Volume 325, number 1,2, 140–145 FEBS 12463 June 1993

Minireview

The mitochondrial genome on its way to the nucleus: different stages of gene transfer in higher plants

Axel Brennicke, Lutz Grohmann, Rudolf Hiesel, Volker Knoop and Wolfgang Schuster

Institut für Genbiologische Forschung, Ihnestraße 63, D-1000 Berlin 33, Germany

Received 2 April 1993

The vast majority of mitochondrial proteins are in all eukaryotes encoded in the nuclear genomes by genes which have been transferred from the original endosymbiont. DNA as well as RNA was and is exchanged between organelles. A functionally successful information transfer, however, requires complex structural and regulatory alterations of the concerned gene. The recently identified variations of the information content in mitochondrial genomes of different plant species represent different stages of the transfer process. These evolutionary intermediates allow a definition of requirements and chances of successful gene transfers.

Endosymbiont hypothesis; Sequence transfer; Plant mitochondria

1. INTRODUCTION

In recent years the endosymbiont hypothesis has become widely accepted to explain the origin of the mitochondrial genome, that represents the residual genome of the former endosymbiotic ancestral organism [1,2]. During evolution most, but not all, essential genetic information of the mitochondrial ancestor has been transferred to the nuclear genome, while part of the information originally contained in the two now cooperating symbionts was no more necessary. The coordination and cohabitation in the symbiosis of two, in plants of three, former independent organisms allowed the loss or divergent evolution of different features and their genetic basis, such as cell walls and compound synthesis for the endosymbiont, and energy conversion for the host cell.

The intermediates of the evolution towards centralizing information in the nucleus have, of course, vanished in the wake of the survivors, but the extent of gene management by the nuclear genome has reached different stages in different species of eukaryotes.

Recent investigations of the mitochondrial genomes in plants show that many more genes have been retained in the mitochondrial genomes of these organisms in comparison to animals [3,4]. The observation of variations in gene content even between different plant mitochondrial genomes has now provided evidence for active gene transfers to the nucleus much more recent than the presumed initial primeval massive gene flux to the nuclear compartment of the host. These observations suggest that a mechanism for translocation of genetic material is still functional and active, albeit possibly distinct from the translocations at the beginning of the endosymbiosis.

Correspondence address: A. Brennicke, Institut für Genbiologische Forschung, Ihnestraße 63, D-1000 Berlin 33, Germany.

2. STEPS OF GENE TRANSLOCATION

The physical gene transfer from mitochondrion to nucleus can be theoretically dissected into several steps (Fig. 1) [5]. These include the availability of genetic material in the organelle, export from the mitochondrion, import into the nucleus, and integration into the nuclear genome. After the now nuclear gene has been activated under the appropriate controls the mitochondrial copy can be inactivated. Its eventual deletion from the organellar genome completes the transfer process.

In different species of higher plants apparent intermediates of the translocation process have been characterized and provide molecular evidence for the physical order of the different steps of the gene transfer process (Fig. 2).

2.1. *Providing the mitochondrial genetic material for transfer*

The genetic unit to be launched from the mitochondrion should be dispensible. This safety measure is inherent in the redundancy of the organelle, which is present a number of times per cell. The organelle itself contains a number of genomes providing a further pool of DNA molecules available for translocation. Additional duplications of complete genes are found within the

 Published by Elsevier Science Publishers B.V.

plant mitochondrial genome. In one instance recombination of a duplicated coding region has occurred within the ATG translation initiation codon [6]. This recombination site would be most appropriate for nuclear integration to ensure an in-frame addition of a mitochondrial target sequence necessary to direct the resulting cytoplasmic protein product to the mitochondrial compartment.

Alternatively the transferred genetic information may be derived from mitochondrial transcripts. In this case the presence of numerous RNA molecules for any transcribed region would ensure an even greater abundance of duplicate copies of the starting material. RNA as starting material would furthermore provide a preselection of transcribed and thus more likely functionally important sequences and ensure that only mature information, i.e. without introns, etc., is translocated.

2.2. *Selection of sequences for transfer*

The mitochondrial sequences identified in the nuclear genomes in plants include different coding regions as well as intron sequences, suggesting that initially any mitochondrial sequence may be a target for the transfer. In this context the investigations of integration of the chloroplast sequences in nuclear genomes are informative. They show that just about the entire plastid genome is integrated in the nuclear genomes of some plants [7,8]. Detailed complete and statistical analysis will be required to detect preferences in selection beyond this observation that any nucleic acid sequence combination can and has been translocated.

2.3. *Transfer of the genetic information*

The physical transfer meets the toughest obstacle in the passage of highly charged nucleic acids through the mitochondrial membranes. Several recent reports provide indirect yet compelling evidence for transport of tRNA molecules into plant mitochondria [9,10]. Not all tRNAs required for translation are encoded in the plant mitochondrial genome and some of the tRNAs found in mitochondria are indeed encoded by nuclear genes. In trypanosomes not only mature nuclear encoded tRNAs, but also their precursors have been found in the mitochondrial compartment [11]. These observations show that transport mechanisms for nucleic acids through the mitochondrial membranes do exist.

Other obvious possibilities are nucleic acids escaping as DNA or RNA before healing and resealing of membrane barriers during occasional disruptions and lesions of the intracellular compartments. Such temporary fusions of normally separate intracellular spaces are reported from electron microscopic observations [12], which are however often difficult to interpret and need further verification in statistically significant numbers.

The final step of the actual transport of nucleic acids is the entry into the nuclear compartment. Since numerous examples of constitutive nucleic acid movement in-

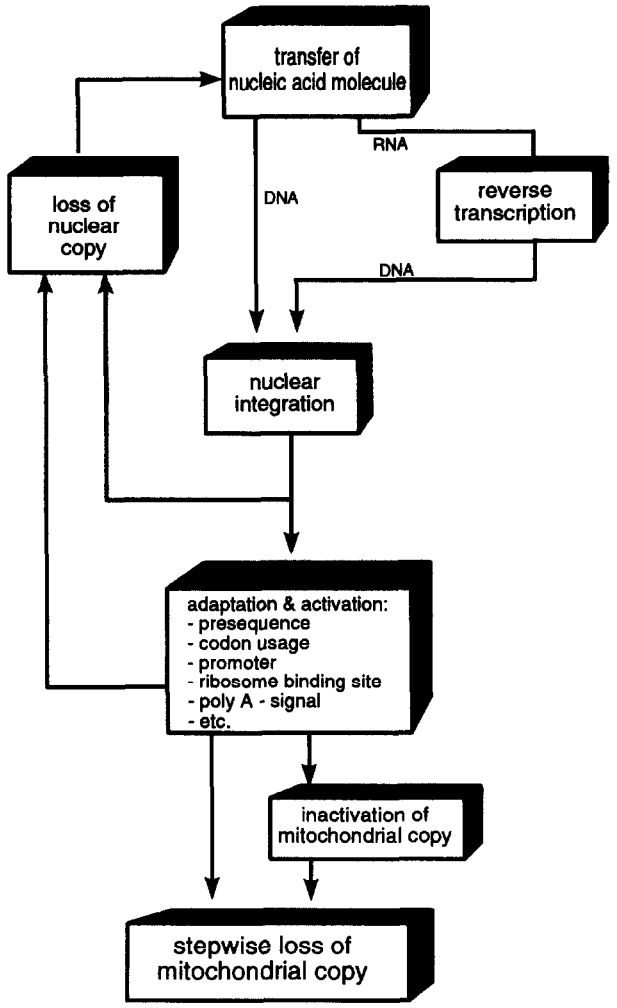

Fig. 1. The transfer of genetic information from mitochondrion to nucleus can be subdivided into different steps. Only complete passage through the entire procedure will result in a successful transfer, with the former organellar gene now actively integrated in the nucleus.

wards through the nuclear pores have been identified, as for example spliceosomal RNAs, the bacterial T-DNA and viral nucleic acids, the mitochondrial nucleic acid should enter the nuclear compartment easily.

2.4. *Integration of the transferred genetic information*

The mechanism and site selection of integration must have an influence on the nature of the 'original' nuclear sequences into which a mitochondrial sequence has been inserted and which now surround this fragment. Hotspots of integration are observed for Ti-plasmids and transposable elements [13], but their modes of selection are unclear at present and may be related to the insertion mechanisms of other, also mitochondrial sequences. The examples of mitochondrial sequences in nuclear DNA reported to date display a great variability of the adjacent sequences, suggesting a random mode of integration. The genomic environments of

successfully transferred genes likewise show no similarity to each other. However, the rapid drift of non-essential nuclear sequences by mutation may distort similarities beyond recognition in evolutionary times. For example, the sequences surrounding the cytochrome oxidase subunit II (*coxII*) genes found in the nucleus of some legumes show in the two investigated examples, cowpea and soybean, very little similarity, although they are presumably derived from a once-upon-a-time singular event during speciation of the legumes [14].

The non-functional mitochondrial sequences presently found in the nuclear genome are located adjacent to genes (V.K., unpublished results). This bias however is most likely due to the bias in sequence data available for the nuclear genome, which consist almost exclusively of coding regions and adjacent sequences. How frequently mitochondrial insertions in other, non-coding genomic nuclear DNA occur, can only be estimated from more random sequence data of that genome or a specific survey of the nuclear DNA for mitochondrial integrations.

2.5. *Activation of the transferred gene*

The translocation of complete genes to the nuclear genome poses special problems with respect to the correct regulation and expression of these genes at their new site.

The product of the now nuclear gene must be correctly targetted to the mitochondrial compartment, which in the case of proteins is most commonly achieved by a signal peptide. In the three reported examples of species-specific successful gene transfers in higher plants, for which sequences upstream of the coreprotein region are available, appropriate extensions of the reading frame have indeed been observed. This upstream extension in both soybean and cowpea, which are likely to be descendants of the same transfer event, is separated from the region homologous to *cox*II by an intron [14,15]. This addition of a functional protein domain via an intron sequence, which is reminiscent of the hypothesis of exon shuffling, possibly allows more freedom for the reading frame transition by evolutionary alterations of the precise setting of the intron borders than a blunt connection of coding sequences. The latter connection, however, has been observed in the third reported gene transfer in higher plants, the gene for the mitochondrial ribosomal protein S12 (*rps12*) in *Oenothera* [16]. Here the 5′ extension of the open reading frame is directly connected to the region homologous to the mitochondrial sequence without a separating intron. In light of the experimental evidence obtained in yeast for the high probability of any open reading frame to act as a correct target sequence for the mitochondrial compartment [17], the chances of inserting a transferred gene into an appropriate frame are fairly good.

It is currently impossible to estimate the chances of creating the correct transcriptional and regulatory environment for a mitochondrial gene in the nucleus of higher plants, since very little data are available. The genomic vicinities of genes coding for mitochondrial proteins such as the adenine nucleotide translocator [18] and the β-subunit of the ATPase [19] show no overt similarities and their regulatory switches yet need to be identified.

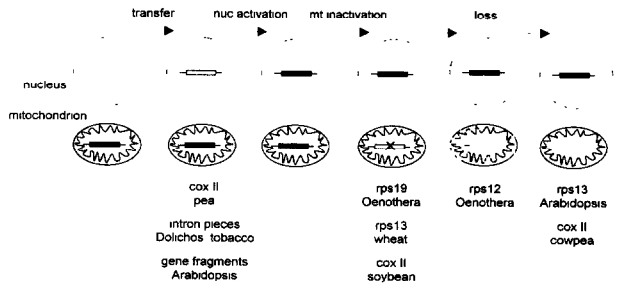

Fig. 2. The different steps of nucleic acid transfer from mitochondrion to nucleus can be seen in intermediates in higher plants. Several originally mitochondrial genes are caught in the process of being translocated to the nucleus with the stages differing between the plant species. Details are given in the text.

The correct regulatory environment is apparently not present a priori in the nuclear genome and just used by insertion of the mitochondrial gene, but it seems to be created during evolution at least in the *coxII* gene in pea [14]. In this plant a nuclear copy of the active mitochondrial gene is not expressed and may thus require adaptation of the genomic environment for functional expression. During the evolutionary drift in the nuclear genome the differences in codon usage between the two compartments, which could also disturb expression, are probably eliminated by selection.

2.6. *The mitochondrial gene becomes superfluous*

2.6.1. Mitochondrial and nuclear genes are active

During the switch between the active genes in the two compartments the duplicated genes in both compartments should be active in a transitory state. As yet no example for this situation has been identified, but most likely will be found eventually in plants.

2.6.2. Inactivation of the mitochondrial gene

After the nuclear gene has been activated, the mitochondrial gene product is no longer required and can be turned off. This situation has first been observed in the fungus *Neurospora* for the *atp9* gene, which is active in the nucleus but present as a silent copy in the mitochondrial genome [20]. In higher plants the example of the *coxII* gene in soybean shows a perfectly intact mitochondrial gene still present [21], while the protein is synthesized from a nuclear gene [15]. Although the mitochondrial copy looks by all criteria functional, little to no transcripts from this gene are detected.

Another potential candidate in higher plants for this intermediate step in gene transfer is the mitochondrial *rps13* gene in wheat [22], for which no transcripts can be detected, while it is actively transcribed in mitochondria of many other plants [4].

Such transcriptional inactivation of apparently intact genes can occur obviously at any level of transcript modulation between initiation of transcription and transcript degradation. Detailed investigations are required to determine the individual features involved in down-regulation of these genes in plant mitochondria.

The causes of interrupted functions are easier to define in several cases of actually disrupted coding regions. In the coding region of the *rps19* gene in *Oenothera* for example, a translational termination codon has been incorporated in the reading frames [23]. The presumably functional nuclear gene copy has not yet been identified.

2.6.3. Partial loss of the mitochondrial gene

Parts of the presumably superfluous mitochondrial gene have been lost from the organelle in one instance of a successful gene transfer. The mitochondrial *rps12* gene has been shuffled to the nucleus in the flowering plant *Oenothera* [16], while about two thirds of the coding region have been deleted from the mitochondrial genome [24]. Homologous recombination across a ten nucleotide repeat has excised upstream sequences and the 5′ half of the open reading frame.

This example suggests that the elimination of the now obsolete mitochondrial gene copy is no active process. It is rather a toleration of no more lethal recombinations of the mitochondrial genome, which is very frequently rearranged by such processes in higher plants. The process of mitochondrial gene elimination may thus occur stepwise by losing bits and pieces before the last fragment has dissappeared. Such partial losses and rearrangements of non-essential gene sequences are also witnessed by the multiple pseudogene sequences present in the mitochondrial genomes of probably all plants besides the intact functional gene copies.

2.6.4. Complete loss of the mitochondrial gene

The process of gene transfer is completed when all evidence of the former mitochondrial gene has dissappeared from the organellar genome. This stage has apparently been reached for most of the mitochondrial proteins, which are now encoded in the nuclear genome of all eukaryotes.

In the lineage of the legumes the presumably single event of the *coxII* transfer has reached this final stage in cowpea and mungbean, in which the mitochondrial copy is made obsolete by a functional nuclear gene [14]. The analogous situation is observed in the *Lycopodiaceae*: the transfer of the *coxIII* gene has allowed the complete elimination of the mitochondrial gene in *Selaginella* (R.H., unpublished results). In *Arabidopsis* the mitochondrial RPS13 protein is most likely encoded by a nuclear locus, since any trace of this gene has dissappeared from the mitochondrial genome (W.S., unpublished observations).

3. WHAT ARE THE CHANCES FOR A FUNCTIONAL GENE TRANSFER?

The examples of intermediates in the gene transfer from mitochondrion to nucleus observed in plants show that the distribution of the genetic material between the different genomes has not yet reached an equilibrium, but is a continuing process. Still unanswered is the academic question why there is a mitochondrial genome left at all and why not all its genes have yet been concentrated in the nucleus. Currently discussed arguments include overloading of the membrane by trespassing hydrophobic – and presumably slow – membrane proteins, unidirectional insertion into the membranes and organized assembly supported by chaperonines, for example, which would require numerous alterations in the switch to import of all mitochondrial proteins.

Some mitochondrial genomes may be frozen in their current state by insurmountable obstacles, as for example the altered genetic codes in most animal mitochondria, which could explain the nearly invariable gene content of these organisms.

The different intermediate stages observed in plants may be interpreted as steps towards this goal of gene transfer. The complete translocation may alternatively not have been achieved simply because evolutionary time has not yet been sufficient. This raises the question of how difficult the transfer process is, how often a successful gene transfer may occur. Such considerations, however, have at present to neglect the problem that certain genes may be disadvantageous for the cell if they are in the nucleus.

3.1. *Functional transfer by RNA*

Specific features of gene expression in plant mitochondria have not been observed in the nucleo-cytoplasmic compartment and thus need to be altered before a mitochondrial gene can be functionally transferred. Mitochondrial introns belong to the organelle specific group II (and in mosses also to group I) [25]. These have never been observed in nuclear genes and therefore presumably have to be eliminated. Posttranscriptional RNA editing considerably alters the genetic information transcribed from the DNA template to ensure synthesis of competent protein products [26–28]. The requirement for RNA editing thus has to be likewise lifted from the transferred gene. The most parsimonious solution to overcome both handicaps appears to be a transfer starting from mature mitochondrial transcripts, which have been spliced and edited. This proposal has been made to explain the observations made with the gene transfer event of *coxII* investigated in the legumes

[14,15] and of *rps12* in *Oenothera* [16]. In these examples the now nuclear gene corresponds in all instances to the edited version of mitochondrial RNA in other plants. For transfer of the genetic unit the mature transcript is assumed to be reverse transcribed either within the organelle, during passage through the cytoplasm or in the nucleus before insertion of the DNA into the nuclear genome (Fig. 1).

Although in many instances the observed mitochondrial or chloroplast nucleic acid sequences in the nuclear genomes of animals and plants are derived by DNA transfer events, a functional gene translocation appears to have to start from mature RNA and to require an intermediate reverse transcription step.

3.2. *Are the translocations random or directed?*

The origin of the features required for correct expression of the transferred gene are more difficult to assess at the moment. The former mitochondrial gene somehow has to acquire the appropriate context for competent expression.

This latter phrase, 'appropriate context', circumscribes a major problem of the transfer which can be solved by two different scenarios, targetted integration and undirected integration subject to trial and error.

In the first pathway distinct recognition between the incoming mitochondrial sequence and the 'appropriate context' must be achieved, to ensure faithful – or at least predominantly 'correct and functional' – integration with little random background of spuriously integrated mitochondrial sequences. In fully targetted and controlled insertion of mitochondrial sequences similar control regions should be found surrounding different inserted former mitochondrial genes. The scattered nuclear genes for mitochondrial ribosomal proteins in yeast indeed show some similarities in binding sites for transcription factors specific for mitochondrial proteins [29]. These elements are, however, very loosely defined and the observed similarities may as likely be the result of convergent evolution as resulting from common insertion sites.

The second possible route of random integration into the nuclear genome postulates many aborted trials that should leave and should have left remnants in different stages of mutational distortion in the nuclear genome. Most of the observations of sequence transfer events are indeed reports of bits and pieces, odds and ends, of the mitochondrial sequences in the nuclear genomes of plants, yeast and animals [30–32]. Here likewise apparently useless and senseless pieces in various stages of deformation have been observed integrated most likely at random in the nuclear genome. The presence of silent nuclear gene copies of *coxIII* and *coxII* in *Lycopodium* (R.H., unpublished observations) and pea, respectively, suggests a gradual gain of the necessary control elements at the site of integration, where these were not a priori present.

This molecular evidence for such potential evolutionary intermediates thus largely sustains the second model of a rather random integration process. However, a preference of direction into target regions cannot be excluded at present.

How often such more or less random insertions have to take place before a functional transfer is achieved, can currently only be debated as an exercise of mind. That such complete gene transfers are not everyday occurrences becomes obvious from the evolutionary time involved in activating the transferred *coxII* gene in the evolutionary line of the legumes [14].

3.3. *How often do sequence transfers occur?*

Sequence transfers from the mitochondrion to the nucleus must be much more frequent than those events that have become manifested and are now detectable, considering the time required to provide the necessary expression signals in the nucleus for a successful gene transfer.

Detection of actual transferred sequences may only be possible to a very low percentage of even stably integrated fragments. Nucleotide sequence drift in the nuclear genome will rapidly evolve non-functional fragments beyond recognition. The few clear-cut examples of nuclear insertions of mitochondrial sequences are indeed recent acquisitions, supporting this scenario. A *nad5* intron fragment in a lectin gene of a bean species is only present in one of the two homologous lectin genes and must have been inserted after the duplication of the ancestral genetic element [33]. Although still comparatively young, this piece of DNA might be retained for its potential involvement in conferring tissue specific expression to the downstream lectin coding region.

The fragment of the mitochondrial *nad1* gene found in one member of the ubiquitin gene family is only present in one *Arabidopsis* ecotype and in none of the other lines [34]. This insertion is thus clearly of very recent origin. It has not had enough evolutionary time to be altered, since it is in both intron and exon sequences nearly identical with the mitochondrial genome. Careful analysis of ascension and ancestors of this ecotype might allow a definite timing of this sequence transfer event.

Further examples of mitochondrial insertions in the nuclear genome might allow a definition of the times of transfer and also of the sequence drift in non-coding regions in the nuclear genome. Such timed mitochondrial sequence transfer may offer a unique chance to determine the actual mutation rate in the nuclear genome, since the evolution of the corresponding mitochondrial original genes is comparatively slow and can be quantitated by comparison to other plant species.

Acknowledgements: The work in the authors' laboratory was supported by grants from the Bundesministerium für Forschung und Technologie and the Deutsche Forschungsgemeinschaft.

REFERENCES

[1] Gray, M.W. (1989) Annu. Rev. Cell Biol. 5, 25–50.
[2] Gray, M.W. (1989) Trends Genet. 5, 294–299.
[3] Leaver, C.J. and Gray, M.W. (1982) Annu. Rev. Plant Physiol. 33, 373–402.
[4] Levings, C.S. III. and Brown, G.G. (1989) Cell 56, 171–179.
[5] Obar, R. and Green, J. (1985) J. Mol. Evol. 22, 243–261.
[6] Schuster, W. and Brennicke, A. (1986) Mol. Gen. Genet. 202, 29–35.
[7] Ayliffe, M.A. and Timmis, J.N. (1992) Mol. Gen. Genet. 236, 105–112.
[8] Pichersky, E., Logson, J.M., McGrath, J.M. and Stasys, R.A. Mol. Gen. Genet. (1991) 225, 453–458.
[9] Marechal-Drouard, L., Weil, J.-H. and Guillemaut, P. (1988) Nucleic Acids Res. 16, 4777–4788.
[10] Small, I., Marechal-Drouard, L., Masson, J., Pelletier, G., Cosset, A., Weil, J.-H. and Dietrich, A. (1992) EMBO J. 11, 1291–1296.
[11] Hancock, K., LeBlanc, A.J., Donze, D. and Hajduk, S.L. (1992) J. Biol. Chem. 267, 23963–23971.
[12] Wildman, S.G., Hongladarum, T. and Honda, S.I. (1962) Science 138, 434–436.
[13] Zambryski, P., Tempe, J. and Schell, J. (1989) Cell 56, 193–201.
[14] Nugent, J.M. and Palmer, J.D. (1991) Cell 66, 473–481.
[15] Covello, P.S. and Gray, M.W. (1992) EMBO J. 11, 3815–3820.
[16] Grohmann, L., Brennicke, A. and Schuster, W. (1992) Nucl. Acids Res. 20, 5641–5646.
[17] Baker, A. and Schatz, G. (1987) Proc. Natl. Acad. Sci. USA 84, 3117–3121.
[18] Winning, B.M., Sarah, C.J., Purdue, P.E., Day, C.D. and Leaver, C.J. (1992) Plant J. 2, 736–773.
[19] Boutry, M. and Chua, N.H. (1985) EMBO J. 5, 2159–2165.
[20] Sebald, W. and Hoppe, J. (1981) Curr. Top. Bioenerg. 12, 1–64.
[21] Grabeau, E.A. (1987) Curr. Genet. 11, 287–293.
[22] Bonen, L. (1987) Nucleic Acids Res. 15, 10393–10404.
[23] Schuster, W. and Brennicke, A. (1991) Nucleic Acids Res. 19, 6923–6928.
[24] Schuster, W., Wissinger, B., Unseld, M. and Brennicke, A. (1990) EMBO J. 9, 263–269.
[25] Ohta, E., Oda, K., Yamato, K., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Ohyama, K. and Michel, F. (1993) Nucleic Acids Res. 21, 1297–1305.
[26] Covello, P.S. and Gray, M.W. (1989) Nature 341, 662–666.
[27] Gualberto, J.M., Lamattina, L., Bonnard, G., Weil, J.H. and Grienenberger, J.M. (1989) Nature 341, 660–662.
[28] Hiesel, R., Wissinger, B., Schuster, W. and Brennicke, A. (1989) Science 246, 1632–1634.
[29] Dorsman, J.C., van Heeswijk, W.C. and Grivell, L. (1988) Nucl. Acids Res. 16, 7287–7301.
[30] Farelly, F. and Butow, R.A. (1983) Nature 301, 296–301.
[31] Gellissen, G., Bradfield, J.Y., White, B.N. and Wyatt, G.R. (1983) Nature 301, 631–634.
[32] Jacobs, H.T., Posakony, J.W., Grula, J.W., Roberts, J.W., Xin, J.-H., Britten, R.J. and Davidson, E.H. (1983) J. Mol. Biol. 165, 609–632.
[33] Knoop, V. and Brennicke, A. (1991) Curr. Genet. 20, 423–425.
[34] Sun, C.-W. and Callis, J. (1993) Plant Cell 5, 97–107.