Running head: MatK gene expression network

Corresponding author: Christian Schmitz-Linneweber, Humboldt University Berlin, Institute of Biology, Molecular Genetics, Chausseestr. 117, 10115 Berlin, Germany. Tel: ++49-30-20938188; Fax: ++49-30-20938141; Email: smitzlic@rz.hu-berlin.de

Research area: Genes, Development and Evolution
Multiple checkpoints for the expression of the chloroplast-encoded splicing factor MatK

Stefanie Hertel¹, Reimo Zoschke², Laura Neumann², Yujiao Qu², Ilka M. Axmann¹, and Christian Schmitz-Linneweber²

¹Institute for Theoretical Biology, Charité - Universitätsmedizin Berlin, Berlin, Germany.
²Molecular Genetics, Institute of Biology, Humboldt-University Berlin, Berlin, Germany.

One-sentence summary:
Expression of a chloroplast encoded splicing factor is regulated on the level of mRNA stability and translation, including a possible auto-regulatory loop.
Financial sources:
This work was funded by the Deutsche Forschungsgemeinschaft (grant number: SCHM 1698/3-1 to CS) and by the German Ministry for Education and Research (BMBF) through the FORSYS partner program (grant no. 0315294 to SH and IMA). RZ is supported by a research fellowship of the German Research Foundation (grant number ZO 302/1-1).

The authors wish it to be known that, in their opinion, the first 2 authors should be regarded as joint First Authors

Present address:

Reimo Zoschke:
Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403, USA.

Laura Neumann:
Charité, Campus Benjamin Franklin, Medizinische Klinik I - Gastroenterologie / Infektiologie / Rheumatologie, Hindenburgdamm 30, D-12200 Berlin, Germany.
ABSTRACT
The chloroplast genome of land plants contains only a single gene for a splicing factor, MatK. To better understand the regulation of matK gene expression, we quantitatively investigated the expression of matK across tobacco (Nicotiana tabacum) development at the transcriptional, post-transcriptional and protein level. We observed striking discrepancies of MatK protein and matK mRNA levels in young tissue suggestive of translational regulation or altered protein stability. We furthermore found increased matK mRNA stability in mature tissue, while other chloroplast RNAs tested showed little changes. Finally, we quantitatively measured MatK-intron interactions and found selective changes in the interaction of MatK with specific introns during plant development. This is evidence for a direct role of MatK in the regulation of chloroplast gene expression via splicing. We furthermore modeled a simplified matK gene expression network mathematically. The model reflects our experimental data and suggests future experimental perturbations to pinpoint regulatory checkpoints.
INTRODUCTION

Two classes of genes dominate the chloroplast genome of land plants: those coding for components of the photosynthetic apparatus and those encoding components of the chloroplast gene expression system. The latter class is represented by a larger number of genes for ribosomal proteins, several genes for subunits of the plastid encoded RNA polymerase and genes for tRNAs and rRNAs. In addition, there is a single gene for a protein involved in splicing, designated matK. matK encodes a protein with homology to bacterial intron maturases. matK, like its bacterial relatives, is located within an intron, in this case within the trnK precursor RNA (Neuhaus and Link, 1987). This intron belongs to a class known as the group II introns, which are characterized by six secondary structure domains (Michel et al., 1989; Lambowitz and Zimmerly, 2004). Maturases are almost always found within domain IV of group II introns. The bacterial maturases are multifunctional proteins that are required for splicing as well as for making their intron mobile (Lambowitz and Zimmerly, 2011). They are with few exceptions specific for their own intron (Carignani et al., 1986; Anziano et al., 1990; Anziano and Butow, 1991; Lambowitz and Zimmerly, 2004) and are capable of down-regulating their own expression (Singh et al., 2002). In the case of MatK, however, genetic and phylogenetic evidence suggest that it aids in splicing multiple introns (Wolfe et al., 1992; Hess et al., 1994; Hübschmann et al., 1996; Vogel et al., 1999; Funk et al., 2007; Duffy et al., 2009; Gao et al., 2009; McNeal et al., 2009). Binding studies indicated an association with the trnK intron at least in vitro (Liere and Link, 1995). Recently, it was shown that MatK associates in vivo with seven group II introns including the trnK intron (Zoschke et al., 2010).

Splicing in land plant chloroplasts is carried out by a multitude of proteins, many of which have been characterized over the last 15 years (Khrouchtchova et al., 2012). Among these splicing factors, MatK is distinguished because it is the only one expressed from the chloroplast genome. All other factors are nuclear-encoded. The matK gene has an unbroken chloroplast heritage in land plants and can in fact also be found in green algae sister groups of embryophytes (Turmel et al., 2006). MatK is a standard molecular marker for phylogenetic studies and has been amplified from tens of thousands of plant species (CBOL_Plant_Working_Group, 2009). It has been lost in parasitic species of the genus Cuscuta (Funk et al., 2007; McNeal et al., 2009) and a parasitic orchid (Delannoy et al., 2011). These plants have also lost the group II introns shown to be associated with the MatK protein, thus making the matK gene dispensable.

Little is known about the expression characteristics of matK or of its intron targets, information that is necessary to appreciate a potential regulatory role of this peculiar splicing factor. We therefore set out to perform a comprehensive expression analysis of MatK and its target RNAs in order to assess possible checkpoints within this small gene expression
network. Furthermore, we developed a mathematical model that was fit to our experimental data. Theoretical predictions derived from simulations using this model indicate that MatK plays a critical role in the splicing of intron-containing chloroplast tRNA genes early in development. Moreover, the model implies an auto-regulatory feedback mechanism required for the idiosyncratic accumulation kinetic of MatK. Testable predictions made by the model are discussed.

RESULTS
MatK protein accumulation is inversely correlated with matK mRNA levels during early tobacco development

It was previously shown by Western analysis of protein extracts prepared from tobacco plants of different ages that MatK displays a complex accumulation pattern across tobacco development (Zoschke et al., 2010).

The main peak of MatK protein accumulation is found at day 7 after imbibition. Here, we extracted RNA from samples of the same age groups used in this previous study to analyze protein accumulation. For plant ages of 3 to 38 days, we used whole seedlings for RNA extraction, excluding roots. For 59-day old plants, only leaf tissue from mature leaves was used. The samples are from plants that had been grown in parallel to those used for protein analyses published previously (Zoschke et al., 2010). With these samples, we performed RNA gel blot hybridization using probes located within the matK reading frame and within exon 2 of tRNA-K(UUU) (Fig. 1). The signals obtained were scored based on their length and whether they are detected with intron- or exon-specific probes. Thus, we identified unspliced precursor RNAs (larger than 2.5 kb; the unspliced tRNA including the matK reading frame is ~2600 nt long), free intron RNA and the mature tRNA. The autoradiographs of two biological replicates were quantified (Fig. 1). The quantification demonstrates that the levels of the mature, spliced tRNA-K(UUU) increase early to reach a plateau after only 7 days of development. By contrast, the precursor RNAs show a more complex accumulation pattern with a main peak in 25 day-old seedlings and a minimum at 7 days. The band at 2.5 kb, presumably the free intron, shows the same pattern as the precursor, suggesting that there is no independent regulation of these two transcript types. Both tRNA precursor and free intron are putative templates for the translation of the matK reading frame. We therefore compared accumulation of these trnK mRNAs with the accumulation of the MatK protein determined previously (Fig. 1, dotted graph). Initially, both mRNA and protein levels start out in parallel (0-3 days post imbibition; hereafter dpi). At day 7 and 11, the two curves show an inverse correlation. In fact, peak MatK protein accumulation coincides with the lowest matK mRNA accumulation at 7 days of tobacco development. Later, mRNA levels increase until 25 days of development while protein levels decline since day 18. This suggests that translation or
protein stability of MatK is independently and, at least during early development, inversely regulated relative to mRNA levels.

**RNA stability of matK increases in mature tobacco seedlings**

The changes identified in the accumulation of tRNA-K(UUU) precursors, i.e. matK mRNA, prompted us to investigate whether they are due to changes in transcription or due to changes in RNA stability. We used run-on transcription assays to analyze transcription rates in 7-day old and 25-day old plants, because these represent the lowest and highest points of matK mRNA accumulation in tobacco development (Fig. 1). Equal amounts of chloroplasts were used for each experiment. Two biological replicates (independent isolations of chloroplasts) were performed and results from one of these replicates are shown in Figure 2A.

Radiolabeled run-on RNA was hybridized to probes representing all chloroplast introns as well as a number of intron-free chloroplast RNAs. Even inspection of the autoradiographs by eye reveals an overall increase in signal in the mature plant sample. This becomes even clearer when the results are quantified and the ratios of run-on transcription signals in 25-day old plants versus 7-day old plants are displayed for the MatK-dependent introns (Fig. 2B): Changes in transcriptional activity vary between 2-fold for trnK and trnA and almost 8-fold for rps12. In order to compare run-on data with changes in steady state levels, we performed RNA gel blot analyses of all seven known MatK target RNAs (trmk, A, I, V, rpl2, rps12, atpF; Fig. 3 A-C; E-G) and one non-MatK target group II intron control (trnG; Fig. 3D). We used the same RNA samples as used to investigate accumulation of trnk and matK (Fig. 1). As for matK, we identified precursor RNAs, spliced RNAs and – where possible – free introns. The pattern for all investigated tRNAs is similar to tRNA-K (UUU) in that the precursors show a pronounced peak at day 25 of development, whereas the mature tRNAs reach a plateau early on during development without much further change. The unspliced mRNA precursors of rpl2, rps12 and atpF also display maximal accumulation on day 18 and day 25 of tobacco development. In contrast to tRNAs, spliced mRNAs closely follow the behavior of unspliced precursor RNAs across development. For rpl2, the free intron does display this behavior as well, while for rps12, the presumable free intron 2 shows an increase in accumulation until the endpoint of the measurements done here. The rps12 intron 2 is the only detected intron here that displays such unusual accumulation behavior, which suggests an active stabilization of this RNA species. Overall, the ratio of spliced to unspliced RNAs is constant for mRNAs across tobacco development but appears regulated for tRNAs.

We next used data points for 7- and 25-day old plants from RNA gel blot hybridization experiments and calculated the 25d/7d ratios of the precursor RNA signals for each probe.
Overall, RNA accumulation ratios resemble ratios from run-on transcription experiments. Both, transcription and RNA accumulation rise in mature tissue for the genes investigated. For most genes, the increase in transcription activity and the increase in RNA accumulation are equal or differ maximally by a factor of two. By contrast, RNA accumulation of the trnK precursor increases 29-fold in the mature tissue, while transcriptional activity only increases 2-fold (Fig. 2B). This suggests that trnK precursor RNA is stabilized in mature tissue. Alternatively, reduced splicing of the precursor in mature tissue increases precursor abundance. In any case, trnK is the most strongly regulated gene in the set investigated here.

**Dynamic association of MatK with different target RNAs across tobacco development**

It has been shown that the RNA binding activity of many RNA binding proteins is regulated (Dreyfuss et al., 2002). This has also been demonstrated for several nuclear splicing factors (e.g. Manley and Tacke, 1996; Huang et al., 2007). In chloroplasts, RNA affinity of certain RNA recognition proteins depends on post-translational modifications in vitro (Lisitsky and Schuster, 1995). With our interest in the regulation of MatK, we sought to determine the association of MatK with its target introns in two developmental stages. We isolated chloroplasts from 7-day old and 25-day old plants carrying an engineered version of MatK with an HA-epitope tag either at the N- or at the C-terminus (transplastomic plants described in Zoschke et al., 2010). The tagged MatK protein was immunoprecipitated from stroma extracts with an anti-HA antibody (Supplemental Fig. S1). RNA was extracted from precipitated and supernatant fractions of the immunoprecipitation and dot-blotted onto a nylon membrane. These membranes were hybridized with radiolabeled probes for the seven known target introns of MatK (Fig. 4A). The experiment was carried out starting from two independent extracts of each, N-terminally and C-terminally tagged MatK:HA plants. Signals obtained after hybridization were quantified using a phosphorimager. While it is difficult to compare the absolute values between pellet signals within individual experiments due to differences in probe lengths and base composition, we can easily score relative enrichment within one developmental stage and also the relative signal changes in the pool of MatK-associated RNAs of 7 day-old plants and 25 day-old plants. To do so, we calculated the mean of pellet-to-supernatant ratios for the two replicate experiments for each probe. This ratio indicates the enrichment of a particular RNA in MatK immunoprecipitation experiments. The ratios are displayed for each of the two developmental stages in a pie chart as fractions of the sum of the ratios of all MatK targets (Fig. 4B). We call these the relative enrichment ratios. They allow scoring differences in immunoprecipitation results within each developmental stage. For example, at day 7, the most strongly enriched intron is the one in trnA, followed by the trnK intron. The least enriched intron is the intron 2 in the rps12 mRNA.
The \textit{atpF} intron is only number five in the list of enrichment ratios at day 7. This changes towards day 25, when enrichment of the \textit{atpF} intron increases and is only topped by the \textit{trnK} intron. To visualize and quantify such changes better, we plotted the ratio of relative enrichment ratios at 25 days over 7 days in Figure 4C. This shows that enrichment of the \textit{atpF} mRNA increases 3-fold in the mature seedlings, which marks the largest change of all MatK-associated introns. Other RNAs show a reduced enrichment like \textit{trnV} and the \textit{rpl2} mRNA. These changes do not correlate with changes in transcript abundance between the two growth stages (Fig. 2B). In sum, this analysis shows that (i) MatK associates with different affinity to different introns and (ii) that affinities change over tobacco development.

**A mathematical model accurately reflects the characteristics of the MatK gene expression network**

The experimental data have shown so far that MatK associates with different RNA precursors forming pre-RNA/MatK complexes including pre-\textit{trnK}/MatK complexes (Zoschke et al., 2010), MatK protein accumulation varies inversely with changes in the \textit{matK} mRNA levels (Fig. 1), and there is no correlation between levels of the spliced \textit{trnK} and the \textit{matK} (\textit{trnK} precursor) RNA (Fig. 1). What mechanism cause these observed inverse correlations? We emphasize that such a question cannot be answered intuitively but requires mathematical models. Here, we developed two models of regulation of the \textit{matK} expression (Fig. 5). In Model 1 (Fig. 5A), we assumed that MatK down-regulates its own production through a negative feedback loop mediated by pre-\textit{trnK}/MatK repression complexes in the light of the fact that direct auto-regulation exists for bacterial maturases (Singh et al., 2002). In Model 2 (Fig. 5C), we considered the case that there is not such feedback regulation of \textit{matK} translation to analyze whether this model can still explain the observed expression dynamics. Both models were built with the intention to use a minimum number of variables and parameters that is nevertheless capable of explaining the dynamics within the \textit{matK} system. Eventually, a model is aimed to identify important regulatory points. To set the stage for our theoretical investigation, we first state our assumptions, also using the simplified models as outlined in Figure 5A/C.

First, we deliberately based our model exclusively on chloroplast genetic information to test whether we still could arrive at an output that is congruent with our experimental data. We do not dispute that nuclear-encoded factors are essential for all steps of chloroplast gene expression including splicing and that these factors can be expected to also exert regulatory control over most individual gene expression steps. However, there is to date little evidence for nuclear factors actively regulating chloroplast gene expression (Raynaud et al., 2007; Boulouis et al., 2011) and chloroplast auto-regulation has been observed for several chloroplast genes (Choquet et al., 2001; Ramundo et al., 2013). Second, we disregarded the
mRNA targets of MatK, such as \textit{atpF}, \textit{rpl2} and \textit{rps12}, and focused only on tRNA-MatK interactions (hereafter called tRNA-MatK splicing network), because these mRNAs are expressed to a much lower level than the tRNAs. In particular, \textit{tmI} and \textit{tmA} are expressed as part of the rRNA operon, which has been shown to have the highest transcription levels of all genes in the plastid genome (Legen et al., 2002; Nakamura et al., 2003). By comparison, \textit{rps12} and \textit{rpl2} transcription and mRNA levels are much lower, close to the detection level (Legen et al., 2002). Third, we assumed that formation of tRNA/MatK complexes is a prerequisite for splicing, since a role of MatK for splicing of these introns is very likely (Zoschke et al., 2010). Dissociation constants of other group II intron binding factors are low and complex recycling is inferred to occur slowly (Saldanha et al., 1999; Wank et al., 1999; Rambo and Doudna, 2004; Ostersetzer et al., 2005). Accordingly, we fourth assumed the same kinetics for tRNA/MatK complexes and neglected any complex dissociation as there are neither data on intron affinities (kDs) for the group II intron binding factor MatK available nor are turnover rates or functions of free introns known. Fifth, we assumed that the tRNAs \textit{tmA}, \textit{tmV}, \textit{tmI} and \textit{tmK} compete for MatK. Sixth, since the levels of \textit{tmA}, \textit{tmV} and \textit{tmI} show similar dynamics, we lumped them into one tRNA species, \textit{tRNA}_{sum} and chose the values of RNA accumulation of \textit{tmA} (Fig. 3A) as a representative example that describes the dynamics of \textit{tRNA}_{sum}. Because of known differences in expression levels between \textit{tmA}, \textit{tmV} and \textit{tmI} and \textit{tmK}, we calculated \textit{tRNA}_{sum} as 10-fold more abundant than \textit{tmK} in our model. Seventh, the processes that cause the observed increase of the \textit{matK} mRNA levels until 25 days of development and their subsequent decline are at present unclear. However, our data on \textit{matK} (\textit{tmK} precursor) RNA and MatK protein accumulation points to the putative rate-limiting role of nuclear-encoded splicing factors at the later stage of development when MatK is no longer rate-limiting (see Discussion). Correspondingly, we modeled this transition in the form of a delay in tRNA/MatK complex formation. This has the advantage of not having to specify all the processes explicitly and thus reduced the numbers of effective variables and parameters to be fixed.

In both models (Fig. 5A/C), the variables \( y_1 \) and \( y_5 \) represent the concentration of the unspliced tRNAs: pre-\textit{tmK-matK} and pre-\textit{tRNA}_{sum}, respectively. The variable \( y_4 \) represents the concentration of the MatK protein. MatK (\( y_4 \)) binds to pre-\textit{tmK-matK} (\( y_1 \)) and pre-\textit{tRNA}_{sum} (\( y_5 \)) transcripts leading to pre-\textit{tmK-matK} (\( y_2 \)) and pre-\textit{tRNA}_{sum}/MatK complexes (\( y_6 \)), respectively. \( y_3 \) describes the concentration of mature (spliced) \textit{tmK} transcripts, \( y_7 \) the amount of mature \textit{tRNA}_{sum} transcripts.

For Model 1, the dynamics of these variables is described by the following system of differential equations:

\[
\frac{dy_1}{dt} = k_{1r} - k_{1d} \cdot y_1(t) - \alpha \cdot (k_{2a} \cdot y_1(t) \cdot y_4(t) + k_{2at} \cdot y_1(t) \cdot y_4((t - \tau)))
\]
\[
\frac{dy_2}{dt} = \alpha \left( k_{2a} \cdot y_1(t) \cdot y_4(t) + k_{2ar} \cdot y_1(t) \cdot y_4(t - \tau) \right) - k_{2d} \cdot y_2(t) \\
\frac{dy_3}{dt} = k_{3} \cdot y_2(t) - k_{3d} \cdot y_3(t) \\
\frac{dy_4}{dt} = \frac{k_{4p} \cdot y_1(t)}{1 + y_2(t)^n} - k_{4d} \cdot y_4(t) \\
\frac{dy_5}{dt} = k_{5} - k_{5d} \cdot y_5(t) - \beta \cdot \left( k_{6a} \cdot y_5(t) \cdot y_4(t) + k_{6ar} \cdot y_5(t) \cdot y_4(t - \tau) \right) - k_{6d} \cdot y_5(t) \\
\frac{dy_6}{dt} = \beta \cdot \left( k_{6a} \cdot y_5(t) \cdot y_4(t) + k_{6ar} \cdot y_5(t) \cdot y_4(t - \tau) \right) - k_{6d} \cdot y_6(t) \\
\frac{dy_7}{dt} = k_{7} \cdot y_6(t) - k_{7d} \cdot y_7(t) \\
\text{with} \\
\alpha = \frac{y_1(t)}{y_1(t) + y_5(t)}, \quad \beta = \frac{y_5(t)}{y_1(t) + y_5(t)}.
\] (8)

Here, the various rate constants parameterize transcription \((k_{1r}, k_{9r})\), degradation \((k_{1d}, k_{7d})\), translation \((k_{4p})\), complex formation \((k_{2a}/k_{2ar}, k_{6a}/k_{6ar})\), and splicing \((k_{3s}, k_{7s})\). The subscript \(\tau\) in \(k_{2ar}\) and \(k_{6ar}\) is the delay introduced in the model to signify the time taken for formation of pre-trnK/matK and pre-tRNA/MatK complexes at the later stage of development. To simulate competition of tRNA precursors for MatK, we include the factors \(\alpha\) and \(\beta\) (Eq.8; see also Eqs. 1, 2, 5 and 6). Hill function with the coefficient \(n\) was used to describe the translational repression term of the MatK protein (Eq. 4).

In Model 2 (Fig. 5C), the equations (1-3, 5-8) are the same but we took auto-regulation out of the equation (4) to analyze whether this model can still explain the observed expression dynamics. The equation (4) changes as follows:

\[
\frac{dy_4}{dt} = k_{4p} \cdot y_1(t) - k_{4d} \cdot y_4(t)
\] (9)

We determined the parameters of both models using an optimization procedure that fit Model 1 and Model 2 to our experimental data including values for RNA accumulation of the trnK-matK precursor and spliced trnK tRNA taken from Northern blot analyses shown in Figure 1 and MatK protein values taken from previous Western blot analyses (Zoschke et al., 2010). The optimal parameters we obtained for both models are listed in the methods section.

With auto-regulation included (Model 1), the dynamical behavior of accumulation of unspliced and spliced transcripts observed later than 7 dpi (Fig. 1) was only achieved when a restart of
complex formation at a later stage (25 dpi) of tobacco development was assumed (Fig. 5B). We modeled this delayed complex formation explicitly as delay differential equation (Eqs. 1, 2, 5 and 6). A model of delay differential equations enables us to group unknown biological processes together to account only for the time required for these processes to occur. Here, the unknown processes are those that cause the expression dynamics of tRNA precursor and mature tRNAs around 25 dpi. As shown in Figure 5B/E, Model 1 produces time courses quantitatively similar to the experimental ones (Figs. 1, 3). Moreover, Model 1 predicts the likely time course of formation of pre-trnK/MatK repression complexes and pre-tRNAsum/MatK complexes. Notably, Model 1 is also able to simulate the time point of maximum production of MatK protein (Fig. 5B/E) although we did not use any direct constraints to arrive at the characteristic shape of its accumulation. We also calculated the ratio to ratio (25d/7d) of the pre-trnK/MatK repression complex (R-R25d/7d=1.2) and again found agreement with experiments (Fig. 4C). Thus, we demonstrate that this minimal regulatory tRNA-MatK network is enough to explain our data.

The model without auto-regulation (Model 2) reflects the expression dynamics of mature tRNAs but fails to reproduce the sharp decline of the MatK protein levels around day 25 (Fig. 5D/E). The observed inverse correlation with the maximum MatK protein accumulation coinciding with minimum matK RNA accumulation at 7 days of tobacco development is lost as well. trnK (matK) precursor RNA and MatK protein show instead a parallel expression dynamics. Their levels increase early in development, reach a plateau and decline after 25 days. Furthermore, Model 2 incorrectly predicts a 2-fold enrichment of pre-trnK/MatK complexes in the mature seedlings. Thus, the model without auto-feedback regulation fails to fit the main characteristics of the MatK gene expression network.

**Modeling suggests that the disruption of auto-regulation leads to an over-accumulation of MatK**

The analysis of the two models suggests that an auto-regulation like in Model 1 is able to explain the distinctive pattern of matK mRNA (trnK precursor RNA) and MatK protein accumulation. We next exploited this model as a predictor for what happens if we perturb the tRNA-MatK splicing system. We were most interested in the putative impact of auto-regulation on the overall expression dynamics predicted by Model 1. To analyze this, we removed auto-regulation from Model 1 (see Methods). Please note that unlike for Model 2 (Fig. 5C/D), this simulation does not involve data fitting and optimization but instead predicts the expression dynamics of the tRNA-MatK network when negative auto-regulation is absent/disrupted.

Experimentally, this prediction could be tested with transplastomic tobacco plants in which matK segments are ectopically over-expressed. If MatK indeed regulates its own translation
it would preferably interact with these over-expressed sequence elements. As a result, less MatK molecules would associate with the endogenous repression site in the \textit{tmK (matK)} precursor RNA causing a disturbance of the negative auto-regulation.

With disrupted auto-regulation, Model 1 predicts a striking over-accumulation of MatK paralleled by a decline in \textit{trnK (matK)} precursor RNA (Supplemental Fig. S2A). Both, the peak of MatK protein levels at 7 days as well as the double peak of \textit{matK} mRNA levels are lost. Moreover, unspliced tRNAs are sharply decreasing while mature tRNAs are increasing which could be explained by increased splicing of tRNAs fostered by MatK and pre-\textit{tmK}/MatK complexes. Most importantly, the simulation suggests that the auto-regulatory feedback seems to prevent an over-production of MatK.

**Development-dependent complex formation could ensure correct timing of intron splicing**

We also asked how increased MatK target tRNA levels (tRNA\textsubscript{sum}) would influence the expression dynamics of the tRNA-MatK splicing network. Constitutive over-expression was simulated through increasing the transcription rate of the pre-tRNA\textsubscript{sum} gene 10-fold and 20-fold (\textit{k}_{5r} in Eq. 5) while keeping the degradation rates (\textit{k}_{1d}, in Eq. 1; \textit{k}_{5d}, in Eq. 5) of the tRNA precursors constant. Model 1 predicts an increase in the levels of tRNA\textsubscript{sum}/MatK complexes and mature tRNA\textsubscript{sum}, respectively (Supplemental Fig. S2B). Furthermore, elevated levels of tRNA\textsubscript{sum} transcripts cause increased accumulation of the MatK protein. An explanation for this finding is that MatK preferentially binds to the tRNA\textsubscript{sum} transcripts due to the excess of tRNA targets. Consequently, the MatK protein synthesis is no longer inhibited by pre-\textit{tmK}/MatK repression complex formation. Interestingly, a delay in maturation of \textit{tmK} transcripts at the early stage of development is predicted for the tRNA\textsubscript{sum} over-expressors relative to the simulated wild type (Supplemental Fig. S2B). Because of this delay, we suppose that early pre-\textit{tmK}/MatK repression complexes are essential to adjust the splicing of RNAs required for chloroplast translation.

In a next step we investigated what effect the absence of early pre-\textit{tmK}/MatK complexes would have on the splicing pattern of other MatK target tRNAs. As expected, the pre-\textit{tmK-matK} level as well as the MatK protein abundance strongly increase, because much less pre-\textit{tmK (matK)} RNAs are spliced (Supplemental Fig. S2C). The RNA is thus free for translation of \textit{matK}. More MatK protein is now available to bind to alternative MatK targets, here tRNA\textsubscript{sum}. As predicted by Model 1, this results in a sharp increase of tRNA\textsubscript{sum}/MatK complexes such that spliced tRNA products already start to saturate at an early stage of plant development. Almost all pre-tRNA\textsubscript{sum} transcripts are used up. When they reach their minimum, MatK binds preferentially to the pre-\textit{tmK-matK} precursors. As a result, the level of
mature \textit{trnK} tRNAs increases rapidly and greatly exceeds the simulated wild-type amount. Interestingly, this simulation strongly indicates that insufficient quantities of pre-\textit{trnK}/MatK repression complexes at an early stage of development lead to over-production of MatK, which is potentially harmful for chloroplast development (see Discussion).

The absence of early tRNA\textsubscript{sum}/MatK complexes as simulated in Supplemental Figure S2D leads to excess quantities of unspliced tRNA\textsubscript{sum} precursors. They are predicted to form complexes with MatK at a later stage of development in order to finally be spliced very fast. By contrast, the pre-\textit{trnK}-\textit{matK} precursors are processed more and earlier in this simulated mutant. Later on, \textit{trnK} maturation is again balanced due to enhanced maturation of the MatK \textit{trans}-tRNA-targets. Interestingly, earlier build-up of MatK/\textit{trnK} repression complexes leads to reduced levels of MatK.

\section*{DISCUSSION AND CONCLUSIONS}

\textbf{MatK gene expression is regulated at the level of protein and RNA accumulation}

The unique position of MatK as the only chloroplast-encoded splicing factor conserved in land plant evolution prompted us to investigate its expression dynamics in plant development to get insights into how MatK might be regulated. We included expression of six further plastid genes in our analysis, because these genes are targets of the splicing factor MatK (Zoschke et al., 2010). Monitoring transcription, RNA accumulation, RNA processing and protein accumulation is necessary to uncover potential regulatory steps in gene expression, i.e. steps where the different processes are not only determined by input of the proceeding process. In this way, we identified a striking discrepancy between the maximum amounts of MatK protein accumulation early in plant development (day 7) and the minimal \textit{matK} mRNA levels at the same time. Most parsimoniously, our data suggest that translation of MatK is induced actively at this stage, although increased protein stability would lead to the same results. Given that it is not easy to detect even steady-state amounts of MatK, it will be a challenge to differentiate between these two options in the future. Noteworthy, 7-day old plants consist almost entirely of cotyledons, while at later stages, primary leaves emerge. As we find comparatively low RNA amounts for all transcripts tested at this stage (Fig. 1, 3), a cotyledon-specific reduction in RNA accumulation has to be considered here. In any case, a strong accumulation of the MatK protein at early stages in plant development seems sensible, as this is a time when chloroplast biogenesis is in full gear. Since this requires plastid translation, MatK as an essential splicing factor for a set of tRNAs and two ribosomal protein coding mRNAs is needed as well.

Next to effects on the protein level, we also found differences between the changes in \textit{matK} transcriptional activity and changes in \textit{matK} RNA accumulation over development. This effect stood out over the much more mild changes found for the other genes tested, suggesting
that matK mRNA is specifically stabilized in mature tissue (Fig. 2B). PPR proteins are known to stabilize selected transcripts in the chloroplasts (e.g. Pfalz et al., 2009), although no information is available that this would be regulated in plant development. For the trnK/matK transcript, a 54kD endonuclease has been shown to bind to the 3‘-UTR and to mediate its processing (Nickelsen and Link, 1993). This or other proteins might contribute to differential stabilization of the matK mRNA.

Age-dependent binding of MatK to target introns

There is to date no evidence that chloroplast RNA splicing would be a rate-limiting step in gene expression in land plants, e.g. no correlation between splicing rates and the amount of the final gene product has been established for chloroplast RNAs. In few cases, tissue specific changes in the abundance of spliced to unspliced mRNAs have been found (Barkan, 1989; McCullough et al., 1992), but whether this is caused by altered splicing rates or via differential stability of spliced RNA versus precursor-RNA was not shown. For tRNAs with introns, changes in splicing would directly impact the amount of the final gene product. Tissue-dependent changes in the ratio of spliced to unspliced transcripts have been described for the mustard trnG intron (Liere and Link, 1994). We here show that precursor mRNAs and mature mRNAs display similar curves of accumulation during development, while tRNA precursors and mature tRNAs show striking differences (Fig. 1A; 3A-D). All five mature tRNAs analyzed reach a plateau early during development, while precursor tRNA levels peak late at 25 dpi. Whether this means that splicing rates decrease later in tobacco development or whether the precursors are stabilized remains to be determined. The observations of potential tissue-specific splicing described above imply changes in the activity of splicing factors. Previously, the amounts of the splicing factor CRS1 have been demonstrate to correlate with splicing of its target intron in the atpF mRNA, incidentally also a target of MatK (Till et al., 2001). MatK levels are high in young tissues when atpF splicing rates are high as well (Fig. 1). Later in development, MatK levels have decreased dramatically, while atpF splicing rates are invariantly high. This on first sight suggests that MatK might be responsible for rate-limiting splicing of atpF only at early stages of development, while later, nuclear factors like CRS1 might limit. We have however gone beyond a simple correlation of splicing factor / target introns and asked whether affinities of MatK for atpF and its other targets change during development. Intriguingly, affinity for atpF increases three-fold in mature tissue, while association with most other introns remains stable or declines slightly (Fig. 4). This suggests that binding of MatK to the atpF intron is a regulated process and that decreased MatK levels could be compensated by changes in RNA affinity. In the absence of indications for MatK protein modifications it will be challenging to understand how such affinity changes could be induced. Again, nuclear factors might be
A model of a simplified MatK gene expression network reflects the experimental data with great precision

We present here a detailed analysis of the kinetics of transcription rates, RNA accumulation, RNA processing, product accumulation and RNA-protein interactions. Identifying key steps inside such a diverse data set is difficult. We therefore used a mathematical approach to gain an improved understanding of the system. To the best of our knowledge, this is the first theoretical investigation applied to problems in organellar gene expression. We investigated two models. In Model 1, we assumed that the MatK protein auto-regulates its own production via pre-trnK/MatK repression complexes. In Model 2, we assumed that there is no such auto-regulation. Both models were fit to our experimental data. We found that only Model 1 could capture the expression dynamics of MatK as well as the MatK targets and even our simulated pre-trnK/MatK-complex ratio (25d/7d) was in perfect agreement with our experimental observations (Fig. 5). A particularly beneficial outcome from simulating of this system is that our mathematical model could produce predictions of the dynamic process of RNA-protein complex formation. Thus, our model predicts a first increase in the amount of RNA/protein complexes at the early stage of development and a second one at day 25. The second temporal peak formation of complexes was explicitly modeled as delay differential equation. Generally, observed dynamics that result from unknown biological processes can be lumped together into a form of time delay (Epstein, 1990). Here, the two different time points of maximum formation of tRNA/MatK complexes may account for a transition in the roles of MatK and nuclear-encoded splicing factors. Indeed, splicing in chloroplast was demonstrated to rely on a multitude of different factors. MatK is only one of them (Khrouchtchova et al., 2012). A simple interpretation is that MatK within the chloroplast “spliceosome” limits the splicing rate at the early stage of development, whereas nuclear-encoded splicing factors become rate-limiting later on, when the MatK protein production is steadily suppressed by pre-trnK/MatK complexes. The explicit delay could reflect the time lag of changes in the composition of RNA/protein complexes with nuclear-encoded splicing factors becoming rate limiting for the splicing process in mature tissue and possibly replacing MatK as the limiting factor for the splicing process. This is a prediction that we propose to test by measuring the temporal variation of composition of the chloroplast splicing machinery on MatK-dependent introns. As more data become available, it will be possible to determine how to include nuclear-encoded splicing factors with intron target ranges overlapping the one of MatK (e.g. RNC1, WTF1, CRS1, WHY1, THA8) into the network model.
**Testable Predictions from the mathematical model**

An intriguing feature of our analysis was that the fit of the model to our experimental data was dramatically improved by the assumption of MatK auto-regulation. The importance of auto-regulation is amenable to straightforward experimental testing, most directly by transferring the gene to the nucleus. Another interesting test of auto-regulation would be *in situ* over-expression of partial *matK* sequences by stable chloroplast transformation. Such an over-expression should lead to a depletion of MatK-repression complexes and thus would free *matK* mRNA for translation. One contact site of MatK has been shown to be located in the 5'-UTR of the *matK* reading frame (Zoschke et al., 2010), a potential site for auto-regulation of intron maturases in general (Singh et al., 2002). More detailed studies on this binding site of MatK would help to choose the right fragment for over-expression. In bacteria, where a maturase in *Lactobacillus lactis* shows auto-regulation, it has been speculated that over-expression of any maturase could be deleterious, because their reverse-transcriptase and endonuclease activities could lead to non-specific harmful effects in the bacterial transcriptome (Singh et al., 2002). However, domains for such activities have been lost in MatK. Nevertheless, MatK could interfere with the splicing of the other non-target group-II introns if over-expressed. Chloroplast introns are known to be degenerated and deviate strongly from the group II consensus sequence, thus potentially allowing for promiscuous binding of splicing factors.

A surprising side-effect of the simulated over-expression of *pre-tRNA*~*sum*~ genes was that the maturation of *trnK* transcripts was slightly delayed, particularly at the early stage of development (see graph "mature *trnK*"; Supplemental Fig. S2B). In contrast, missing quantities of early *pre-tRNA*~*sum*~/MatK complexes could counteract this delay and cause splicing to occur slightly earlier (see graph "mature *trnK*"; Supplemental Fig. S2D). Most importantly, even insufficient amounts of early *pre-trnK*/MatK repression complexes gave rise to potentially harmful over-production of MatK (Supplemental Fig. S2C). Therefore, we reason that early *pre-trnK*/MatK repression complexes appear to (1) assure accumulation of spliced tRNAs in time before they are needed for bulk photosystem translation and (2) prevent excess of MatK proteins. The predicted delay in maturation indicates a developmentally persistent molecular defect in plants with reduced *pre-trnK*/MatK complexes within the first days of chloroplast development. Such mutants are supposed to be delayed in their development from germination onwards. One potential additional role of MatK next to auto-regulation could therefore be to ensure tight translation regulation in chloroplast, starting from the very early developmental stage of the tobacco plant. We chose a minimal set of parameters, which regulates chloroplast tRNA splicing via MatK. Additional components of the chloroplast MatK splicing mechanism almost certainly remain to be considered, but we
believe that our minimal mathematical model is a significant step forward in understanding the role of MatK in the chloroplast splicing system. It aids experimental design and allows the identification of sensitive nodes in the network or the analysis of perturbation effects on the system. The detailed kinetics of RNA/protein complex formation has not been measured. Consequently, our model cannot be regarded as a precise quantitative model of the tRNA-MatK splicing network. However, our model can be extended stepwise by different kinetic mechanisms in order to gain deeper insights into the various underlying processes of MatK-dependent tRNA splicing in chloroplasts.

MATERIALS AND METHODS

Plant material

Previously described transplastomic Nicotiana tabacum (cv. petit havanna) plants with in vivo HA-epitope labeled MatK and control plants with unlabeled MatK were grown in a growth chamber with long day regime (16 h light, 300 μE, 27°C; 8 h dark, 27°C; Zoschke et al., 2010). Plant material was harvested 3 d, 7 d, 11 d, 18 d, 25 d, 38 d, and 59 d after imbibition (in each case 1 h after the end of the dark period). Additionally dry seeds were harvested as pre-imbibition material.

RNA gel blot analyses and run-on transcription

5 μg total RNA was used for RNA gel blot analyses as described previously (Beick et al., 2008). PCR probes were radioactive body labeled with [α-32P]dCTP using the DecaLabel DNA labeling Kit (Fermentas). Oligonucleotide probes were end labeled with [γ-32P]ATP using T4 polynucleotide kinase (Fermentas). The sequences of oligonucleotide probes and the primers used for the amplification of PCR probes are given in Supplemental Table S1. Chloroplasts were isolated from seedlings/plants 7d/25d after imbibition as described previously (Zoschke et al., 2010). 5 x 107 chloroplasts were used for a 10 minute in vivo run-on labeling of transcribed transcripts with [α-32P]UTP at 25 °C following a prior published protocol (Zoschke et al., 2007). As described previously, labeled transcripts were hybridized to a custom macroarray with spotted PCR probes for all plastid introns and several control genes (Beick et al., 2008). Signals were detected and quantified using a PhosphorImaging system and its software (Bio-Rad). Primers used for amplification of PCR probes are shown in Supplemental Table S1.
RNA co-immunoprecipitation and dot-blot analyses

Isolated chloroplasts of 7-d and 25-d old seedlings and plants were lysed and stroma fractions were separated by centrifugation as described previously (Zoschke et al., 2010). A volume of stroma fractions corresponding to a total protein content of 5 mg and 5 μl of anti-HA antiserum (H3663, Sigma-Aldrich) were used for each immunoprecipitation of HA-epitope labeled MatK with Dynabeads protein G (Invitrogen) following manufacturer’s instructions. Dot-blot analyses of coprecipitated RNA of stroma from 7-d/25-d old transplastomic seedlings/plants were carried out as previously described slot-blot analyses (Zoschke et al., 2010). Signals were quantified using a phosphorimager system and its software (Bio-Rad). Enrichment ratios of pellet signals over supernatant signals were calculated for two biological replicates (transplastomic tobacco lines with N- and C-terminal HA-epitope labeled MatK) and compared between experiments with 7-d old seedlings and 25-d old plants. Primers used for amplification of PCR probes are shown in Supplemental Table S1.

Protein protocols

Isolation of proteins from plant tissues was carried out as described (Barkan, 1998). Protein amount of HA-epitope labeled MatK was analyzed by SDS-PAGE and immunoblotting (anti-HA antiserum, Sigma-Aldrich). Quantifications of immunological analyses were carried out as described (Zoschke et al., 2010).

Computational methods

1. Optimization process

Model 1 and Model 2 were implemented using Matlab (R2011b, Mathworks, Cambridge, UK), with the dde23 solver. In order to investigate whether the models sufficiently explain the data we fit these models to the measured data points (see Results) using a least-squares method. The cost function to be optimized is the sum of the least squares

\[ E = \sum_{i=PR, R, P} \left( X_i (t_n) - X_i^{exp} (t_n) \right)^2 \]

with

\[ (9) \]
PR = precursor RNA, R = mature RNA, and P = protein.

We repeated the parameter search from three different initial conditions. For each tested model, three parameter sets were determined. An optimal parameter set was chosen from these three based on how a model fits best to the data points.

In the following, the parameters of the optimal sets are given:
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auto-regulation model</td>
<td>No auto-regulation model</td>
<td></td>
</tr>
<tr>
<td>$k_1r$</td>
<td>0.35 $d^{-1}$</td>
<td>0.34 $d^{-1}$</td>
</tr>
<tr>
<td>$k_1d$</td>
<td>0.02 $d^{-1}$</td>
<td>0.13 $d^{-1}$</td>
</tr>
<tr>
<td>$k_2a$</td>
<td>0.62 $d^{-1}$</td>
<td>0.97 $d^{-1}$</td>
</tr>
<tr>
<td>$k_{2at}$</td>
<td>3.79 $d^{-1}$</td>
<td>8.02 $d^{-1}$</td>
</tr>
<tr>
<td>$k_{2d}$</td>
<td>0.29 $d^{-1}$</td>
<td>0.29 $d^{-1}$</td>
</tr>
<tr>
<td>$k_{3s}$</td>
<td>0.19 $d^{-1}$</td>
<td>0.19 $d^{-1}$</td>
</tr>
<tr>
<td>$k_{3d}$</td>
<td>0.30 $d^{-1}$</td>
<td>0.26 $d^{-1}$</td>
</tr>
<tr>
<td>$k_{4p}$</td>
<td>0.33 $d^{-1}$</td>
<td>0.47 $d^{-1}$</td>
</tr>
<tr>
<td>$k_{4d}$</td>
<td>0.04 $d^{-1}$</td>
<td>0.32 $d^{-1}$</td>
</tr>
<tr>
<td>$n$</td>
<td>90</td>
<td>-</td>
</tr>
<tr>
<td>$k_{5r}$</td>
<td>3.51 $d^{-1}$</td>
<td>2.17 $d^{-1}$</td>
</tr>
<tr>
<td>$k_{5d}$</td>
<td>0.29 $d^{-1}$</td>
<td>0.21 $d^{-1}$</td>
</tr>
<tr>
<td>$k_{6a}$</td>
<td>0.24 $d^{-1}$</td>
<td>0.19 $d^{-1}$</td>
</tr>
<tr>
<td>$k_{6at}$</td>
<td>0.93 $d^{-1}$</td>
<td>0.20 $d^{-1}$</td>
</tr>
<tr>
<td>$k_{7s}$</td>
<td>10.0 $d^{-1}$</td>
<td>10.0 $d^{-1}$</td>
</tr>
<tr>
<td>$k_{7d}$</td>
<td>2.08 $d^{-1}$</td>
<td>0.74 $d^{-1}$</td>
</tr>
<tr>
<td>$\tau$</td>
<td>25 d</td>
<td>25 d</td>
</tr>
</tbody>
</table>

2. \textit{Simulation of the tRNA-MatK gene expression network with disrupted MatK auto-regulation} (Supplemental Fig. S2A)

We took auto-regulation out of the equation (4). The equation (4) changes to equation (9).

3. \textit{Simulation of the tRNA-MatK gene expression network with MatK trans-target genes over-expressed} (Supplemental Fig. S2B)
To simulate an over-production of tRNA\textsubscript{sum} precursor transcript we constitutively increased the rate $k_5$, 10-fold and 20-fold:

for $10X[\text{pre-tRNA}_{\text{sum}}]$, $k_5 = 35.106$ d$^{-1}$,  
for $20X[\text{pre-tRNA}_{\text{sum}}]$, $k_5 = 70.212$ d$^{-1}$.

4. Simulation of the tRNA-MatK gene expression network in the absence of early tRNA/MatK complexes (Supplemental Fig. S2C, D)

To simulate the absence of early pre-trnK/MatK complexes and tRNA\textsubscript{sum}/MatK complexes, the formation rates of pre-trnK/MatK complexes ($k_{2a}$) in equation (2) and pre-tRNA\textsubscript{sum}/MatK complexes ($k_{6a}$) in equation (6) were set to zero.

ACKNOWLEDGEMENTS
The authors would like to thank Thomas Börner and Kenny Watkins for critical reading of the manuscript.

LITERATURE CITED


Till B, Schmitz-Linneweber C, Williams-Carrier R, Barkan A (2001) CRS1 is a novel group II intron splicing factor that was derived from a domain of ancient origin. RNA 7: 1227-1238


Figure Legends

Figure 1. RNA gel blot analysis of matK across tobacco development.
Total RNA (5 μg / lane) from seedlings of different age (days post imbibition – dpi) and from leaves of older plants were extracted, separated on an agarose gel, blotted to a Nylon membrane and hybridized with different radioactive probes. An RNA gel blot analyses with an exon and an intron probe, respectively, were carried out. Bands that could be identified as unspliced precursor RNAs (p), free intron (i), and mature, spliced RNA (m) are indicated on the left. Probe positions for each autoradiogram are indicated on the right. The labeled bands were densitometrically quantified on a phosphorimager. Values were normalized to the top value of all seedling samples, which was arbitrarily set to 1. These values are displayed in the chart shown on top of the panel (black graph = precursor RNA; grey graph = free intron; dashed graph = mature tRNA). The experiment was replicated with independent RNA samples (not shown), which allowed determination of standard deviations shown here. As a loading control, blots were stained with methylene blue prior to hybridization; this is shown below the autoradiograms.
Protein values (dotted line) are taken from Western analyses published previously (Zoschke et al., 2010). Arrows indicate the largest discrepancies in the dynamics of protein and mRNA levels.

Figure 2. Run-on-transcription analysis of all intron-containing genes in the chloroplast and additional controls.
A) RNA was extracted from run-on reactions, which were carried out with chloroplasts purified from seven and 25-day old plants and hybridized on macroarrays. The array comprised probes for all chloroplast introns (In) plus additional controls, including a negative control (pBlue = pBluescript II SK (+)). Probes were spotted as duplicates (only one spot shown). The signals shown here are an example of two biological replicates. The data were used to generate charts in (B). For better visualization, different exposures are shown for different probes (short = dashed line; intermediate = grey lines; and long exposure times =
black lines).

B) Comparison of transcription rates and RNA steady-state levels in 25-day old versus 7-day old material. mRNA levels were assayed by Northern and quantified from two independent experiments (grey bars; Figures 1, 3). Transcription rates were assayed by run-on and quantified from two independent assays (black bars, see A). Note the logarithmic scale of the Y-axis. Assayed genes are indicated below. Standard deviations were derived from two biological replicates.

**Figure 3.** RNA gel blot analysis of MatK target RNAs and *trnG* across tobacco development. Total RNA (5 μg / lane) from seedlings of different age (days post imbibition – dpi) and from leafs of older plants were extracted, separated on an agarose gel, blotted to a Nylon membrane and hybridized with different radioactive probes. Each panel includes two RNA gel blot analyses with an exon and an intron probe, respectively. Bands that could be identified as unspliced precursor RNAs (p), free intron (i), and mature, spliced RNA (m) are indicated on the left. Probe positions for each autoradiogram are indicated on the right. The labeled bands were densitometrically quantified on a phosphorimager. Values were normalized to the top value of all seedling samples, which was arbitrarily set to 1. These values are displayed in the charts shown on top of each panel (black graphs = precursor RNAs; gray graphs [where shown] = free introns; dashed graphs = mature RNAs). The experiment was replicated with independent RNA samples (not shown), which allowed determination of standard deviations shown here. As a loading control, blots were stained with methylene blue prior to hybridization; this is shown below the autoradiograms. (A-D) Probes for tRNAs; (E-G) Probes for mRNAs.

**Figure 4.** Association of MatK with its target RNAs changes across tobacco development.
A) Dot blot analysis of the RNA from pellet (P) and supernatant fraction (S) of MatK:HA immunoprecipitations from the stroma of 7- and 25-day old tobacco seedlings. C+ = plants tagged at the Cterminus of MatK; N+ = plants tagged at the N-terminus of MatK.

B) The ratios of pellet and supernatant signals (P/S) were calculated and summarized for experiments with N+ and C+ plants. Means of these experiments are represented in a pie chart as ratio of the sum of all MatK targets.

C) The ratio to ratio (25d/7d) is indicative of the developmental dependency of RNA-MatK interactions. This chart was derived from the experiment shown in A with immunoprecipitations from C- and N-terminally plants considered as replicate experiments (standard deviations shown as error bars). RNAs with values below 1 show a decrease in the association with MatK:HA in 25-day old plants when compared with 7-day old plants;
opposite for values above 1.

**Figure 5.** Model 1 with auto-regulation of MatK reproduces the observed expression dynamics.

Two possible models of *matK* gene expression were analyzed in respect of how well both models fit to our experimental data.

A) Reaction scheme of the tRNA-MatK splicing network for Model 1. The MatK trans-targets *trnA*, *trnV* and *trnI* are lumped into one tRNA species assigned as tRNA\textsubscript{sum}. The values of *trnA* accumulation (Fig. 3A) were used to describe the expression dynamics of tRNA\textsubscript{sum} and tRNA\textsubscript{sum} is assumed to be 10-fold more abundant than *trnK* (details see Results). \(k_{1r}\) and \(k_{5r}\) are the transcription rates of the tRNA precursors pre-trnK-matK \((y_1)\) and pre-tRNA\textsubscript{sum} \((y_5)\), respectively. The protein MatK \((y_4)\) is encoded within the intron of the *tmK-matK* precursor gene and translated with the rate \(k_{4p}\). As the levels of the tRNA precursors, \(y_1\) and \(y_5\), as well as the MatK protein \((y_4)\) increase, they form pre-trnK/MatK \((y_2)\) and pre-tRNA\textsubscript{sum}/MatK complexes \((y_6)\), respectively. The pre-trnK/MatK repression complex \((y_2)\) inhibits translation of matK via a negative feedback loop (red line with blunt end). The subscript \(\tau\) in \(k_{2ar}\) and \(k_{6ar}\) is the delay introduced in the models to signify the time taken for formation of pre-trnK/MatK and pre-tRNA\textsubscript{sum}/MatK complexes at day 25 of tobacco development. Within the tRNA/protein complexes, MatK splices the introns with rates \(k_{3s}\) and \(k_{7s}\) leading to spliced tRNAs: mature *trnK* \((y_3)\) and mature tRNA\textsubscript{sum} \((y_7)\). Competition of pre-trnK-matK \((y_1)\) and tRNA\textsubscript{sum} \((y_5)\) for MatK \((y_4)\) is assigned by the factors \(\alpha = \frac{y_1}{y_1 + y_5}\) and \(\beta = \frac{y_5}{y_1 + y_5}\). To simplify the model the involvement of nuclear-encoded splicing factors was not considered. Complex dissociation was assumed to be slow and therefore neglected. Dashed arrows represent degradation of RNAs, complexes and the MatK protein with rates \(k_{1d}-k_{7d}\).

B) Accumulation of tRNA precursors, mature tRNAs, MatK protein, and tRNA/MatK complexes during tobacco development derived from theoretical predictions of Model 1 in comparison with Northern blot and Western blot analyses, respectively (E). (top panel) Model 1 fits certain experimental data displayed in (E), such as the inverse correlation of MatK expression and *matK* mRNA (pre-matK-trnK) accumulation around day 7, the sharp decline of the MatK protein level after day 7, and the enrichment of pre-trnK/MatK complexes in mature seedlings. The model consistently reproduces the ratio-to-ratio value between of pre-trnK/MatK complexes at day 25 and day 7 \((RR_{25d/7d}=1.24)\) indicated with black arrows. (lower panel) Temporal expression profile of tRNA\textsubscript{sum} \((trnA, trnV\) and *trnI*). The days of maximal accumulation of tRNA\textsubscript{sum} are the same as those for *trnK* and those experimentally observed, respectively.

C) Model 2 shows the same reaction scheme depicted in (A) but negative auto-regulation is
removed from the system.

D) Accumulation of tRNA precursors, mature tRNAs, MatK protein, and tRNA/MatK complexes during tobacco development derived from theoretical predictions of Model 2 (C) in comparison with Northern blot and Western blot analyses, respectively (E). Model 2 (without auto-regulation) fails to capture the distinctive expression dynamics of matK mRNA (pre-trnK-matk) and MatK protein.

E) Experimental data. Accumulation of tRNA precursors, mature tRNAs, and MatK protein during tobacco development derived from Northern blot and Western blot analyses, respectively. The accumulation of MatK is plotted relative to the maximum obtained 7 days post imbibitions plants (d), while the abundance of the RNA precursor (pre-trnK-matk) and mature trnK is plotted relative to the maximum 25 days and 59 days post imbibitions plants, respectively (set to 1). Reference parameters of the reaction kinetics of Model 1 and Model 2 are given in the method section. Note the different scaling.

SUPPLEMENTAL DATA

Supplemental Data are available online: Supplemental Figure 1, 2; Supplemental Table 1.

Supplemental Figure 1: Immunological analysis of the immunoprecipitation of MatK:HA from stroma fractions of epitope-tagged tobacco plants (C+, N+). About 1/40 of the precipitated and supernatant fractions from 7 and 25 day old tagged lines were separated by SDS-PAGE, blotted onto nitrocellulose membranes and hybridized with an HRP-coupled HA-antibody. The MatK:HA-specific signal is found only in the precipitated fraction (P) of the immunoprecipitations, not in the supernatant (S), indicating complete pull-down of the tagged MatK protein. A Ponceau-stain of the membrane is shown below. kDa: kilodalton. RNA from these immunoprecipitations was used to generate the dot-blot data shown in Figure 4.

Supplemental Figure S2. Predictions of Model 1.
A, Effects of a disrupted translational negative auto-feedback loop on the expression dynamics of the tRNA-MatK splicing network. Model 1 predicts a potentially harmful over-accumulation of MatK. B, Effects of over-expression of MatK target genes on the overall accumulation dynamics within the tRNA-MatK gene expression network. The transcription rate of pre-tRNA_{sum} (trmA, trnV, trnI) was increased 10-fold (10X k_{5r} rate) or 20-fold (20X k_{5r} rate). As a result, tRNAs mature timely delayed and MatK over-accumulates. C,
Accumulation dynamics of the model species in response to missing pre-trnK/MatK repression complexes at the early stage of development. Model 1 predicts an early saturation of spliced tRNA_{sum} transcripts and an increased accumulation of the MatK protein. D, Effects of missing tRNA_{sum}/MatK complexes at the early stage of development on the expression dynamics of the model species. Model 1 predicts that trnK-matK precursors are processed more and earlier whereas excess quantities of tRNA_{sum} precursors occur that are spliced after a time lag.
A)

Group IIA
- trnA  In
- trnl  In
- trnV  In
- atpF  In
- rpl2  In
- rps12l In
- trnK  In
- clpP  In

Group IIB/Group I*
- petB  In
- petD  In
- rps16 In
- rpoC1 In
- ycf3 In1
- clpP In1
- rps12ln1
- rpl16 In2

Controls
- ndhB In
- ndhA In
- ycf3 In1
- trnG In
- psbA
- rnl16
- psbE
- psbD
- trnEYD
- rm23
- rbcl
- pBlue.

B)

Graph showing the ratio of 25d/7d for various genes:
- run-on
- Northern

The graph compares the expression levels of the following genes:
- trnA
- trnl
- trnV
- atpF
- rpl2
- rps12ln1
- trnK
- trnG