, Rosenberg AH, Zubay G et al. Consecutive low-usage leucine codons block translation only when near the 5' end of a message in Escherichia coli. J Mol Biol 1995; 245(5):467-73.
4. Novy R, Drott D, Yeager K. Overcoming the codon bias of E. coli for enhanced protein expression. Innovations 2001; 12:1-3.
5. Sinclair G, Choy FY. Synonymous codon usage bias and the expression of human glucocerebrosidase in the methylotrophic yeast, Pichia pastoris. Protein Expr Purif 2002; 26(1):96-105.
6. Janatova I, Costaglioli P, Wesche J et al. Development of a reporter system for the yeast Schwanniomyces occidentalis: Influence of DNA composition and codon usage. Yeast 2003; 20(8):687-701.
7. Friberg M, von Rohr P, Gonnet G. Limitations of codon adaptation index and other coding DNA-based features for prediction of protein expression in Saccharomyces cerevisiae. Yeast 2004; 21(13):1083-93.

8. Viotti A, Balducci C, Weil JH. Adaptation of the tRNA population of maize endosperm for zein synthesis. Biochim Biophys Acta 1978; 517(1):125-32.
9. Chiapello H, Lisacek F, Caboche M et al. Codon usage and gene function are related in sequences of Arabidopsis thaliana. Gene 1998; 209(1-2):GC1-GC38.
10. Carlini DB, Stephan W. In vivo introduction of unpreferred synonymous codons into the Drosophila Adh gene results in reduced levels of ADH protein. Genetics 2003; 163(1):239-43.
11. Slimko EM, Lester HA. Codon optimization of Caenorhabditis elegans GluCl ion channel genes for mammalian cells dramatically improves expression levels. J Neurosci Methods 2003; 124(1):75-81.
12. Apt KE, Kroth-Pancic PG, Grossman AR. Stable nuclear transformation of the diatom Phaeodactylum tricornutum. Mol Gen Genet 1996; 252(5):572-9.
13. Falciatore A, Casotti R, Leblanc C et al. Transformation of nonselectable reporter genes in marine diatoms. Mar Biotechnol (NY) 1999; 1(3):239-251.
14. Montsant A, Jabbari K, Maheswari U et al. Comparative genomics of the pennate diatom Phaeodactylum tricornutum. Plant Physiol 2005; 137(2):500-13.
15. Zaslavskaia LA, Lippmeier JC, Kroth PG et al. Transformation of the diatom Phaeodactylum tricornutum (Bacillariophyceae) with a variety of selectable marker and reporter genes. Journal of Appl Phycol 2000; 36(2):379-386.
16. Zaslavskaia LA, Lippmeier JC, Shih C et al. Trophic conversion of an obligate photoautotrophic organism through metabolic engineering. Science 2001; 292(5524):2073-2075.
17. Leon-Banares R, Gonzalez-Ballester D, Galvan A et al. Transgenic microalgae as green cell-factories. Trends Biotechnol 2004; 22(1):45-52.
18. Franklin S, Ngo B, Efuet E et al. Development of a GFP reporter gene for Chlamydomonas reinhardtii chloroplast. Plant J 2002; 30(6):733-44.
19. Mayfield SP, Schultz J. Development of a luciferase reporter gene, luxCt, for Chlamydomonas reinhardtii chloroplast. Plant J 2004; 37(3):449-58.
20. Mayfield SP, Franklin SE, Lerner RA. Expression and assembly of a fully active antibody in algae. Proc Natl Acad Sci USA 2003; 100(2):438-42.
21. Mayfield SP, Franklin SE. Expression of human antibodies in eukaryotic micro-algae. Vaccine 2005; 23(15):1828-32.
22. Fuhrmann M, Oertel W, Hegemann P. A synthetic gene coding for the green fluorescent protein (GFP) is a versatile reporter in Chlamydomonas reinhardtii. Plant J 1999; 19(3):353-61.
23. Ender F, Godl K, Wenzl S et al. Evidence for autocatalytic cross-linking of hydroxyproline-rich glycoproteins during extracellular matrix assembly in Volvox. Plant Cell 2002; 14(5):1147-60.
24. Fuhrmann M, Hausherr A, Ferbitz L et al. Monitoring dynamic expression of nuclear genes in Chlamydomonas reinhardtii by using a synthetic luciferase reporter gene. Plant Mol Biol 2004; 55(6):869-81.
25. Zorin B, Hegemann P, Sizova I. Nuclear-gene targeting by using single-stranded DNA avoids illegitimate DNA integration in Chlamydomonas reinhardtii. Eukaryot Cell 2005; 4(7):1264-72.